THE EXPERT'S VOICE® IN BIG DATA

Beginning Big Data with Pover BI and Excel 2013

TRANSPORT

Neil Dunlop



Beginning Big Data with Power BI and Excel 2013



Neil Dunlop

Apress[®]

Beginning Big Data with Power BI and Excel 2013

Copyright © 2015 by Neil Dunlop

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

ISBN-13 (pbk): 978-1-4842-0530-3

ISBN-13 (electronic): 978-1-4842-0529-7

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image, we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director: Welmoed Spahr Lead Editor: Jonathan Gennick Development Editor: Douglas Pundick Technical Reviewer: Kathi Kellenberger Editorial Board: Steve Anglin, Mark Beckner, Gary Cornell, Louise Corrigan, Jim DeWolf, Jonathan Gennick, Robert Hutchinson, Michelle Lowman, James Markham, Susan McDermott, Matthew Moodie, Jeffrey Pepper, Douglas Pundick, Ben Renow-Clarke, Gwenan Spearing, Matt Wade, Steve Weiss Coordinating Editor: Jill Balzano Copy Editor: Michael G. Laraque Compositor: SPi Global Indexer: SPi Global Artist: SPi Global Cover Designer: Anna Ishchenko

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a Delaware corporation.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales-eBook Licensing web page at www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this text is available to readers at www.apress.com. For additional information about how to locate and download your book's source code, go to www.apress.com/source-code/.

Contents at a Glance

About the Author	xiii
About the Technical Reviewer	XV
Acknowledgments	xvii
Introduction	xix
Chapter 1: Big Data	1
Chapter 2: Excel As Database and Data Aggregator	<mark>15</mark>
Chapter 3: Pivot Tables and Pivot Charts	<mark>35</mark>
Chapter 4: Building a Data Model	<mark>55</mark>
Chapter 5: Using SQL in Excel	77
Chapter 6: Designing Reports with Power View	99
Chapter 7: Calculating with Data Analysis Expressions (DAX)	127
Chapter 8: Power Query	145
Chapter 9: Power Map	173
Chapter 10: Statistical Calculations	<mark>203</mark>
Chapter 11: HDInsight	<mark>225</mark>
Index	

Contents

About the Author	XIII
About the Technical Reviewer	XV
Acknowledgments	xvii
Introduction	xix
Chapter 1: Big Data	1
Big Data As the Fourth Factor of Production	1
Big Data As Natural Resource	1
Data As Middle Manager	<mark>2</mark>
Early Data Analysis	<mark>2</mark>
First Time Line	<mark>2</mark>
First Bar Chart and Time Series	3
Cholera Map	3
Modern Data Analytics	4
Google Flu Trends	4
Google Earth	<mark>5</mark>
Tracking Malaria	<mark>5</mark>
Big Data Cost Savings	5
Big Data and Governments	<mark>5</mark>
Predictive Policing	<mark>5</mark>
A Cost-Saving Success Story	<mark>6</mark>
Internet of Things or Industrial Internet	<mark>6</mark>
Cutting Energy Costs at MIT	<mark>6</mark>

The Big Data Revolution and Health Care	6
The Medicalized Smartphone	7
Improving Reliability of Industrial Equipment	8
Big Data and Agriculture	8
Cheap Storage	8
Personal Computers and the Cost of Storage	8
Review of File Sizes	8
Data Keeps Expanding	9
Relational Databases	9
Normalization	9
Database Software for Personal Computers	10
The Birth of Big Data and NoSQL	11
Hadoop Distributed File System (HDFS)	11
Big Data	11
The Three V's	12
The Data Life Cycle	12
Apache Hadoop	12
CAP Theorem	13
NoSQL	13
Spark	14
Microsoft Self-Service BI	14
Summary	14
Chapter 2: Excel As Database and Data Aggregator	
From Spreadsheet to Database	15
Interpreting File Extensions	
Using Excel As a Database	
Importing from Other Formats	
Opening Text Files in Excel	
Importing Data from XML	19

	Importing XML with Attributes	<mark>20</mark>
	Importing JSON Format	<mark>22</mark>
	Using the Data Tab to Import Data	<mark>23</mark>
	Importing Data from Tables on a Web Site	<mark>23</mark>
	Data Wrangling and Data Scrubbing	<mark>25</mark>
	Correcting Capitalization	<mark>25</mark>
	Splitting Delimited Fields	<mark>26</mark>
	Splitting Complex, Delimited Fields	<mark>29</mark>
	Removing Duplicates	<mark>30</mark>
	Input Validation	31
	Working with Data Forms	32
	Selecting Records	34
	Summary	34
l	Chapter 3: Pivot Tables and Pivot Charts	35
	Recommended Pivot Tables in Excel 2013	35
	Defining a Pivot Table	<mark>36</mark>
	Defining Questions	37
	Creating a Pivot Table	37
	Changing the Pivot Table	<mark>39</mark>
	Creating a Breakdown of Sales by Salesperson for Each Day	40
	Showing Sales by Month	41
	Creating a Pivot Chart	42
	Adjusting Subtotals and Grand Totals	43
	Analyzing Sales by Day of Week	43
	Creating a Pivot Chart of Sales by Day of Week	45
	Using Slicers	47
	Adding a Time Line	48
	Adding a Time Line Importing Pivot Table Data from the Azure Marketplace	48 49
	Adding a Time Line Importing Pivot Table Data from the Azure Marketplace Summary	48 49 54

CONTENTS

Chapter 4: Building a Data Model55
Enabling PowerPivot55
Relational Databases57
Database Terminology57
Creating a Data Model from Excel Tables58
Loading Data Directly into the Data Model62
Creating a Pivot Table from Two Tables66
Creating a Pivot Table from Multiple Tables67
Adding Calculated Columns70
Adding Calculated Fields to the Data Model72
Summary74
Chapter 5: Using SQL in Excel
History of SQL77
NoSQL
NoSQL
NoSQL 77 NewSQL 77 SQL++ 78
NoSQL 77 NewSQL 77 SQL++ 78 SQL Syntax 78
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79Subtotals79
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79Subtotals79Joining Tables80
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79Subtotals79Joining Tables80Importing an External Database80
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79Subtotals79Joining Tables80Importing an External Database80Specifying a JOIN Condition and Selected Fields86
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79Subtotals79Joining Tables80Importing an External Database80Specifying a JOIN Condition and Selected Fields86Using SQL to Extract Summary Statistics89
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79Subtotals79Joining Tables80Importing an External Database80Specifying a JOIN Condition and Selected Fields86Using SQL to Extract Summary Statistics89Generating a Report of Total Order Value by Employee91
NoSQL77NewSQL77SQL++78SQL Syntax78SQL Aggregate Functions79Subtotals79Joining Tables80Importing an External Database80Specifying a JOIN Condition and Selected Fields86Using SQL to Extract Summary Statistics89Generating a Report of Total Order Value by Employee91Using MSQuery94

Chapter 6: Designing Reports with Power View	<mark>99</mark>
Elements of the Power View Design Screen	
Considerations When Using Power View	
Types of Fields	
Understanding How Data Is Summarized	
A Single Table Example	
Viewing the Data in Different Ways	
Creating a Bar Chart for a Single Year	
Column Chart	
Displaying Multiple Years	
Adding a Map	
Using Tiles	
Relational Example	
Customer and City Example	115
Showing Orders by Employee	
Aggregating Orders by Product	
Summary	
Chapter 7: Calculating with Data Analysis Expressions (DAX)	127
Understanding Data Analysis Expressions	
DAX Operators	
Summary of Key DAX Functions Used in This Chapter	
Updating Formula Results	1 <mark>2</mark> 8
Creating Measures or Calculated Fields	
Analyzing Profitability	132
Using the SUMX Function	
Using the CALCULATE Function	

CONTENTS

Calculating the Store Sales for 2009	138
Creating a KPI for Profitability	140
Creating a Pivot Table Showing Profitability by Product Line	142
Summary	144
Chapter 8: Power Query	145
Installing Power Query	145
Key Options on Power Query Ribbon	146
Working with the Query Editor	146
Key Options on the Query Editor Home Ribbon	147
A Simple Population	149
Performance of S&P 500 Stock Index	151
Importing CSV Files from a Folder	155
Group By	1 <u>60</u>
Importing JSON	1 <mark>62</mark>
Summary	172
Chapter 9: Power Map	173
Installing Power Map	173
Plotting a Map	173
Key Power Map Ribbon Options	174
Troubleshooting	175
Plotting Multiple Statistics	
Adding a 2D Chart	184
Showing Two or More Values	191
Creating a 2D Chart	193
Summary	<mark>201</mark>

Chapter 10: Statistical Calculations	<mark>203</mark>
Recommended Analytical Tools in 2013	203
Customizing the Status Bar	205
Inferential Statistics	<mark>206</mark>
Review of Descriptive Statistics	<mark>206</mark>
Calculating Descriptive Statistics	<mark>207</mark>
Measures of Dispersion	<mark>207</mark>
Excel Statistical Functions	<mark>208</mark>
Charting Data	<mark>208</mark>
Excel Analysis ToolPak	208
Enabling the Excel Analysis ToolPak	<mark>208</mark>
A Simple Example	210
Other Analysis ToolPak Functions	<mark>21</mark> 4
Using a Pivot Table to Create a Histogram	<mark>21</mark> 4
Scatter Chart	219
Summary	<mark>224</mark>
Chapter 11: HDInsight	<mark>22</mark> 5
Getting a Free Azure Account	225
Importing Hadoop Files into Power Query	226
Creating an Azure Storage Account	<mark>226</mark>
Provisioning a Hadoop Cluster	<mark>229</mark>
Importing into Excel	<mark>23</mark> 4
Creating a Pivot Table	<mark>238</mark>
Creating a Map in Power Map	<mark>239</mark>
Summary	
Index	<mark>243</mark>

About the Author



Neil Dunlop is a professor of business and computer information systems at Berkeley City College, Berkeley, California. He served as chairman of the Business and Computer Information Systems Departments for many years. He has more than 35 years' experience as a computer programmer and software designer and is the author of three books on database management. He is listed in Marquis's *Who's Who in America*. Check out his blog at http://bigdataondesktop.com/.

About the Technical Reviewer



Kathi Kellenberger, known to the Structured Query Language (SQL) community as Aunt Kathi, is an independent SQL Server consultant associated with Linchpin People and an SQL Server MVP. She loves writing about SQL Server and has contributed to a dozen books as an author, coauthor, or technical editor. Kathi enjoys spending free time with family and friends, especially her five grandchildren. When she is not working or involved in a game of hide-and-seek or Candy Land with the kids, you may find her at the local karaoke bar. Kathi blogs at www.auntkathisql.com.

Acknowledgments

I would like to thank everyone at Apress for their help in learning the Apress system and getting me over the hurdles of producing this book. I would also like to thank my colleagues at Berkeley City College for understanding my need for time to write.

Introduction

This book is intended for anyone with a basic knowledge of Excel who wants to analyze and visualize data in order to get results. It focuses on understanding the underlying structure of data, so that the most appropriate tools can be used to analyze it. The early working title of this book was "Big Data for the Masses," implying that these tools make Business Intelligence (BI) more accessible to the average person who wants to leverage his or her Excel skills to analyze large datasets.

As discussed in Chapter 1, big data is more about volume and velocity than inherent complexity. This book works from the premise that many small- to medium-sized organizations can meet most of their data needs with Excel and Power BI. The book demonstrates how to import big data file formats such as JSON, XML, and HDFS and how to filter larger datasets down to thousands or millions of rows instead of billions.

This book starts out by showing how to import various data formats into Excel (Chapter 2) and how to use Pivot Tables to extract summary data from a single table (Chapter 3). Chapter 5 demonstrates how to use Structured Query Language (SQL) in Excel. Chapter 10 offers a brief introduction to statistical analysis in Excel.

This book primarily covers Power BI—Microsoft's self-service BI tool—which includes the following Excel add-ins:

- 1. *PowerPivot*. This provides the repository for the data (see Chapter 4) and the DAX formula language (see Chapter 7). Chapter 4 provides an example of processing millions of rows in multiple tables.
- 2. *Power View*. A reporting tool for extracting meaningful reports and creating some of the elements of dashboards (see Chapter 6).
- **3.** *Power Query.* A tool to Extract, Transform, and Load (ETL) data from a wide variety of sources (see Chapter 8).
- 4. *Power Map*. A visualization tool for mapping data (see Chapter 9).

Chapter 11 demonstrates how to use HDInsight (Microsoft's implementation of Hadoop that runs on its Azure cloud platform) to import big data into Excel.

This book is written for Excel 2013, but most of the examples it includes will work with Excel 2010, if the PowerPivot, Power View, Power Query, and Power Map add-ins are downloaded from Microsoft. Simply search on download and the add-in name to find the download link.

Disclaimer All links and screenshots were current at the time of writing but may have changed since publication. The author has taken all due care in describing the processes that were accurate at the time of writing, but neither the author nor the publisher is liable for incidental or consequential damages arising from the furnishing or performance of any information or procedures.

CHAPTER 1

Big Data

The goal of business today is to unlock intelligence stored in data. We are seeing a confluence of trends leading to an exponential increase in available data, including cheap storage and the availability of sensors to collect data. Also, the Internet of Things, in which objects interact with other objects, will generate vast amounts of data.

Organizations are trying to extract intelligence from unstructured data. They are striving to break down the divisions between silos. Big data and NoSQL tools are being used to analyze this avalanche of data.

Big data has many definitions, but the bottom line involves extracting insights from large amounts of data that might not be obvious, based on smaller data sets. It can be used to determine which products to sell, by analyzing buying habits to predict what products customers want to purchase. This chapter will cover the evolution of data analysis tools from early primitive maps and graphs to the big data tools of today.

Big Data As the Fourth Factor of Production

Traditional economics, based on an industrial economy, teaches that there are three factors of production: land, labor, and capital. The December 27, 2012, issue of the *Financial Times* included an article entitled "Why 'Big Data' is the fourth factor of production," which examines the role of big data in decision making. According to the article "As the prevalence of Big Data grows, executives are becoming increasingly wedded to numerical insight. But the beauty of Big Data is that it allows both intuitive and analytical thinkers to excel. More entrepreneurially minded, creative leaders can find unexpected patterns among disparate data sources (which might appeal to their intuitive nature) and ultimately use the information to alter the course of the business."

Big Data As Natural Resource

IBM's CEO Virginia Rometty has been quoted as saying "Big Data is the world's natural resource for the next century." She also added that data needs to be refined in order to be useful. IBM has moved away from hardware manufacturing and invested \$30 billion to enhance its big data capabilities.

Much of IBM's investment in big data has been in the development of Watson—a natural language, question-answering computer. Watson was introduced as a *Jeopardy*! player in 2011, when it won against previous champions. It has the computing power to search 1 million books per second. It can also process colloquial English.

One of the more practical uses of Watson is to work on cancer treatment plans in collaboration with doctors. To do this, Watson received input from 2 million pages of medical journals and 600,000 clinical records. When a doctor inputs a patient's symptoms, Watson can produce a list of recommendations ranked in order of confidence of success.

Data As Middle Manager

An April 30, 2015, article in the *Wall Street Journal* by Christopher Mims entitled "Data Is Now the New Middle Manager" describes how some startup companies are substituting data for middle managers. According to the article "Startups are nimbler than they have ever been, thanks to a fundamentally different management structure, one that pushes decision-making out to the periphery of the organization, to the people actually tasked with carrying out the actual business of the company. What makes this relatively flat hierarchy possible is that front line workers have essentially unlimited access to data that used to be difficult to obtain, or required more senior management to interpret." The article goes on to elaborate that when databases were very expensive and business intelligence software cost millions of dollars, it made sense to limit access to top managers. But that is not the case today. Data scientists are needed to validate the accuracy of the data and how it is presented. Mims concludes "Now that every employee can have tools to monitor progress toward any goal, the old role of middle managers, as people who gather information and make decisions, doesn't fit into many startups."

Early Data Analysis

Data analysis was not always sophisticated. It has evolved over the years from the very primitive to where we are today.

First Time Line

In 1765, the theologian and scientist Joseph Priestley created the first time line charts, in which individual bars were used to compare the life spans of multiple persons, such as in the chart shown in Figure 1-1.

hart o 300 30 20 50 Thurch MEN of LEARNING Demonthere Polybie Saller tophane Aristal Hippo Enie crate Catullu STATESMEN Curtes Philip Cicer Alenana misius nilln.s Sull 300 20 20 20 20 8 0 20

Figure 1-1. An early time line chart

First Bar Chart and Time Series

The Scottish engineer William Playfair has been credited with inventing the line, bar, and pie charts. His time-series plots are still presented as models of clarity. Playfair first published *The Commercial and Political Atlas* in London in 1786. It contained 43 time-series plots and one bar chart. It has been described as the first major work to contain statistical graphs. Playfair's *Statistical Breviary*, published in London in 1801, contains what is generally credited as the first pie chart. One of Playfair's time-series charts showing the balance of trade is shown in Figure 1-2.



Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.

The Bottom line is divided into Years, the Right hand line into 1.10,000 each. Sub water 196 by W? Poplar

Figure 1-2. Playfair's balance-of-trade time-series chart

Cholera Map

In 1854, the physician John Snow mapped the incidence of cholera cases in London to determine the linkage to contaminated water from a single pump, as shown in Figure 1-3. Prior to that analysis, no one knew what caused cholera. This is believed to be the first time that a map was used to analyze how disease is spread.



Figure 1-3. Cholera map

Modern Data Analytics

The Internet has opened up vast amounts of data. Google and other Internet companies have designed tools to access that data and make it widely available.

Google Flu Trends

In 2009, Google set up a system to track flu outbreaks based on flu-related searches. When the H1N1 crisis struck in 2009, Google's system proved to be a more useful and timely indicator than government statistics with their natural reporting lags (*Big Data* by Viktor Mayer-Schonberger and Kenneth Cukier [Mariner Books, 2013]). However, in 2012, the system overstated the number of flu cases, presumably owing to media attention about the flu. As a result, Google adjusted its algorithm.

Google Earth

The precursor of Google Earth was developed in 2005 by the computer programmer Rebecca Moore, who lived in the Santa Cruz Mountains in California, where a timber company was proposing a logging operation that was sold as fire prevention. Moore used Google Earth to demonstrate that the logging plan would remove forests near homes and schools and threaten drinking water.

Tracking Malaria

A September 10, 2014, article in the *San Francisco Chronicle* reported that a team at the University of California, San Francisco (UCSF) is using Google Earth to track malaria in Africa and to track areas that may be at risk for an outbreak. According to the article, "The UCSF team hopes to zoom in on the factors that make malaria likely to spread: recent rainfall, plentiful vegetation, low elevations, warm temperatures, close proximity to rivers, dense populations." Based on these factors, potential malaria hot spots are identified.

Big Data Cost Savings

According to a July 1, 2014, article in the *Wall Street Journal* entitled "Big Data Chips Away at Cost," Chris Iervolino, research director at the consulting firm Gartner Inc., was quoted as saying "Accountants and finance executives typically focus on line items such as sales and spending, instead of studying the relationships between various sets of numbers. But the companies that have managed to reconcile those information streams have reaped big dividends from big data."

Examples cited in the article include the following:

- Recently, General Motors made a decision to stop selling Chevrolets in Europe based on an analysis of costs compared to projected sales, based on analysis that took a few days rather than many weeks.
- Planet Fitness has been able to analyze the usage of their treadmills based on their location in reference to high-traffic areas of the health club and to rotate them to even out wear on the machines.

Big Data and Governments

Governments are struggling with limited money and people but have an abundance of data. Unfortunately, most governmental organizations don't know how to utilize the data that they have to get resources to the right people at the right time.

The US government has made an attempt to disclose where its money goes through the web site USAspending.gov. The city of Palo Alto, California, in the heart of Silicon Valley, makes its data available through its web site data.cityofpaloalto.org. The goal of the city's use of data is to provide agile, fast government. The web site provides basic data about city operations, including when trees are planted and trimmed.

Predictive Policing

Predictive policing uses data to predict where crime might occur, so that police resources can be allocated with maximum efficiency. The goal is to identify people and locations at increased risk of crime.

A Cost-Saving Success Story

A January 24, 2011, *New Yorker* magazine article described how 30-something physician Jeffrey Brenner mapped crime and medical emergency statistics in Camden, New Jersey, to devise a system that would cut costs, over the objections of the police. He obtained medical billing records from the three main hospitals and crime statistics. He made block-by-block maps of the city, color-coded by the hospital costs of the residents. He found that the two most expensive blocks included a large nursing home and a low-income housing complex. According to the article, "He found that between January 2002 and June of 2008 some nine hundred people in the two buildings accounted for more than four thousand hospital visits and about two hundred million dollars in health-care bills. One patient had three hundred and twenty-four admissions in five years. The most expensive patient cost insurers \$3.5 million." He determined that 1% of the patients accounted for 30% of the costs.

Brenner's goal was to most effectively help patients while cutting costs. He tried targeting the sickest patients and providing preventative care and health monitoring, as well as treatment for substance abuse, to minimize emergency room visits and hospitalization. He set up a support system involving a nurse practitioner and a social worker to support the sickest patients. Early results of this approach showed a 56% cost reduction.

Internet of Things or Industrial Internet

The Internet of Things refers to machine to machine (M2M) communication involving networked connectivity between devices, such as home lighting and thermostats. CISCO Systems uses the term *Internet* of *Everything*.

An article in the July/August 2014 issue of *Fast Company* magazine described how General Electric Corporation (GE) is using the term *Industrial Internet*, which involves putting intelligence into machines to collect and analyze their data in order to predict failure, so that problems can be anticipated and corrected. One example is the GE jet engine. GE plans to use data generated from engines to predict when an engine part requires repairs, so that the problem can be corrected before a failure occurs. GE's assumption is that small gains in productivity and reliability can drive massive economic benefits.

GE is also working on trip-optimizer, an intelligent cruise control for locomotives, which use trains' geographical location, weight, speed, fuel consumption, and terrain to calculate the optimal velocity to minimize fuel consumption.

Cutting Energy Costs at MIT

An article in the September 28, 2014, *Wall Street Journa*l entitled "Big Data Cuts Buildings' Energy Use" describes how cheap sensors are allowing collection of real-time data on how energy is being consumed. For example, the Massachusetts Institute of Technology (MIT) has an energy war room in which energy use in campus buildings is monitored. Energy leaks can be detected and corrected.

The Big Data Revolution and Health Care

An April 2013 report from McKinsey & Company entitled "The big-data revolution in US health care: Accelerating value and innovation" describes how a big data revolution is occurring in health care, based on pharmaceutical companies aggregating years of research into medical databases and health care providers digitizing patient records. There is a trend toward evidence-based medicine based on data. Big data algorithms are being used to provide the best evidence.

Health care spending currently accounts for more than 17% of US gross domestic product (GDP). McKinsey estimates that implementation of these big data strategies in health care could reduce health expenses in the United States by 12% to 17%, saving between \$300 billion to \$450 billion per year.

"Biological research will be important, but it feels like data science will do more for medicine than all the biological sciences combined," according to the venture capitalist Vinod Khosla, speaking at the Stanford University School of Medicine's Big Data in Biomedicine Conference (quoted in the *San Francisco Chronicle*, May 24, 2014). He went on to say that human judgment cannot compete against machine learning systems that derive predictions from millions of data points. He further predicted that technology will replace 80%–90% of doctors' roles in decision making.

The Medicalized Smartphone

A January 10, 2015, *Wall Street Journal* article reported that "the medicalized smartphone is going to upend every aspect of health care." Attachments to smartphones are being developed that can measure blood pressure and even perform electrocardiograms. Wearable wireless sensors can track blood-oxygen and glucose levels, blood pressure, and heart rhythm. Watches will be coming out that can continually capture blood pressure and other vital signs. The result will be much more data and the potential for virtual physician visits to replace physical office visits.

In December 2013, IDC Health Insights released a report entitled "U.S. Connected Health 2014 Top 10 Predictions: The New Care Delivery Model" that predicts a new health care delivery model involving mobile health care apps, telehealth, and social networking that will provide "more efficient and cost-effective ways to provide health care outside the four walls of the traditional healthcare setting." According to the report, these changes will rely on four transformative technologies:

- 1. Mobile
- 2. Big data analytics
- 3. Social
- 4. Cloud

The report cites the Smartphone Physical project at Johns Hopkins. It uses "a variety of smartphonebased medical devices that can collect quantitative or qualitative data that is clinically relevant for a physical examination such as body weight, blood pressure, heart rate, blood oxygen saturation, visual acuity, optic disc and tympanic membrane images, pulmonary function values, electrocardiogram, heart and lung sounds and ultrasound visualization (such as carotid artery imaging)."

According to the report, "With greater consumer uptake of mobile health, personal health, and fitness modeling, and social technologies, there will be a proliferation of consumer data across diverse data sources that will yield rich information about consumers."

A May 2011 paper by McKinsey & Company entitled "Big data: The next frontier for innovation, competition, and productivity" posits five ways in which using big data can create value, as follows:

- 1. Big data can unlock significant value by making information transparent and usable in much higher frequency.
- **2.** As organizations create and store more transactional data in digital form, they can collect more accurate and detailed performance information on everything from product inventories to sick days, and therefore boost performance.
- **3.** Big data allows ever-narrower segmentation of customers and, therefore, much more precisely tailored products or services.
- 4. Sophisticated analytics can substantially improve decision making.
- 5. Big data can be used to improve the development of the next generation of products and services.

Improving Reliability of Industrial Equipment

General Electric (GE) has made implementing the Industrial Internet a top priority, in order to improve the reliability of industrial equipment such as jet engines. The company now collects 50 million data points each day from 1.4 million pieces of medical equipment and 28,000 jet engines. The goal is to improve the reliability of the equipment. GE has developed Predix, which can be used to analyze data generated by other companies to build and deploy software applications.

Big Data and Agriculture

The goal of precision agriculture is to increase agricultural productivity to generate enough food as the population of the world increases. Data is collected on soil and air quality, elevation, nitrogen in soil, crop maturity, weather forecasts, equipment, and labor costs. The data is used to determine when to plant, irrigate, fertilize, and harvest. This is achieved by installing sensors to measure temperature and the humidity of soil. Pictures are taken of fields that show crop maturity. Predictive weather modeling is used to plan when to irrigate and harvest. The goal is to increase crop yields, decrease costs, save time, and use less water.

Cheap Storage

In the early 1940s, before physical computers came into general use, *computer* was a job title. The first wave of computing was about speeding up calculations. In 1946, the Electrical Numerical Integrator and Computer (ENIAC)—the first general purpose electronic computer—was installed at the University of Pennsylvania. The ENIAC, which occupied an entire room, weighed 30 tons, and used more than 18,000 vacuum tubes, had been designed to calculate artillery trajectories. However, World War II was over by 1946, so the computer was then used for peaceful applications.

Personal Computers and the Cost of Storage

Personal computers came into existence in the 1970s with Intel chips and floppy drives for storage and were used primarily by hobbyists. In August 1981, the IBM PC was released with 5¼-inch floppy drives that stored 360 kilobytes of data. The fact that IBM, the largest computer company in the world, released a personal computer was a signal to other companies that the personal computer was a serious tool for offices. In 1983, IBM released the IBM-XT, which had a 10 megabyte hard drive that cost hundreds of dollars. Today, a terabyte hard drive can be purchased for less than \$100. Multiple gigabyte flash drives can be purchased for under \$10.

Review of File Sizes

Over the history of the personal computer, we have gone from kilobytes to megabytes and gigabytes and now terabytes and beyond, as storage needs have grown exponentially. Early personal computers had 640 kilobytes of RAM. Bill Gates, cofounder of Microsoft Corporation, reportedly said that no one would ever need more than 640 kilobytes. Versions of Microsoft's operating system MS-DOS released during the 1980s could only address 640 kilobytes of RAM. One of the selling points of Windows was that it could address more than 640 kilobytes. Table 1-1 shows how data is measured.

Unit	Power of 2	Approximate Number
Kilobytes	2^10	Thousands
Megabytes	2^20	Millions
Gigabytes	2^30	Billions
Terabytes	2^40	Trillions
Petabytes	2^50	Quadrillions
Exabytes	2^60	Sextillions
Zettabytes	2^70	Septillions

Table 1-1. Measurement of Storage Capacity

Data Keeps Expanding

The New York Stock Exchange generates 4 to 5 terabytes of data every day. IDC estimates that the digital universe was 4.4 petabytes in 2013 and is forecasting a tenfold increase by 2920, to 44 zettabytes.

We are also dealing with exponential growth in Internet connections. According to CISCO Systems, the 15 billion worldwide network connections today are expected to grow to 50 billion by 2020.

Relational Databases

As computers became more and more widely available, more data was stored, and software was needed to organize that data. Relational database management systems (RDBMS) are based on the relational model developed by E. F. Codd at IBM in the early 1970s. Even though the early work was done at IBM, the first commercial RDBMS were released by Oracle in 1979.

A relational database organizes data into tables of rows and columns, with a unique key for each row, called the primary key. A database is a collection of tables. Each entity in a database has its own table, with the rows representing instances of that entity. The columns store values for the attributes or fields.

Relational algebra, first described by Codd at IBM, provides a theoretical foundation for modeling the data stored in relational databases and defining queries. Relational databases support selection, projection, and joins. *Selection* means selecting specified rows of a table based on a condition. *Projection* entails selecting certain specified columns or attributes. *Joins* means joining two or more tables, based on a condition.

As discussed in Chapter 5, Structured Query Language (SQL) was first developed by IBM in the early 1970s. It was used to manipulate and retrieve data from early IBM relational database management systems (RDBMS). It was later implemented in other relational database management systems by Oracle and later Microsoft.

Normalization

Normalization is the process of organizing data in a database with the following objectives:

- 1. To avoid repeating fields, except for key fields, which link tables
- 2. To avoid multiple dependencies, which means avoiding fields that depend on anything other than the primary key

There are several normal forms, the most common being the Third Normal Form (3NF), which is based on eliminating transitive dependencies, meaning eliminating fields not dependent on the primary key. In

CHAPTER 1 🔳 BIG DATA

other words, data is in the 3NF when each field depends on the primary key, the whole primary key, and nothing but the primary key.

Figure 1-4, which is the same as Figure 4-11, shows relationships among multiple tables. The lines with arrows indicate relationships between tables. There are three primary types of relationships.

- 1. One to one (1-1) means that there is a one-to-one correspondence between fields. Generally, fields with a one-to-one correspondence would be in the same table.
- 2. One to many means that for each record in one table, there are many records in the corresponding table. The many is indicated by an arrow at the end of the line. Many means zero to n. For example, as shown in Figure 1-4, for each product code in the products table, there could be many instances of Product ID in the order details table, but each Product ID is associated with only one product code in the products table.
- **3.** Many to many means a relationship from many of one entity to many of another entity. For example, authors and books: Each author can have many books, and each book can have many authors. Many means zero to n.



Figure 1-4. Showing relations among tables

Database Software for Personal Computers

In the 1980s, database programs were developed for personal computers, as they became more widely used. dBASE II, one of the first relational programmable database systems for personal computers, was developed by Wayne Ratliff at the Jet Propulsion Lab (JPL) in Pasadena, California. In the early 1980s, he partnered with George Tate to form the Ashton-Tate company to market dBASE II, which became very successful. In the mid-1980s, dBASE III was released with enhanced features.

dBASE programs were interpreted, meaning that they ran more slowly than a compiled program, where all the instructions are translated to machine language at once. Clipper was released in 1985 as a compiled version of dBASE and became very popular. A few years later, FoxBase, which later became FoxPro, was released with additional enhanced features.

The Birth of Big Data and NoSQL

The Internet was popularized during the 1990s, owing in part to the World Wide Web, which made it easier to use. Competing search engines allowed users to easily find data. Google was founded in 1996 and revolutionized search. As more data became available, the limitations of relational databases, which tried to fit everything into rectangular tables, became clear.

In 1998, Carlo Strozzi used the term *NoSQL*. He reportedly later regretted using that term and thought that *NoRel*, or non-relational, would have been a better term.

Hadoop Distributed File System (HDFS)

Much of the technology of big data came out of the search engine companies. Hadoop grew out of Apple Nutch, an open source web search engine that was started in 2002. It included a web crawler and search engine, but it couldn't scale to handle billions of web pages.

A paper was published in 2003 that described the architecture of Google's Distributed File System (GFS), which was being used by Google to store the large files generated by web crawling and indexing. In 2004, an open source version was released as the Nutch Distributed File System (NDFS).

In 2004, Google published a paper about MapReduce. By 2005, MapReduce had been incorporated into Nutch. MapReduce is a batch query process with the ability to run ad hoc queries against a large dataset. It unlocks data that was previously archived. Running queries can take several minutes or longer.

In February 2006, Hadoop was set up as an independent subproject. Building on his prior work with Mike Cafarella, Doug Cutting went to work for Yahoo!, which provided the resources to turn Hadoop into a system that could be a useful tool for the Web. By 2008, Yahoo! based its search index on a Hadoop cluster. Hadoop was titled based on the name that Doug Cutting's child gave a yellow stuffed elephant. Hadoop provides a reliable, scalable platform for storage and analysis running on cheap commodity hardware. Hadoop is open source.

Big Data

There is no single definition of big data, but there is currently a lot of hype surrounding it, so the meaning can be diluted. It is generally accepted to mean large volumes of data with an irregular structure involving hundreds of terabytes of data into petabytes and higher. It can include data from financial transactions, sensors, web logs, and social media. A more operational definition is that organizations have to use big data when their data processing needs get too big for traditional relational databases.

Big data is based on the feedback economy where the Internet of Things places sensors on more and more equipment. More and more data is being generated as medical records are digitized, more stores have loyalty cards to track consumer purchases, and people are wearing health-tracking devices. Generally, big data is more about looking at behavior, rather than monitoring transactions, which is the domain of traditional relational databases. As the cost of storage is dropping, companies track more and more data to look for patterns and build predictive models.

CHAPTER 1 🔳 BIG DATA

The Three V's

One way of characterizing big data is through the three V's:

- 1. Volume: How much data is involved?
- 2. Velocity: How fast is it generated?
- 3. Variety: Does it have irregular structure and format?

Two other V's that are sometimes used are

- 1. *Variability*: Does it involve a variety of different formats with different interpretations?
- 2. Veracity: How accurate is the data?

The Data Life Cycle

The data life cycle involves collecting and analyzing data to build predictive models, following these steps:

- 1. Collect the data.
- 2. Store the data.
- 3. Query the data to identify patterns and make sense of it.
- 4. Use visualization tools to find the business value.
- 5. Predict future consumer behavior based on the data.

Apache Hadoop

Hadoop is an open source software framework written in Java for distributed storage and processing of very large datasets stored on commodity servers. It has built-in redundancy to handle hardware failures. Hadoop is based on two main technologies: MapReduce and the Hadoop Distributed File System (HDFS). Hadoop was created by Doug Cutting and Mike Cafarella in 2005.

MapReduce Algorithm

MapReduce was developed at Google and released through a 2004 paper describing how to use parallel processing to deal with large amounts of data. It was later released into the public domain. It involves a two-step batch process.

- 1. First, the data is partitioned and sent to mappers, which generate key value pairs.
- 2. The key value pairs are then collated, so that the values for each key are together, and then the reducer processes the key value pairs to calculate one value per key.

One example that is often cited is a word count. It involves scanning through a manuscript and counting the instances of each word. In this example, each word is a key, and the number of times each word is used is the value.

Hadoop Distributed File System (HDFS)

HDFS is a distributed file system that runs on commodity servers. It is based on the Google File System (GFS). It splits files into blocks and distributes them among the nodes of the cluster. Data is replicated so that if one server goes down, no data is lost. The data is immutable, meaning that tables cannot be edited; it can only be written to once, like CD-R, which is based on write once read many. New data can be appended to an existing file, or the old file can be deleted and the new data written to a new file. HDFS is very expandable, by adding more servers to handle more data.

Some implementations of HDFS are append only, meaning that underlying tables cannot be edited. Instead, writes are logged, and then the file is re-created.

Commercial Implementations of Hadoop

Hadoop is an open source Apache project, but several companies have developed their own commercial implementations. Hortonworks, which was founded by Doug Cutting and other former Yahoo! employees, is one of those companies that offers a commercial implementation. Other such companies include Cloudera and MapR. Another Hadoop implementation is Microsoft's HD Insight, which runs on the Azure cloud platform. An example in Chapter 11 shows how to import an HD Insight database into Power Query.

CAP Theorem

An academic way to look at NoSQL databases includes the CAP Theorem. CAP is an acronym for Consistency, Availability, and Partition Tolerance. According to this theorem, any database can only have two of those attributes. Relational databases have consistency and partition tolerance. Large relational databases are weak on availability, because they have low latency when they are large. NoSQL databases offer availability and partition tolerance but are weak on consistency, because of a lack of fixed structure.

NoSQL

As discussed elsewhere in this book, NoSQL is a misnomer. It does not mean "no SQL." A better term for this type of database might be *non-relational*. NoSQL is designed to access data that has no relational structure, with little or no schema. Early versions of NoSQL required a programmer to write a custom program to retrieve that data. More and more implementations of NoSQL databases today implement some form of SQL for retrieval of data. A May 29, 2015, article on Data Informed by Timothy Stephan entitled "What NoSQL Needs Most Is SQL" (http://data-informed.com/what-nosql-needs-most-is-sql/) makes the case for more SQL access to NoSQL databases.

NoSQL is appropriate for "web scale" data, which is based on high volumes with millions of concurrent users.

Characteristics of NoSQL Data

NoSQL databases are used to access non-relational data that doesn't fit into relational tables. Generally, relational databases are used for mission-critical transactional data. NoSQL databases are typically used for analyzing non-mission-critical data, such as log files.

Implementations of NoSQL

There are several categories of NoSQL databases:

- *Key-Value Stores*: This type of database stores data in rows, but the schema may differ from row to row. Some examples of this type of NoSQL database are Couchbase, Redis, and Riak.
- *Document Stores*: This type of database works with documents that are JSON objects. Each document has properties and values. Some examples are CouchDB, Cloudant, and MongoDB.
- *Wide Column Stores*: This type of database has column families that consist of individual column of key-value pairs. Some example of this type of database are HBase and Cassandra.
- *Graph*: Graph databases are good for social network applications. They have nodes that are like rows in a table. Neo4j is an example of this type of database.

One product that allows accessing data from Hadoop is Hive, which provides an SQL-like language called HiveQL for querying Hadoop. Another product is Pig, which uses Pig Latin for querying. The saying is that 10 lines of Pig Latin do the work of 200 lines of Java.

Spark

Another technology that is receiving a lot of attention is Apache Spark—an open source cluster computing framework originally developed at UC Berkeley in 2009. It uses an in-memory technology that purportedly provides performance up to 100 times faster than Hadoop. It offers Spark SQL for querying. IBM recently announced that it will devote significant resources to the development of Spark.

Microsoft Self-Service BI

Most of these big data and NoSQL technologies came out of search engine and online retailers that process vast amounts of data. For the smaller to medium-size organizations that are processing thousands or millions of rows of data, this book shows how to tap into Microsoft's Power BI running on top of Excel.

Much has been written about the maximum number of rows that can be loaded into PowerPivot. The following link to a blog post at Microsoft Trends documents loading 122 million records: www.microsofttrends.com/2014/02/09/how-much-data-can-powerpivot-really-manage-how-about-122-million-records.

Summary

This chapter provides a brief history of the evolution of data analysis, culminating with the big data tools of today. It covers the evolution of data structures, from relational databases to the less structured data that NoSQL addresses.

CHAPTER 2

Excel As Database and Data Aggregator

Spreadsheets have a long history of making data accessible to ordinary people. This chapter chronicles the evolution of Excel from spreadsheet to powerful database. It then shows how to import data from a variety of sources. Subsequent chapters will demonstrate how Excel with Power BI is now a powerful Business Intelligence tool.

From Spreadsheet to Database

The first spreadsheet program, VisiCalc (a contraction of *visible calculator*), was released in 1979 for the Apple II computer. It was the first "killer app." A saying at the time was that "VisiCalc sold more Apples than Apple." VisiCalc was developed by Dan Bricklin and Bob Franken. Bricklin was attending Harvard Business School and came up with the idea for the program after seeing a professor manually write out a financial model. Whenever the professor wanted to make a change, he had to erase the old data and rewrite the new data. Bricklin realized that this process could be automated, using an electronic spreadsheet running on a personal computer.

In those days, most accounting data was trapped in mainframe programs that required a programmer to modify or access. For this reason, programmers were called the "high priests" of computing, meaning that end users had little control over how programs worked. VisiCalc was very easy to use for a program of that time. It was also very primitive, compared to the spreadsheets of today. For example, all columns had to be the same width in the early versions of VisiCalc.

Success breeds competition. VisiCalc did not run on CP/M computers, which were the business computers of the day. CP/M, an acronym originally for *Control Program/Monitor*, but that later came to mean "Control Program for Microcomputers," was an operating system used in the late 1970s and early 1980s. In 1980, Sorcim came out with SuperCalc as the spreadsheet for CP/M computers. Microsoft released Multiplan in 1982. All of the spreadsheets of the day were menu-driven.

When the IBM PC was released in August 1981, it was a signal to other large companies that the personal computer was a serious tool for big business. VisiCalc was ported to run on the IBM PC but did not take into account the enhanced hardware capabilities of the new computer.

Seeing an opportunity, entrepreneur Mitch Kapor, a friend of the developers of VisiCalc, founded Lotus Development to write a spreadsheet specifically for the IBM PC. He called his spreadsheet program Lotus 1-2-3. The name 1-2-3 indicated that it took the original spreadsheet functionality and added the ability to create graphic charts and perform limited database functionality such as simple sorts.

Lotus 1-2-3 was the first software program to be promoted through television advertising. Lotus 1-2-3 became popular hand-in-hand with the IBM PC, and it was the leading spreadsheet through the early 1990s.

Microsoft Excel was the first spreadsheet using the graphical user interface that was popularized by the Apple Macintosh. Excel was released in 1987 for the Macintosh. It was later ported to Windows. In the early 1990s, as Windows became popular, Microsoft packaged Word and Excel together into Microsoft Office and priced it aggressively. As a result, Excel displaced Lotus 1-2-3 as the leading spreadsheet. Today, Excel is the most widely used spreadsheet program in the world.

More and more analysis features, such as Pivot Tables, were gradually introduced into Excel, and the maximum number of rows that could be processed was increased. Using VLOOKUP, it was possible to create simple relations between tables. For Excel 2010, Microsoft introduced PowerPivot as a separate download, which allowed building a data model based on multiple tables. PowerPivot ships with Excel 2013. Chapter 4 will discuss how to build data models using PowerPivot.

Interpreting File Extensions

The file extension indicates the type of data that is stored in a file. This chapter will show how to import a variety of formats into Excel's native .xlsx format, which is required to use the advanced features of Power BI discussed in later chapters. This chapter will deal with files with the following extensions:

.xls: Excel workbook prior to Excel 2007

.xlsx: Excel workbook Excel 2007 and later; the second x was added to indicate that the data is stored in XML format

.xlsm: Excel workbook with macros

.xltm: Excel workbook template

.txt: a file containing text

.xml: a text file in XML format

Using Excel As a Database

By default, Excel works with tables consisting of rows and columns. Each row is a record that includes all the attributes of a single item. Each column is a field or attribute.

A table should show the field names in the first row and have whitespace around it—a blank column on each side and a blank row above and below, unless the first row is row 1 or the first column is A. Click anywhere inside the table and press Ctrl+T to define it as a table, as shown in Figure 2-1. Notice the headers with arrows at the top of each column. Clicking the arrow brings up a menu offering sorting and filtering options.

CHAPTER 2 🔲 EXCEL AS DATABASE AND DATA AGGREGATOR

×	1 5	19		Ŧ		c	h2fig1 - Exce	1				TABLE TOOL	S	
FILE	H	IOME H	11	NSERT PAGE	P	ORMULAS D	ATA REV	/IEW R	VIEW POW	B	Team			
Table Table ·@· Re	Name: 1 size Tabl	e 😳	Ren Cor	nmarize with Piv nove Duplicates nvert to Range	otTable Insert Slicer	Export Refresh	Copen in	es Browser	 Header R Total Rov Banded R 	ow First v Last lows Ban	: Column : Column ded Colum	I Filter	Button	
Pro	perties			Tools		Extern	al Table Data			Table S	tyle Option	s		
C1		*	: [$\times \checkmark f_{2}$	last name	2								
al.	A		В	С	D	E	F	G	Н	I.	J	K	L	M
1				last name 💌	first name 💌	address 🛛 💌	city 💌	state	🕶 zip 💌	amount 💌				
2				Johnson	Ann	1601 Bonita St.	Berkeley	CA	94704	60				
3				Smith	Frank	456 Shattuck	Berkeley	CA	94704	50				
4				Jones	Inger	235 Hearst St	Berkeley	CA	94706	50				
5				Jones	John	123 Main	Oakland	CA	94122	60				
6				Anderson	Mary	234 Eighth St.	Oakland	CA	94134	80	2			
7											「三			
8											000			
9														

Figure 2-1. An Excel table

Note that, with the current version of Excel, if you sort on one field, it is smart enough to bring along the related fields (as long as there are no blank columns in the range), as shown in Figure 2-2, where a sort is done by last name. To sort by first and last name, sort on first name first and then last name.

XI	1 2		;		c	h2fig2 - Exce	1				TABLE TOOLS
FILE	н	OME IN	SERT PAG		ORMULAS DA	ATA REV	/IEW VI	EW POW	/ERPIVOT	Team	DESIGN
ľ	& Cut		Calibri	• 11 •		= %-	🗟 Wrap	o Text	Text		- 🖁 👪
Paste	Forr	y 🔹 nat Painter	в <u>и</u>	• 🖽 • 🔷	- <u>A</u> - = =	≡ ∉ #	E 🖽 Merg	ge & Center	- \$ - 9	6 ° .00	Condition
	Clipboar	d G	1	Font	rs.	Alig	Inment		ra N	umber	G
C1		• :)	×	last nam	e						
.4	A	В	с	D	E	F	G	Н	L	J	к
1			last name 🚽	first name 🔽	address 🛛 🔽	city 💌	state 💌	zip 💌	amount 💌		
2			Anderson	Mary	234 Eighth St.	Oakland	CA	94134	80		
3			Johnson	Ann	1601 Bonita St.	Berkeley	CA	94704	60		
4			Jones	Inger	235 Hearst St	Berkeley	CA	94706	50		
5			Jones	John	123 Main	Oakland	CA	94122	60		
6			Smith	Frank	456 Shattuck	Berkeley	CA	94704	50		
7										<u>/=</u>	
8											
9											

Figure 2-2. Excel Table sorted by first and last name

Importing from Other Formats

Excel can be used to import data from a variety of sources, including data stored in text files, data in tables on a web site, data in XML files, and data in JSON format. This chapter will show you how to import some of the more common formats. Chapter 4 will cover how to link multiple tables using Power BI.

Opening Text Files in Excel

Excel can open text files in comma-delimited, tab-delimited, fixed-length-field, and XML formats, as well as in the file formats shown in Figure 2-3. When files in any of these formats are opened in Excel, they are automatically converted to a spreadsheet.



Figure 2-3. File formats that can be imported into Excel

When importing files, it is best to work with a copy of the file, so that the original file remains unchanged in case the file is corrupted when it is imported.

Figure 2-4 shows a comma-delimited file in Notepad. Note that the first row consists of the field names. All fields are separated by commas. Excel knows how to read this type of file into a spreadsheet. If a text file in this format is opened with Excel, it will appear as a spreadsheet, as shown in Figure 2-1.



Figure 2-4. Comma-separated values format

Importing Data from XML

Extensible Markup Language (XML) is a popular format for storing data. As in HTML, it uses paired tags to define each element. Figure 2-5 shows the data we have been working with in XML format. The first line is the declaration that specifies the version of XML, UTF encoding, and other information. Note the donors root tag that surrounds all of the other tags.

		contribuxmIrev - Notepad	. 🗆 🗙
<u>File Edit Form</u>	nat <u>V</u> iew <u>H</u> elp		
xml version="</td <td>1.0" encoding="UTF-8" standalone="no" ?></td> <td></td> <td>^</td>	1.0" encoding="UTF-8" standalone="no" ?>		^
<	donor>		
	<lname>Johnson</lname>		
	<fname>Ann</fname>		
	<stradur>1601 Bonita St</stradur>		
	(statesCac/states		
	<zip>94704</zip>		
	<amount>60.0</amount>		
<	donor >		
	<lrame>Smith</lrame>		
	<trame>Frank</trame>		
	ccitysBerkelevc/citys		
	<state>CAc/state></state>		
	<zip>94704</zip>		
	<amount>50.0</amount>		
<	donor>		
	<iname>Jones</iname>		
	<trane>inger</trane>		
	<city>Berkelevc/city></city>		
	<state>CA</state>		
	<zip>94706</zip>		
	<amount>50.0</amount>		
<	donor>		
	<iname>Jones</iname>		
	(straddes122 Main/(straddes		
	<city>Oakland</city>		
	<state>CA</state>		
	<zip>94122</zip>		
	<amount>60.0</amount>		
<	donor>		
	cfname>Marvc/fname>		
	<straddr>234 Eighth St.</straddr>		
	<city>Oakland</city>		
	<state>CA</state>		
	<zip>94134</zip>		
	<amount>80.0</amount>		
(Idapane)			
			*
<			>
Contraction of the second seco			

Figure 2-5. XML format

The data is self-descriptive. Each row or record in this example is surrounded by a donor tag. The field names are repeated twice for each field in the opening and closing tags. This file would be saved as a text file with an .xml extension.

An XML file can be opened directly by Excel. On the first pop-up after it opens, select "As an XML table." Then, on the next pop-up, click yes for "Excel will create the schema."

As shown in Figure 2-6, XML files can be displayed in a web browser such as Internet Explorer. Note the small plus and minus signs to the left of the donor tab. Click minus to hide the details in each record. Click plus to show the details.



Figure 2-6. XML file displayed in a web browser

Importing XML with Attributes

As does HTML, XML can have attributes. Figure 2-7 shows an example of adding a gender attribute as part of the donor tag

```
<donor gender = "female">
...
</donor>
```

	contribuxmlrevattributes.xml - Notepad	 ×	
<u>File Edit Format View H</u> elp			7
<pre>Elle Edit Fgrmat View Help <donors> <donor gender="female"> <donor gender="female"> <donor femes="">Johnson <finame>Johnson <finame>Johnson <finame>Johnson <finame>Johnson <donor gender="mail"> <loop and="" state="" the="">Collection <donor gender="female"> <loop and="" state="" the="">Collection <donor gender="female"> <loop and="" state="" the="">Collection </loop></donor> <donor gender="female"> <loop and="" state="" the="">Collection <donor gender="female"> <loop and="" state="" the="">Collection </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </loop></donor> </finame></finame></finame></finame></donor> </donor> </donor> <td>contribuxmlrevattributes.xml - Notepad</td><td> ~</td><td></td></donors></pre>	contribuxmlrevattributes.xml - Notepad	 ~	
<pre></pre>			
<city>Oakland</city>		>	

Figure 2-7. XML with attributes

There are two schools of thought about using attributes. Some people believe that attributes should be avoided and that all values should be presented in paired tags. However, XML with attributes may be encountered. Excel can interpret either format. Figure 2-8 shows the Excel spreadsheet that would be displayed by opening the XML file shown in Figure 2-7.

X) FILE			ERT	PAGE LAYOUT	FORMULAS	fig8 DATA	attribu - Ex	cel W VIEW	POWER	PIVOT	Team	TABL				1						? 🗈
Paste	X Cu Do Co	t py ~	Calibri B I	- 11 - U - 🖂 - 🖄	A A • <u>A</u> •	= = =	- ₩-	Wrap Text	enter -	Text \$ - 9	6,	•	Conditional	Format as	Cell	e Insert	Delete	Format	∑ A	utoSum •	AZY Sort &	Find &
*	Clipboa	ind G		Font	5		Align	ment	G.	N	umber	5	Formatting •	Table * Styles	Styles *	*	Cells		20	Edi	Filter * S ting	elect *
D3		+ : >	< <	<i>f</i> _X 456 Shat	tuck																	
14	A	8	с	D	E	F	G	н	0	J	к	0.0	L M	1	4	0	р		Q	R	s	т
1 8	ender 🔹	Iname 💌	fname	 straddr 	city 💌	state 💌	zip 💌 an	nount 💌														
2 16	male	Jonnson	Ann	1601 Bonita St	Berkeley	CA	94704	60														
4 fe	male	lones	Incer	235 Hearst	Berkeley	CA	94704	50														
5 m	ale	Jones	John	123 Main	Oakland	CA	94122	60														
6 fe	male	Anderson	Mary	234 Eighth St.	Oakland	CA	94134	80														
7																						
8																						
9																						
9																						

Figure 2-8. XML with attributes imported into Excel
Importing JSON Format

Much NoSQL data is stored in JavaScript Object Notation (JSON) format. Unfortunately, Excel does not read JSON files directly, so it is necessary to use an intermediate format to import them into Excel. Figure 2-9 shows our sample file in JSON format. Note the name-value pairs in which each field name is repeated every time there is an instance of that field.



Figure 2-9. JSON format

Because Excel cannot read JSON files directly, you would have to convert this JSON-formatted data into a format such as XML that Excel can read. A number of free converters are available on the Web, such as the one at http://bit.ly/1s20i6L.

The interface for the converter is shown in Figure 2-10. (Remember to work with a copy of the file, so that the original will not be corrupted.) The JSON file can be pasted into the JSON window and then converted to XML by clicking the arrow. The result is an XML file having the same name and that you can open directly from Excel.

CHAPTER 2 EXCEL AS DATABASE AND DATA AGGREGATOR



Figure 2-10. XML and JSON

Using the Data Tab to Import Data

So far, we have been importing data from text files directly into Excel to create a new spreadsheet. The options on the Data tab can be used to import data from a variety of formats, as part of an existing spreadsheet.

Importing Data from Tables on a Web Site

As an example, let's look at importing data from a table on a web site. To import data from a table on a web site, do the following:

- 1. Click the Data tab to display the Data ribbon.
- 2. Click From Web.
- 3. A browser window will appear. Type in the address of the web site you want to access. We will access a Bureau of Labor Statistics example of employment based on educational attainment. For this example, we will use http://www.bls.gov/emp/ep_table_education_summary.htm.
- 4. As shown in Figure 2-11, a black arrow in a small yellow box will appear next to each table. Click the black arrow in the small yellow box at the upper left-hand corner of the table, and it will change to a check mark in a green box. Click Import. The Import Data dialog will appear as shown in Figure 2-12.

CHAPTER 2 EXCEL AS DATABASE AND DATA AGGREGATOR

									- 1.0		
k in next to the tables you w	want to select, then click Impor	t.			t a d	2022				_	
EP NEWS RELEASES	assignment,	2012 8	and pr	ojec	tea	2022					
EP TABLES	Other available formats	: (<u>XLS</u>)									
EP PUBLICATIONS	• Employment by sur	nmary ed	ucation a	nd trai	ining a	assignme	ent, 201	2 and pr	ojected 202	2	
EP FAQS	(Numbers in thousa	ands)		10101000		3					
CONTACT EP		Employment			Cha	nge	Total jo	b openings	;		
SEARCH EP	Education work	Nun	Perc	ent	2012	2-22	replacer 20	nent needs, 12–22	Med		
EP TOPICS	experience, and on- the-job training	2012	2022	2012	2022	Number	Percent	Number	Percent distribution	201	
AGGREGATE ECONOMY	Typical entry-level education										
CLASSIFICATIONS AND	Total, all occupations	145,355.8	160,983.7	100.0	100.0	15,628.0	10.8	50,557.3	100.0	\$34,	
CROSSWALKS	Doctoral or professional degree	4,002.4	4,640.8	2.8	2.9	638.4	16.0	1,426.8	2.8	\$96,	
	Master's degree	2,432.2	2,880.7	1.7	1.8	448.5	18.4	950.8	1.9	\$63,	
REQUIREMENTS MATRIX	Bachelor's degree	26,033.0	29,176.7	17.9	18.1	3,143.6	12.1	8,618.7	17.0	\$67,	
FACTOR ANALYSIS	Associate's degree	5,954.9 7,000.9		4.1	4.3	1,046.0	17.6	2,269.5	4.5	\$57,	
The summer and	· · ·										

Figure 2-11. Importing data from the Web



Figure 2-12. Import Data dialog

5. Select new Worksheet and then OK. The table will be imported, as shown in Figure 2-13. Some formatting may be required in Excel to get the table to be precisely as you want it to be for further analysis.

CI R MARTINGE AVOUT FORMULAS DATA REVIEW VIEW	Book1	- Excel Team		Ц			? 10 - 5 ×
From From Other Existing Connections 21 [2] A Sort Filer To Research Connections 21 [2] Sort Filer To Research Connections 20 [2] Sort Filer To Research Connections 21 [2] Sort Filer To Research Connections 21 [2] Sort Filer To Research Connections 20 [2] Sort Filer To Research Con	ear apply dvanced Colum	o Flash F ins Fill Di	ternove Data Cor aplicates Validation * Data Tools	+= vsolidat	e What-If Analysis	Relations	Show Detail
A1 $\cdot \cdot \cdot \times \checkmark f_x$							*
A	В	с	D	E	F	G	н
1 Employment by summary education and training assignment, 2012 and projected 2022							
2 (Numbers in thousands)	1 P						
3 Education, work experience, and on-the-job training	Employment				Change		Total job openings due to growth and replacemer
4	Number		Percent distribution		2012-22		
5	2012	2022	2012	2022	Number	Percent	Number
6 Typical entry-level education							
7 Total, all occupations	145,355.80	160,983.70	100	100	15,628.00	10.8	
8 Doctoral or professional degree	4,002.40	4,640.80	2.8	2.9	638.4	16	
9 Master's degree	2,432.20	2,880.70	1.7	1.8	448.5	18.4	
10 Bachelor's degree	26,033.00	29,176.70	17.9	18.1	3,143.60	12.1	
11 Associate's degree	5,954.90	7,000.90	4.1	4.3	1,046.00	17.6	
12 Postsecondary non-degree award	8,554.20	9,891.20	5.9	6.1	1,337.10	15.6	
13 Some college, no degree	1,987.20	2,212.20	1.4	1.4	225	11.3	
14 High school diploma or equivalent	58,264.40	62,895.20	40.1	39.1	4,630.80	7.9	
15 Less than high school	38,127.60	42,286.00	26.2	26.3	4,158.40	10.9	
16							
17 Previous work experience in a related occupation							
18 Total, all occupations	145,355.80	160,983.70	100	100	15,628.00	10.8	
19 5 years or more	4,831.90	5,091.80	3.3	3.2	259.9	5.4	
20 Less than 5 years	16,167.70	17,663.50	11.1	11	1,495.90	9.3	
21 None	124,356.20	138,228.40	85.6	85.9	13,872.20	11.2	
22							
23 Typical on-the-job training needed to attain competency							
() Sheet1 (+)	141 355 44	400.003.30	1	100	10.000.00	10.0	
READY			· · · ·				■ ■ ■ + 1075
the second se			The state of the s	_		-	and the second sec

Figure 2-13. Data from the Web imported into Excel

Data Wrangling and Data Scrubbing

Imported data is rarely in the form in which it is needed. Data can be problematic for the following reasons:

- It is incomplete.
- It is inconsistent.
- It is in the wrong format.

Data wrangling involves converting raw data into a usable format. *Data scrubbing* means correcting errors and inconsistencies. Excel provides tools to scrub and format data.

Correcting Capitalization

One of the functions that Excel provides to correct capitalization is the Proper() function. It capitalizes the first letter of each word and lowercases the rest. Figure 2-14 shows some text before and after executing the function. To use the Proper() function, as shown in Figure 2-14, perform the following steps:

- 1. In a new column, type "=Proper(."
- 2. Click the first cell that you want to apply it to and then type the closing parentheses ()). Then press Enter.

- **3.** Click the cell in which you just entered the formula and then drag the fill handle in the lower right corner of the cell, to copy the formula down the column.
- 4. Convert the content of the cells from formulas to values. For a small number of cells, do that by selecting each cell, pressing F2 to edit, then pressing F9 to convert to values, and then pressing Enter.
- **5.** For a larger number of cells, select the column, right-click, and select copy. Then select the top cell in the range to be copied to, press Ctrl+Alt+V, to open the Paste Special Box, select Values, and click OK.
- 6. Then you can delete the original column by right-clicking anywhere in the column and selecting Delete ➤ Entire Column.

FIL		HOME INS	ERT	PAGE LAYOU	T FORM	JLAS [DATA REVI	EW	VIEW PO	WER	Book1 - Exce PIVOT Tear	:l n				NII.		
(Å	Xa	ut	Calibri	-	11 - A /	: = =	· - 16 =	₿ v	Vrap Text		General					+=	*	Σ Aut
Past	Fo	opy * ormat Painter	BI	<u>u</u> • ⊞ •	<u>ð</u> - <u>A</u>	• = =	:≡ (€ 1 2		/lerge & Center	•	\$ - % ,	00. 0 00. 00.	Conditional Formatting *	Format as Table *	Cell Styles *	Insert *	Delete Forma	t Clea
	Clipbo	ard 1%		Font		rs.	Align	ment		r,	Number	5		Styles			Cells	
13		* : >	< 🗸	$f_x = ($	PROPER(F3)													
24	A	В	С	D	E	F	G	н	1		JK	L	M	N		0	P	Q
1																		
2																		
3						iones, fra	nk a		Jones, Fra	nk A								
4						anderson	, mary j		Anderson,	Ma	y J							
5						smith, jan	ef.		Smith, Jan	e F.								
6										2								
7																		
8																		
9																		

Figure 2-14. Use of the Proper() function

Splitting Delimited Fields

In some cases, first and last names are included in the same field and are separated by commas. These and similar fields can be split using the Text to Columns option on the Data tab.

- 1. Highlight the range of cells that you want to convert.
- 2. Click Text to Columns on the Data tab. In the pop-up window, select Delimited, as shown in Figure 2-15. Then click Next.

Convert Text to Columns Wizard - Step 1 of 3 ? ×	
The Text Wizard has determined that your data is Delimited.	
If this is correct, choose Next, or choose the data type that best describes your data.	
Original data type	
Choose the file type that best describes your data:	
Even width Fields are aligned in columns with spaces between each field	
O rixed width - rields are alighed in coldinits with spaces between each neid.	
Preview of selected data: 3 Jones, Frank A 4 Anderson, Mary K 5 Smith, Jane H. 6	
Cancel < Back <u>N</u> ext > <u>F</u> inish]

Figure 2-15. Convert Text to Columns Dialog

3. On the next pop-up window, shown in Figure 2-16, select Comma and Space as the delimiters and review the preview. If the preview looks right, click Next. The reason for including the space as a delimiter is to accommodate middle initials, which are, in this example, separated from their respective first names by a space.

C	convert Text to Columns Wizard - Step 2 of 3	? ×
This screen lets you a in the preview below Delimiters ☐ Iab ☐ Semicolon ✔ Comma ✔ Space ☐ Other: Data preview	set the delimiters your data contains. You can see how your tex v.	t is affected
Jones Fran) Anderson Mary Smith Jane	: A K H.	~
	Cancel < <u>B</u> ack <u>N</u> ext >	<u>F</u> inish

Figure 2-16. Data preview of Text to Columns

4. In the next pop-up window, select each column in the Data Preview window and click Text as the format, as shown in Figure 2-17. Enter the destination where you want the results to appear, then click Finish.

	Co	onvert Te	ext to Co	olumr	ns Wi	zard - S	Step	3 of 3		?	×
This screen let Column data <u>G</u> eneral <u>Text</u> <u>D</u> ate: Do not <u>i</u>	ts you se a format MDY mport co	elect each o	'Gen to da	d set ti eral' co ites, ar	he Data onverts nd all r	a Format numeric emaining <u>A</u> dv	value i value vance	es to nui es to tex d	mber: t.	s, date	values
D <u>e</u> stination: Data <u>p</u> reviev	SHS3										
Text Jones Anderson Smith	Text Frank Mary Jane	Text A K H.									× · · · · · · · · · · · · · · · · · · ·
			Cance	1	<]	<u>B</u> ack		Next >		Ein	ish

Figure 2-17. Last step of Convert Text to Columns

Splitting Complex, Delimited Fields

More complex cases involving multiple data elements in a single field can often be solved through using Excel's built-in functions. Some of the functions that are available for string manipulation are shown in Table 2-1.

Table 2-1. Available Functions for String Manipulation

Function	Description
LEN(text)	Returns the number of characters in the text string
LEFT(text,num_chars)	Returns the leftmost number of characters specified by num_chars
RIGHT(text,num_chars)	Returns the rightmost number of characters specified by num_chars
MID(text,start,num_chars)	Returns the number of characters specified by num_chars, starting from the position specified by start
<pre>SEARCH(find_text, within_text,start_num)</pre>	Returns the position at which the specified find_text starts in within_text when reading from left to right. Not case sensitive. start_num specifies the occurrence of the text that you want. The default is 1.

(continued)

Function	Description
TRIM(text)	Removes all leading and trailing spaces from a text string, except single spaces between words
CLEAN(text)	Removes all non-printable characters from a text string
LOWER(text)	Returns string with all characters converted to lower case
UPPER(text)	Returns string with all characters converted to upper case
PROPER(text)	Returns string with first letter of each word in upper case

Table 2-1. (continued)

Functions can be nested. For example, we could extract the last name by using the following formula:

=LEFT(F7,SEARCH(",",F7)-1)

This function would extract the leftmost characters up to the position of the comma. To extract the first name and middle initial, use the following:

=RIGHT(F7,SEARCH(", ",F7)+1)

Removing Duplicates

It is not unusual for imported data to include duplicate records. Duplicate records can be removed by using the Remove Duplicates option on the Data tab.

It is always a good idea to work with a copy of the original files, to prevent valid data from being deleted. To delete duplicates, simply highlight the columns that contain duplicate values and click Remove Duplicates, as shown in Figure 2-18. Confirm the columns that contain duplicates and click OK. Only the rows in which all the values are the same will be removed.



Figure 2-18. Removing duplicates

Input Validation

Excel provides tools to check for valid input. If data is not valid, the result can be Garbage In, Garbage Out (GIGO). Excel provides data validation tools that are usually associated with data entry; however, they can be used to flag cells from data that has already been entered. In the example shown in Figure 2-19, the valid range for sales figures is 0 to \$20,000, as defined through the Data Validation option on the Data tab.

	н н	• ¢ • OME	# INSERT	PAGE LAYOU	T FOR	MULAS DA	TA R	eview v	IEW	POWER	Book1 PIVOT	- Excel Team				hh ,				()) ?	Nei
From	From Web	From F Text Get Exte	From Other Sources -	Existing Connections	Refresh All -	Connections Properties Edit Links nnections		iort Filter	Filte	Clear Reapply Advanced	Text to Column	o Flash ns Fill	Remove Duplicates	Data Validation Data	Consolid • Tools	ate What- Analysis	Relationships	Group	Ungroup S	tline *∃ s	Show Detail Hide Detail
F2		• :	$\times \checkmark$	fx 100	000																
Sel 1	А	В	с	D	Ε	F	G	н		1	J	к	L	м	N	0	P	Q	R	S	т
1						sales 10000						Data	Validatio	on	?	×					
3 4						8232 7343				Settings	Input Mes	sage E	rror Alert								
5						5678				Validation	criteria										
6						-120				Allow:	umber		I Ignor	e blank							
8						14897				Data:		12									
9										between		~]								
10										Minimum				100.00	<u>.</u>						
11										0	_			181							
13										20000				196	12						
14																					
15										Apply th	iese chang	les to all o	ither cells w	ith the same	e settings						
16										Clear All				OK	Canc	el					
17																					
18																					
19																					
20																					
21																					
22																					

Figure 2-19. Data Validation dialog

Any values outside of the range can be highlighted by selecting the Circle Invalid Data option under Data Validation on the Data tab. The result is shown in Figure 2-20.

K) FILI		Ь - с⇒ номе	- # INSERT	PAGE LAYOU	T FORMUL	AS DATA	REVIEV	v vi	ew poweri	Book1 -	Excel Team				h J					? 🗉
From	From From We	m From b Text Get D	From Other Sources * ternal Data	Existing Connections	Refresh All - Connec	onnections roperties dit Links tions	J Sort	Filter	Te Clear To Reapply Advanced	Text to Columns	Flash Fill	Remove Duplicates	Data Validation Data 1	Consolidate	What-If Analysis	Relationships	Group	Ungroup :	Subtotal	⊟ Show E ∃ Hide D
F2		•	XV	fx 100	000															
1	А	B	c	D	E	F		н	1	J	к	L	Form	nula Bar	0	P	Q	R	S	1.4
1					sa	les														
2						10000														
3						8232														
4						7343														
5						1200														
7						12654														
8						14897														
9					-															
10																				
11																				
12																				

Figure 2-20. Invalid data flagged with red circle

Working with Data Forms

One of the hidden and less documented features in Excel is the Data Forms feature. Sometimes it is easier to read data in which each record is displayed as a form rather than just in tabular format. By default, there is no Data Forms icon on the toolbars. Data Forms can be accessed by clicking a cell in a table with row headings and pressing Alt+D and then O (as in Oscar). Data Forms can also be added to the Quick Access Toolbar. This can be done via the following steps:

1. Click the down arrow at the rightmost end of the Quick Access Toolbar (located at the top-left section of the screen) and click More Commands, as shown in Figure 2-21.

■ 5· ♂·	ch2fig2 - Ex	el	100010	TABLE TOOLS		? 11 -
HOME II Customize Quick Access Toolbar Image: Copy - Paste New New Clipboard Save Email		Wrap Text Wrap Center - Ignment	PIVOT Team	DESIGN Conditional Format as Ce Formatting * Table * Style Styles	Il Insert Delete Format	Na ∑ AutoSum * Ary ## ↓ Fill * Sort & Find & Clear * Filter * Select * Editing
C1 C1 Vice Print Print Print Preview and Print	E F	G H	I J	K L M	N O	P Q R S
2	ghth St. Oakland Bonita St. Berkele	CA 94134 CA 94704	80 60			
5 Sort Descending 6 Touch/Mouse Mode	lain Oakland hattuck Berkele	CA 94122 CA 94704	60 50			
More Commands Show Below the Ribbon						
11						
13 14 15						

Figure 2-21. Customize Quick Access Toolbar menu

2. After clicking More Commands, select "Commands Not in the Ribbon," at the top of the screen. The screen shown in Figure 2-22 will appear. Find and select the icon for forms and press the Add ➤ button in the middle to put it on the Quick Access Toolbar.

CHAPTER 2 EXCEL AS DATABASE AND DATA AGGREGATOR

	ch2fig1 - Excel	TABLE TOOLS	? ×	2 2 - 2 Nei -
FIE HOME INSERT PAGE LAY File HOME INSERT PAGE LAY General Calibri General Format B I II Format Format Format Cipboard G Format I Johnson Anc Jones Jones Jones S Jones Jones G Anderson Mar		ARE TOOLS INS Cuttomice Quick Access Toolbar: For all documents (defauit) Save Sundo Redo Redo	? ×	AutoSum * Are Atom Sort & Find
9 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 24 25 25 25 25 25 25 25 25 25 25	Envire Envire Envire Envire Envire Envire Exse Border Exponentiation Sign Exponentiation Sign Fifther Connections Fifther Connections Fight Protical Fight Protical Fight Protical Formad Fighther Connections Formad Spow Quick Access Toolbar below the Ribbon	Add >> << Bernove	Cancel	

Figure 2-22. Customize the Quick Access Toolbar menu

3. To use data forms, click the row that contains the data to be viewed in the form. Then click the Form icon in the Quick Access Toolbar to see the display shown in Figure 2-23. Fields can be edited in the data form. Click the Find Next and Find Previous buttons to move through the records.

	5· 0·	=			ch2fig1 - E	acel				TAI	BLE TOOL	s		1			? 9	1 -1.5%
FILE	HOME	INSERT PAG	SE LAYOUT	FORMULAS DA	ATA RE	VIEW 1	VIEW PC	WERPIVOT	Tear	m I	DESIGN							Neil -
1 ×	Cut	Calibri	- 11 -	A* A* ≡ =	= %	₽ Wr	ap Text	Gene	ral			4	1		*	∑ AutoSum	27 M	
Paste 💉	Copy * Format Painte	віц	- III - <u>@</u>	• A • = =	≡ (2 +	E B Me	erge & Cente	- \$ -	%,	*3 .3	Cond	itional Forma	t as Cell	Insert	Delete Format	Clear -	Sort & Find & Filter - Select -	
Clipt	board	ra i	Font	15	Alij	gnment		6	Number		5	Styles			Cells	Ec	liting	
13	• : [XVI	fx 50															
1.1.1.72				-			97 SI			3 11	2	12 12 17	101	12.2				
A 1	5	Last name	first name a	E addross a	eitu -	G	H I	amount		J	ĸ	L	M	N	0	P Q	к	5
2		Johnson	Ann	1601 Bonita St.	Berkeley	CA	94704	amount		h2com	natevt	? ×						
3		Smith	Frank	456 Shattuck	Berkeley	CA	94704	_		necconn	DOTEAL							
4		Jones	Inger	235 Hearst St	Berkeley	CA	94706	last name:	Johnso	d	^	1 of 5						
5		Jones	John	123 Main	Oakland	CA	94122	first name:	Ann			New						
6		Anderson	Mary	234 Eighth St.	Oakland	CA	94134					Dulate	-					
7		-						addrgss:	1601 84	onita st.		Deiere						
8								cifte	Berkele	57		Restore						
9								state:	CA									
10								Tin	04704			Find Prev						
11								T.P.	24104			Find Next						
12								amount	60									
13												⊆riteria						
14												Class						
15												close	-					
16																		
17										_	Ŷ							
18																		
19																		
20																		

Figure 2-23. Using a data form

Selecting Records

Data forms can be used to selectively look at certain records based on a criterion by following these steps:

- 1. Click a cell in the table.
- 2. Click the Form icon in the Quick Access Toolbar or press Alt+D, then O.
- **3.** In the form, click Criteria and enter the criteria you want to use, such as City=Oakland, as shown in Figure 2-24.

FILE		• c≯ - HOME	INSERT PA	GE LAYOUT	FORMULAS D	figure21 - E ATA RE	Excel VIEW	VIEW PO	OWERPIVOT	Te	TA	BLE TOOL DESIGN	s						? .	Neil •
Paste	K Cu Ca Fo	it ipy = rmat Paint	Calibri B I U	• 11 • • ⊞• ≜	A A ≡ ≡	= ≫ · = ॡ ₹	₽ Wi	rap Text erge & Cente	Gene	eral 96 •	*.0 .0	Cond Form	itional Forma	t as Cell	e linsert	Delete Format	∑ Auto ▼ Fill •	oSum *	Azv A	
	Clipbo	ard	6	Font	15	Alig	gnment		-ra)	Numbe	tr	G.	Styles			Cells		Editi	ng	
14		*	XV:	fx 50																
al-	A	В	c	D	E	F	G	н	- E	14	1	к	L	м	N	0	P	Q	R	s
1			last name	 first name 	address	city 💌	state	✓ zip	 amount 	¥										
2			Johnson	Ann	1601 Bonita St.	Berkeley	CA	94704			ch2com	matext	? ×							
3			Smith	Frank	456 Shattuck	Berkeley	CA	94704		1		1.12	Collegia							
4			Jones	Inger	235 Hearst St	Berkeley	CA	94706	last name:			1	Criteria							
5			Jones	John	123 Main	Oakland	CA	94122	fjrst name	5			New	1						
6			Anderson	Mary	234 Eighth St.	Oakland	CA	94134	address:				Çlear							
7										Online	- 4		-	5						
8									ciD:	Caka	ina		Bestore							
9									state:			_	Find Prev							
10								1	gipt				time Free							
11									amount				Find Next							
12									- <u>-</u>			-	Form							
13								-					Parte	-						
14													Close							
15																				
16												~								
17									-											
18																				
19																				
20																				

Figure 2-24. Using a data form to select records based on a criterion

4. Then click Find Next, and Find Previous will only show rows that meet the designated criteria.

Summary

This chapter has shown you how to import data into Excel. Figures 2-1, 2-2, 2-4, 2-5, 2-6, 2-7, 2-9, and 2-10 illustrate different ways of representing the same data. Excel can read all these formats, with the exception of JSON, which can be imported by converting it to XML as an intermediate step. You also saw how to use Tables in Excel as flat files. Chapter 3 will demonstrate how to use Pivot Tables to extract summary data from large datasets. Chapter 4 will show you how to set up relational data models.

CHAPTER 3

Pivot Tables and Pivot Charts

This chapter covers extracting summary data from a single table, using Pivot Tables and Pivot Charts. The chapter also describes how to import data from the Azure Data Marketplace, using Pivot Tables. Chapter 4 will show how to extract summary data from multiple tables, using the Excel Data Model.

Recommended Pivot Tables in Excel 2013

Excel 2013 tries to anticipate your needs by recommending conditional formatting, charts, tools, and Pivot Tables. When data is highlighted or Ctrl+Q is pressed with any cell in the data range selected, a Quick Analysis Lens button is displayed at the lower right of the range. Clicking the lens will show the Analysis Tool window, which displays various conditional formatting options, as shown in Figure 3-1. Moving the cursor over a conditional formatting option will show a preview of that conditional formatting in the data. Clicking the Tables tab at the top right of the Analysis Tool window will result in several suggested Pivot Tables being shown. Hovering over each Pivot Table presents a preview, as shown in Figure 3-2.

X	🖯 🔈 🖓 🖓	÷		firstprivott	ablesept3 - I	Excel			TABL	E TOOLS			Ч.					? 🖻
F	LE HOME INSER	T PAGE LAY	OUT FORMU	LAS DATA	REVIEW	VIEW	POWERPIN	/OT Team	n Di	ESIGN								
-	K Cut	alibri	• 11 • A a	===	87 - 1	Wrap Text	C	Date	*				-		Ċ	AutoSum	* A	#
Pat	te S Format Dainter	3 I U - E	- A . A .		€ 1 2 E	Herge & Ce	nter -	\$ - % ,	***	Conditional	Format as	Cell	Insert	Delete For	rmat	Closes	Sort &	Find &
	· Pormat Painter									Formatting	* Table *	Styles *			*	Cieda -	Filter * S	Select *
	Clipboard 5	Por	nt i	91	Alignmer	nt	14	Number	5		Styles			Cells		E	siting	
A4	• : X	V fx	1/2/2014															
	and data and daulation	-		les la seles		1.00	1 12	1.1		1.0				0		0	D	
	1/2/2014 Eriday	anderson	Prive	24222 0000	• 0	n		,	~		IVI	14	-	0		ų	R.	5
6	1/3/2014 Privay	lonor	Corolla	19999 rod														
7	1/5/2014 Sunday	smith	Campy	24654 green														
8	1/6/2014 Monday	anderson	Prine	25464 silver														
0	1/6/2014 Monday	anderson	Camry	26312 white			-			-		-						
10	1/7/2014 Tuesday	iones	Corolla	19434 red														
11	1/8/2014 Wednesday	smith	Camry	26756 white														
12	1/9/2014 Thursday	anderson	Prius	26545 white								-						
13	1/9/2014 Thursday	smith	Corolla	19345 red														
14	1/10/2014 Friday	iones	Corolla	19434 green														
15	1/11/2014 Saturday	smith	Camry	27654 silver														
16	1/12/2014 Sunday	anderson	Prius	26543 red														
17	1/12/2014 Sunday	iones	Camry	27345 white														
18	1/12/2014 Sunday	smith	Corolla	18566 blue	1			-										
19	1/13/2014 Monday	jones	FORMATTIN	CHARTS	TOTALS	TARIES	SPARKLIN	JES										
20	1/14/2014 Tuesday	smith	, oranger the	a croatio	TOTALS	THULLS	ar cumpi											
21	1/15/2014 Wednesday	y and erson	858	F			Final A											
22	1/15/2014 Wednesday	jones					E.											
23	1/16/2014 Thursday	jones	Data Bars	Color Icon S	et Greate	r Top 10%	Clear											
24	1/17/2014 Friday	smith																
25	1/18/2014 Saturday	jones	Conditional E	armatting user sul	er to biobli	abt interaction	data											
26	1/19/2014 Sunday	smith	Conditional P	ormatting uses ru	es to rugnui	gin interesting	uata.											
27	1/20/2014 Monday	anderson	Camry	25463 green	, in													
2.0	a los loca e Tranders	1	Delive	acare athens	T Shills													

Figure 3-1. Analysis Tool window

E I	LE HOME INSEF	≂ RT PAGE LAY	OUT FORMU	firstprivottab JLAS DATA	lesept3 - Exc REVIEW	el VIEW	POWERP	NVOT Tear	TABU n DE	e tools Sign							? E - Neil
Pas	Cut	Calibri	- 11 - A' A	* = = *	-	Wrap Text		Date	• 0 .00	Conditiona	Format as	Cell	Insert Delete	Format	∑ AutoSum ↓ Fill +	- AT	Find &
-	💖 Format Painter	b I U + C	. <u></u>			vierge & Cer	iter -	3 . 30 .	100 -6.0	Formatting	 Table * 	Styles *	• •		2 Clear *	Filter * 1	ielect -
	Clipboard 12	For	nt	n.	Alignment		15	Number	5		Styles		Cells		Ec	siting	
A4	• : X	1 fr	1/2/2014														
		- JA		13 A.S. 18 A.													
4	sell date 👻 dayiofwe	salesperson	model - pr	rice 👻 color 👻	G	н	1	1	K	L	M	N	0	P	Q	R	S
5	1/3/2014 Friday	anderson	Prius	24323 green													
6	1/4/2014 Saturday	jones	Corolla	18998 red	-					_							
7	1/5/2014 Sunday	smith	Camry	24654 green													
8	1/6/2014 Monday	anderson	Prius	25464 silver													
9	1/6/2014 Monday	anderson	Camry	26312 white													
10	1/7/2014 Tuesday	jones	Corolla	19434 red													
11	1/8/2014 Wednesda	ay smith	Camry	26756 white	L,												
12	1/9/2014 Thursday	anderson	Row Label	Sum of price													
13	1/9/2014 Thursday	smith	Campy	A78905													
14	1/10/2014 Friday	jones	Carolla	470000													
15	1/11/2014 Saturday	smith	Drive	428007													
16	1/12/2014 Sunday	anderson	Crowd Tota	420307													
17	1/12/2014 Sunday	jones	Grand Tota	1292002													
18	1/12/2014 Sunday	smith	Corolla	18565 blue	Γ												
19	1/13/2014 Monday	jones	FORMATTIN	G CHARTS	TOTALS	TABLES	SPARKL	INES									
20	1/14/2014 Tuesday	smith															
21	1/15/2014 Wednesda	y anderson		[0.000]	[[wante]											
22	1/15/2014 Wednesda	y jones		<u>?</u>	?	.?											
23	1/16/2014 Thursday	Jones	Table F	votTable PivotTabl	e PivotTable	t More											
24	1/17/2014 Friday	smith															
25	1/18/2014 Saturday	jones															
26	1/19/2014 Sunday	smith	rables help y	ou son, niter, and su	mmarize dat												
27	1/20/2014 Monday	anderson	Camry	25463 green	. Č												

Figure 3-2. Pivot Table preview

Defining a Pivot Table

This section covers how to define a Pivot Table based on a simple transaction table of hypothetical Toyota sales data that lists transactions, including date of sale, model, salesperson, color, and selling price, as shown in Figure 3-3. Tables on which Pivot Tables are based should contain both numeric data and category fields, so that the numeric data can be sliced and diced by categories.

RI FI	LE HOME IN	SERT PAGE	LAYOUT FC	DRMULAS	DATA	REVIEW	VIEW	firstpri POWER (vottablejuly105 QUERY PO	dsx - Excel WERPIVO	I T							? 🗹	- 8 > Nel - 0	×
9	Cut	Calibri	- 11 -	A* A* =	= = •		Wrap Text		General					#	< 🚺	∑ AutoSun	· · AT	4		
Pas	te te	BIU-	B. D.	A - 1	==+	E 4E E	Merge & Cer	nter -	5 - % +	18 -18	Condition	al Format as	Cell	Insert Delet	e Format	e ci	Sort &	Find &		
	Pormat Painter			-							Formatting	r Table S	Styles *	• •	•	Clear *	Filter *	Select *		
	Clipboard 5		Font	19		Alignme	nt	- Fa	Number	- F9		Styles		Cells			Editing			^
K8	• = 7	$\times \checkmark f_x$,	v
	A B	с	D	E	F	G	н	1	J.	K	L	м	N	0	P	Q	R	s	T	*
1																				
2	sell date 💌 salespe	re 💌 model	▼ price	 color 	*															
3	1/1/2014 jones	Corolla	1943	32 white																
4	1/2/2014 smith	Camry	2678	silver																
5	1/3/2014 anderso	n Prius	2432	23 green																
6	1/4/2014 jones	Corolla	1899	98 red																
7	1/5/2014 smith	Camry	2465	54 green																
8	1/6/2014 anderso	n Prius	2546	54 silver																
9	1/6/2014 anderso	n Camry	2631	12 white																
10	1/7/2014 jones	Corolla	1943	34 red																
11	1/8/2014 smith	Camry	2675	56 white																
12	1/9/2014 anderso	n Prius	2654	15 white																
13	1/9/2014 smith	Corolla	1934	15 red																
14	1/10/2014 jones	Corolla	1943	34 green																
15	1/11/2014 smith	Camry	2765	54 silver																
16	1/12/2014 anderso	n Prius	2654	13 red																
17	1/12/2014 jones	Camry	2734	15 white																
18	1/12/2014 smith	Corolla	1856	56 blue																
19	1/13/2014 jones	Corolla	1976	55 white																
20	1/14/2014 smith	Camry	2675	59 white																
21	1/15/2014 anderso	n Prius	2456	54 siler																
22	1/15/2014 jones	Corolla	1932	28 green																
23	1/16/2014 jones	Corolla	1897	76 silver																
- 44	+ /+ = / = = + + + + + + + + + + + + + +	e	-		T. Harrison	1			-	1										٠
	Sheet1	5 Sheet16	Sheet17	Sheet18	Sheet19	Sheet	1 Sheet2	Sheet	3 Sheet9)	4							•	
REA	DY															III 🔟	D		+ 1009	16

Figure 3-3. Auto sales transaction table

Creating a Pivot Table doesn't change the underlying data; rather, it takes a snapshot of the data. When the underlying data is changed, the Pivot Table does not automatically refresh. To refresh, click anywhere within the Pivot Table to highlight the Pivot Table Tools Analyze tab and click Refresh.

Defining Questions

It is good practice to think through what questions you want answered by the Pivot Table. Some typical questions would be the following:

- What are the average sales per day?
- Who is the best performing salesperson?
- What days of the week have the most sales?
- What months have the highest sales?

Creating a Pivot Table

The steps to define a Pivot Table are as follows:

- 1. Make sure that the underlying data is in table format, with the first row showing the field names, and that the data is surrounded by blank rows and columns or starts in row 1 or column A. Press Ctrl+T to define the data, as a table as shown in Figure 3-3.
- 2. Click anywhere within the table and highlight the Insert tab. Click PivotTable. The screen shown in Figure 3-4 will appear. Note that the current table is selected and that by default, the Pivot Table will be placed in a new worksheet. Click OK.

🕼 🖯 🕹 - 🔅	□• •			firstpri	vottablejuly1	Oudse - Exe	el			TABLE	rools						? 🖽 -	264
HOME INSE InvotTable Recommended T ProtTables Tables	able Pictures	Online Pictures	Store	DATA pps - 20 pps	Recommen Charts	view ill - ded 💭 - Cha	POWER QU 量 · 索 · 益 · 前 · p 达 ·	PivotChart	Power View Reports	Map Nower Map	GN Line Col Spari	umn Win/ Loss dines	Slicer Tim Filters	eline Hype	erlink nks	π Equa Ω Syml	tion = bol	Nel * po
Table1 🔹 i 📉	$\checkmark f_x$	anderso	n															
sell date - salespers	- model	+ price	+ color	* F	G	н	1	J	K	L	M	N	0	P	Q	R	s	T
3 1/1/2014 jones	Corolla	19	432 white	1							-	-						
4 1/2/2014 smith	Camry	26	789 silver			100		Cri	eate Pivo	tTable	?	×						
5 1/3/2014 anderson	Prius	24	323 green			0	oose the data	that you w	ant to anab									
6 1/4/2014 jones	Corolla	18	998 red	1			Calact a tab	le or ranne										
7 1/5/2014 smith	Camry	24	654 green				Table/P	anger Tab	-			1						
8 1/6/2014 anderson	Prius	25	464 silver					ange. 180	ei .			1000						
9 1/6/2014 anderson	Camry	26	312 white				O give an exce	ernal uaca se	urce									
10 1/7/2014 jones	Corolla	19	434 red															
1 1/8/2014 smith	Camry	26	756 white				Connec	tion name:										
2 1/9/2014 anderson	Prius	26	545 white			G	hoose where y	ou want the	PivotTable	report to be	placed							
3 1/9/2014 smith	Corolla	19	345 red				New Works	sheet										
4 1/10/2014 jones	Corolla	19	434 green	1			O Existing Wo	orksheet										
5 1/11/2014 smith	Camry	27	654 silver				Location	n:				18						
6 1/12/2014 anderson	Prius	26	543 red			C	noose whether	r you want t	o analyze m	ultiple table	5							
7 1/12/2014 jones	Camry	27	345 white				Add this da	ta to the Da	ta Model									
8 1/12/2014 smith	Corolla	18	566 blue							or		ncel						
9 1/13/2014 Jones	Corolla	19	765 white	1						- Ch	-	incer .						
0 1/14/2014 smith	Camry	26	759 white					-										
1/15/2014 anderson	Prius	24	564 siler															
22 1/15/2014 jones	Corolla	19	328 green															
1/16/2014 jones	Corolla	18	976 silver															
4 1/17/2014 smith	Camry	26	543 red															
1/18/2014 jones	Prius	24	324 white															
a landance	daralla.		A		1			1.000	-									
Sheet15	Sheet16	Sheet17	Sheet18	Sheet19	Sheet1	Sheet2	Sheet3	Sheet9			4							Þ

Figure 3-4. Create PivotTable dialog

3. At this point, you have to think through what type of data you want to extract. We will start with a summary of sales per day, with a breakout of models sold per day. We will do this by dragging sell date to the Columns box, model to Rows, and price to Values, as shown in Figure 3-5. Note that the default display is Sum of price.

FLE	HOME	.⇒ - 🛄 ∓ INSERT PAGE	LAYOUT	FORMU	firstpri	vottablesep A REV	it3 - Excel 1EW VI	EW POV	VERPIVOT	Team	PIVOTI ANALYZ	ABLE TOOLS	4					?	00 - 0 Neil -
PivotT PivotT Og Ph	able Name: A Fable2 ptions * I rotTable	Active Field: Sum of price Dril Driv Driv Dow Active Field	✓ ↑ 1 II Drill -9 In Up -	i → Gr ea Ur I Gr	roup Selectio ngroup roup Field Group	n Insert Slicer	Insert Timeline O Filter	Filter Connections	Refresh	Change Data Source * Data	Clear	* t * : PivotTable tions	民 Fields, 前, OLAP 1 回日 Relatio Cal	Items, & Sets - Tools - inships culations	PivotCh	hart Rec Pi Tool	commended ivotTables	Field +/- List Button Show	Field Field Headers
A3	٠	$: \times \checkmark f_x$	Sum of	price															
1	A	В	с	D	E	F	G	н	1	J	К	L	м	N	0	^	PivotTa	ble Field	ds •
3 Si	um of price ow Labels 👻	Column Labels 💌 1/1/2014	1/2/2014 1/	3/2014	1/4/2014 1	/5/2014 :	1/6/2014 :	1/7/2014 1	/8/2014	1/9/2014 1/	10/2014 1	/11/2014 1	/12/2014	1/13/2014 1,	/14/2014	1/1	Choose fields	to add to rep	port:
5 C	amry orolla rius	19432	26789	24323	18998	24654	26312 25464	19434	26756	19345 26545	19434	27654	27345 18566 26543	19765	26759		dayiofwe	ek on	
6 G	rand Total	19432	26789	24323	18998	24654	51776	19434	26756	45890	19434	27654	72454	19765	26759		price color MORE TABLE	5	
3 4 5																	Drag fields b	etween areas	below:
5 7 8																	T FILTERS	se	ell date •
9																	T ROWS	Σ	VALUES
1																	model	• Si	im of price
3	•	Sheető Sheet7	Sheet8	Sheet10	Sheet12	Sheet	14 She	et11	(+)	4							Defer Lay	out Update	UPDAT

Figure 3-5. Basic Pivot Table showing count of car sales by model

4. To change Sum of price to Count of price, right-click cell A3 or pull down the arrow on Count of price under Values and select "Summarize Values by" and count. The result is shown in Figure 3-6.

otTable2 Cou Options • PrvotTable	ve Field: unt of price Field Settings Active Field	nill −∃ Up+	学 Gro 创 Ung ⑦ Gro	up Selection proup up Field Group	Insert Slicer	Insert Timeline Filter	Filter Connections	Refresh	Change Data Source * Data	Clear Select	• • PivotTable tions	時た Fields, 前た OLAP 中居 Relatio Cal	Items, & Sets * Tools * anships culations	PivotChi	art Rec Pi Tool	ommended Field List But s	+/- Field ttons Headers Show
3 * :	$\times \checkmark f_x$	Count of	price														
A	В	c	D	E	F	G	н	1	J	К	L	м	N	0	-	PivotTable Fi	ields -
Count of price	olumn Labels 💌	to boost a t		. labora a	In local a	e le loos e		In loss a		leabors -				a trace of	. /.	Choose fields to add t	o report:
Camry Corolla	1/1/2014 1	1	1	1	1	1/6/2014	1/1/2014 1	1	1/9/2014 1/	1	1	1/12/2014	1/13/2014 1/	14/2014	1/1	sell date dayiofweek salesperson model	
Grand Total	1	1	1	1	1	2	1	1	2	1	1	3	1	1		Color MORE TABLES	
																Drag fields between a	reas below:
																T FILTERS	sell date
																= ROWS	Σ VALUES
																model -	Count of price

Figure 3-6. Pivot Table showing value of car sales by model

Changing the Pivot Table

It is very easy to make changes to the Pivot Table. To change the horizontal axis from dates to salesperson, simply click sell date from the Columns box back and select "remove field." Then drag salesperson from the Fields list to the Columns box. The result is shown in Figure 3-7.

X	ILE HOME	insert	₽ PAG	E LAYOU	T F	firstp ORMULAS D4	rivottablesept ATA REVI	3 - Excel EW VIE	EW POV	NERPIVO	T Team	PIVOT ANALYZ	TABLE TOOLS DESIG	N						? 🗹 – 🖗 Neil •
Pive	tTable Name: A otTable2 Options - PivotTable	Active Field: Sum of price	tings Do Active Fie	ill Dril wn Up		→ Group Selecti I Ungroup C Group Field Group	on Insert Slicer	Insert Timeline C Filter	Filter	Refres	h Change Data Source * Data	Clear	t • e PivotTable ctions	民 Field 版 OLA 电子 Rela	ds, Items, & Sets • P Tools • tionships Calculations	PivotChart	Reco Pin Tools	ommended votTables	Field List B	+/- Field uttons Headers Show
A		: ×	V 1	x Su	m of pr	ice														
1	A	В		с	D	E	F	G	н	1	1	к	L	м	N	0	Ê	PivotT	able F	ields 🔹
2 3 4	Sum of price	Column L	abels 💌	iones	smith	Grand Total												Choose fie	lds to add	o report:
5	Camry Corolla	Juniterioun	184657 38411	134993	15915	5 478805 9 384350												☐ sell dat ☐ dayiof	e week erson	
8	Prius Grand Total		127439 350507	50089 453802	25137 48775	9 428907 3 1292062												✓ model ✓ price		
10																		MORE TAR	ILES	
12 13																		D		
14																		Drag field	between	reas below:
16																		T FILTE	85	III COLUMNS
17 18																				salesperson
19																		E ROWS	3	Σ VALUES
20 21 22																		model	•	Sum of price •
23																	¥			and the second
	< >	Sheet6	Sheet7	Sheet	8 Sh	eet10 Sheet	2 Sheet	4 She	et11		1					Þ]	Defer l	ayout Upd	ate UPDAT

Figure 3-7. Pivot Table showing sales by model and salesperson

Creating a Breakdown of Sales by Salesperson for Each Day

The Pivot Table can be modified to show the sales by salesperson and model for each day. Simply drag the model field to the Columns box and salesperson and sell date fields to the Rows box, so that both sell date and salesperson are shown in the Rows box. The results are shown in Figure 3-8.

F	E 5 · C ² LE HOME	- 🛄 = INSERT PAGE	LAYOUT	F	firstpriv ORMULAS DAT	vottablese A RE	pt3 - Exce VIEW	I VIEW POV	VERPIVOT	Team	PIVOTTA ANALYZE	BLE TOOLS DESIG	N	1.1					Neil -
Pivo	tTable Name: Act tTable2 Su Options - RevotTable	tive Field: m of price Field Settings Active Fiel	n Up -	ы њ	→ Group Selection Ungroup Group Field Group	Inset	Insert Timelin Filt	Filter Filter ter	Refresh	Change Data Source * Data	Clear * E Select * Move P Acti	wotTable	民 Field 版 OLA 尼 Relat	ls, Items, & Sets = P Tools = tionships alculations	PivotChart	Reco Pir Tools	ommended votTables	Field List Bu	+/- Field ttons Headers thow
A3	•	$\times \checkmark f_x$	Sun	n of pr	ice														
4	A	В	с	D	E	F	G	н	1	J	к	L	м	N	0		Dista	his r	
1 2 3 4	Sum of price	Column Labels 💌 Camry	Corolla	Prius	Grand Total												Choose field	dDIE Fi	o report:
5	anderson	184657	38411	12743	39 350507												dayiofw	eek	
6	1/3/2014			2432	23 24323												Salesper	rson	
7	1/6/2014	26312		2546	54 51776												✓ model		
8	1/9/2014			2004	20343												✓ price		
10	1/15/2014			2004	54 24564														
11	1/20/2014	25463		2.404	25463												MORE TABL	.ES	
12	2/2/2014		19435		19435														
13	2/7/2014		18976		18976														
14	2/14/2014	25662			25662												Drag fields	between a	reas below:
15	2/17/2014	26543			26543												T FILTER	5	III COLUMNS
16	2/20/2014	26432			26432														model
17	2/23/2014	27456			27456														
18	2/26/2014	26789	200320	5004	26789														
19	= jones	134993	208720	5008	453802												= ROWS		Σ VALUES
21	1/4/2014		19432		19432												salesperso	n •	Sum of price
22	1/7/2014		19434		19434												sell date	-	
23	1/10/2014		19434		19434														
	. lastan.	34.650		1	34.000	1			-								Defer La	wout Und	te UPD

Figure 3-8. Pivot Table showing breakdown by salesperson and model for each day

Showing Sales by Month

To display a sales report by month, right-click any cell in the sell date column and select Months from the Grouping window, as shown in Figure 3-9, and click OK.

Gr	ouping	?	×
Auto			
✓ <u>Starting</u> at:	1/1/2014		
✓ Ending at:	2/28/2014		
Ву			
Seconds Minutes Hours Days			^
Months Quarters Years			~
Nun	ber of day	s: 1	×
	Ж	Can	cel

Figure 3-9. Dialog to select grouping by month

The resulting Pivot Table is shown in Figure 3-10.

K Cut Copy • Format P. Clipboard	ainter 5	- 1 Font	1 • A	∧* = = A • = = 5	■ ※ ・ = (Ξ + Al	₩ E I N gnment	lrap Text lerge & Cente	5	Date \$ - % * Number	***************************************	Conditiona Formatting	al Format as • Table • Styles	Cell Styles *	Insert Dele	te Format	N N	AutoSum * Acar Z Fill * Z Clear * Filter * S Editing	ind & elect *
A	B B	C	D	E	F	G	н	1	J	к	L	М		N C	>		PivotTable F	ields
um of price	Column Labels 💌																Choose fields to add t	to report:
ow Labels 👻	Camry	Corolla	Prius	Grand Total													Sell date	
anderson	184057	38411	12/439	350507													dayiofweek	
Jan	51//5	20411	12/439	179214													✓ salesperson	
reo	132882	30411	50000	1/1293													i model	
lan	55241	135367	50089	453802													i price	
Feb	79752	133353	50005	213105														
mith	159155	77219	251379	487753													MORE TABLES	
Jan	159155	77219	25431	261805														
Feb			225948	225948														
and Total	478805	384350	428907	1292062													Drag fields between a	reas below:
																	T FILTERS	II COLUN
																		model
																	I ROWS	Σ VALUE
																	salesperson 🔻	Sum of pr
																	cell data 🖉	

Figure 3-10. Pivot Table showing sales by salesperson and model by month

Creating a Pivot Chart

To create a Pivot Chart from a Pivot Table, with the Pivot Table displayed on the screen, do the following:

- 1. Go to the Pivot Table Tools Analyze tab and select Pivot Chart.
- 2. From the Insert Chart pop-up shown in Figure 3-11, select the type of chart desired and click OK. For this example, we will use a stacked column chart.

L	Insert Chart	? ×
All Charts		
Recent Templates		
Column Image: Column	Stacked Column	
	ОК	Cancel

Figure 3-11. Insert Chart dialog

3. The result will be displayed as shown in Figure 3-12.

X F	LE HOME	INSERT PAG	E LAYOUT	FO	fin RMULAS	itprivottables DATA R	sept3 - Exce IEVIEW	NEW P	OWERPIVO)T Tear	n ANALYZ	PIVOTCHA E DES	RT TOOLS	ORMAT				0	? 🗷 -	- 5 Neil - F
Chi M	rt Area Format Selection Recet to Match Style		Change	Ab	c Abc	Abc	 Sha Sha Sha Sha Sha 	pe Fill * pe Outline * pe Effects *	Α	Α	A	A - 1	Bring For Send Bac Selection	ward - 🗭 A kward - 🕅 G Pane A B	l lign - iroup -	\$]] 3*		:		
-	Current Selection	Insert Sh	apes			Shape Styles		6		WordArt	Styles	5		Arrange		Size		G		
Ch	art 1 👻 i	$\times \checkmark f$	e																	
4	A	В	с	D	E	F	G	н	I.	J	К	ι	м	N	0		-	PivotChart E	olds	*
2																		FIVORCHAILT	leius	
3	Sum of price Co Row Labels 👻 Ca	lumn Labels 💌 mry	Corolla	Prius	Grand Tota													Choose fields to add t	o report:	8
5	anderson	184657	38411	127439	35050	7												daviofweek		
6	Jan	51775		12743	Sum of price							+						☑ salesperson		
7	Feb	132882	38411		100000	5												✓ model		
8	∋jones	134993	268720	50085	300000							1						☑ price		
9	Jan	55241	135367	50085	250000													color		
10	Feb	79752	133353		200000		_													
11	∃ smith	159155	77219	25137							model *	Ľ								
12	Jan	159155	77219	2543:	150000						Prius									
13	Feb			22594	100000						Corolla									
14	Grand Total	478805	384350	42890	50000						Camry							Drag fields between a	reas below:	
15																		T FILTERS	III LEGEN	ID (SERIE
16					0													T THETENS	an coort	en factor
17						Jan re	60 10	in rep	Jan	1.60									model	
18						anderson		jones		smith										
19					salesperson	▼ sell date	- *											= AVIS (CATEG	T VALUE	
20																			Sumafin	den .
21																		salesperson •	aum or p	nue •
22																		Sei Dare,		
23																	-			

Figure 3-12. Pivot Chart with Pivot Table

Adjusting Subtotals and Grand Totals

The subtotals can be adjusted on the Pivot Table Tools Design tab. Click Subtotals to see the options.

- Do not show subtotals
- Show all subtotals at bottom of group
- Show all subtotals at top of group

Click Grand Totals to see the following options:

- Off for rows and columns
- On for rows and columns
- On for rows only
- On for columns only

Analyzing Sales by Day of Week

It can be useful to analyze sales by day of the week. In this example, we have added a column for day of week, using the following steps:

- 1. Insert a column to the left of salesperson. Name the new column "day of week."
- 2. Enter the following formula in cell B3, to extract the day of week:

=TEXT(A3,"dddd")

3. The formula will be automatically copied down the column, because this is a table. The result is as shown in Figure 3-13.

X) Fi	E HOME INSER	♥ T PAGE LAY	OUT FO	firstprive RMULAS DATA	ttablesept REVII	3 - Excel EW VI	EW POW	VERPIVOT	Team	TABLE	TOOLS							2 00 -
Pas	te Copy • Clipboard G	3 I <u>U</u> - ⊞ Fon	- 11 - 1 1 - <u>@</u> - t		参・ 住在 Align	Wrap Merg ment) Text je & Center	- Date - \$ -	% + +	- 8 48 14	Conditional Formatting *	Format as Table ~ S Styles	Cell Styles -	nsert Delete	Format	∑ AutoSum ↓ Fill - ℓ Clear - E	Sort & F Filter - S	ind & elect *
PR	oper 🔹 : 🗙	$\checkmark f_x$	=TEXT(A3,"	'dddd")														
	A B	с	D	E		G	н	а II.	J	к	ι	м	N	0	Р	Q	R	s
1			-	-														
2	sell date * dayiofw(*	salesperson *	model	price color														
4	1/2/2014 T TEXT(value	format text)	Campy	26789 silve	-													
5	1/3/2014 Friday	anderson	Prius	24323 gree	2													
6	1/4/2014 Saturday	iones	Corolla	18998 red														
7	1/5/2014 Sunday	smith	Camry	24654 gree	n													
8	1/6/2014 Monday	anderson	Prius	25464 silve	r													
9	1/6/2014 Monday	anderson	Camry	26312 whit	2													
10	1/7/2014 Tuesday	jones	Corolla	19434 red														
11	1/8/2014 Wednesday	smith	Camry	26756 whit	2													
12	1/9/2014 Thursday	anderson	Prius	26545 whit	e													
13	1/9/2014 Thursday	smith	Corolla	19345 red														
14	1/10/2014 Friday	jones	Corolla	19434 gree	n													
15	1/11/2014 Saturday	smith	Camry	27654 silve	r													
16	1/12/2014 Sunday	anderson	Prius	26543 red	1													
17	1/12/2014 Sunday	jones	Camry	27345 whit	2													
18	1/12/2014 Sunday	smith	Corolla	18566 blue														
19	1/13/2014 Monday	jones	Corolla	19765 whit	9													
20	1/14/2014 Tuesday	smith	Camry	26759 whit	5													
21	1/15/2014 Wednesday	anderson	Prius	24564 siler														
22	1/15/2014 Wednesday	jones	Corolla	19328 gree	n													
23	1/16/2014 Thursday	jones	Corolla	18976 silve	r													
	• Sheet24	Sheet25 S	Sheet26	Sheet27 Sheet	28 Sh	eet29	Sheet30	Sheet31	Sheet1			1						

Figure 3-13. Extracting day of week

- 4. Save the spreadsheet.
- **5.** On the Insert tab, click PivotTable. Click OK in the Create PivotTable pop-up window to accept the default.
- **6.** If necessary, click Refresh on the Pivot Table Analyze tab to display the new column.
- 7. On the PivotTable Fields tab, drag model to the Columns box, day of week to Rows, and Sum of price to Values to get the Pivot Table shown in Figure 3-14.

FILE F	• ¢• 🔳	≠ RT PAG	E LAYOUT	r FG	firstp ORMULAS DA	rivottabl ATA	esept3 - Excel REVIEW	VIEW POV	VERPIVOT	Team	PWOTT ANALYZE	ABLE TOOLS	N	9 Y U					? 00 - 0 Neil -
PivotTable Nar PivotTable1	Active Fiel Sum of pr	ettings Dor Active Fie	ill Drill wn Up - Id	$d_{\rm H}=d_{\rm H}$	→ Group Selecti I Ungroup C Group Field Group	ion In St	sert Insert icer Timeline Filte	Filter Connections	Refresh	Change Data Source * Data	Clear	• • PivotTable tions	院 Field 示 OLAI C	s, Items, & Sets + ? Tools + ionships alculations	PivotChart	Reco Pit	ommended votTables	Field List Bu	+/- Field attons Headers Show
A3	• : X	√ f.	sur	n of pr	ice														
A A 1 2 .		B	c	D	E	F	G	н	1	J	к	L	м	N	0	-	PivotT Choose fiel	able F	ields 🔹
4 Row Labe 5 Sunday 6 Monday 7 Tuesday 8 Wednese 9 Thursday 10 Friday 11 Saturday 12 Grand To	day tal	79455 106183 50211 53545 81117 52205 56089 478805	Corolla 77111 39639 37331 96631 57276 38410 37952 384350	Prius 5308 4892 7684 4799 7765 7518 4922 42890	Grand Total 6 209652 194742 194742 10 164382 16 198172 17 216050 16 165801 12 143263 17 1292062												sell dati dayiofv salesper model price color MORE TAB	t veek rson LES	
14																	Drag fields	between a	rreas below:
16 17 18																	T FILTER	5	model
19																	ROWS		Σ VALUES
21 22																	dayiofwe	ek 🔻	Sum of price
23	Sheet24	Sheet25	Shee	et26	Sheet27 Sh	eet28	Sheet29	Sheet30	. 🕀	: 4					•	¥	🗌 Defer L	ayout Upd	ate UPDAT

Figure 3-14. Pivot Table of sales by day of week and model

Creating a Pivot Chart of Sales by Day of Week

To get a visual representation of sales by day of the week, create a Pivot Chart from a Pivot Table. With the Pivot Table displayed on the screen, do the following:

- 1. Go to the Pivot Table Tools Analyze tab and select Pivot Chart.
- 2. From the Insert Chart pop-up shown in Figure 3-15, select the type of chart desired.

	Insert Chart	? ×
All Charts		
All Charts		
	ОК	Cancel

Figure 3-15. Insert Chart dialog

3. If a clustered column chart is selected, the results will be as shown in Figure 3-16.

Thart Area Format Selection	Ae Insert Sh	Change Shape *	Abc Abc	Abc ane Styles	 Shape Shape Shape Shape Shape 	e Fill * e Outline * e Effects *	Α	A		A - 151 A - 151 A - 151	Bring Forwa Send Backw Selection Pa	rd - F A ard - F G ne A R	dign • iroup • otate •	3" 5" Size	:	
Chart 1 *	$\times \sqrt{f}$	r														
A 1 2 3 3 3 3 3 3 4 4 8 5 5 5 5 1 5 4 5 5 5 5 5 1 5 5 5 5 5 5 5 5 5 5 5 5 5	8 Column Labels ▼ Camry 79455 106183 50211 53545 81117 52205 56089	Corolla Pri 77111 5 39639 4 37331 7 96631 4 57276 7 38410 7 37952 4	D E us Grand Total 3086 209652 892(50000 518(80000 518(80000 518)	F	G	н	1		K model •	ι +	M	N	0	F A	PivotChart F Choose fields to add sell date dayiofweek salesperson model price color	to report:
2 Grand Total 3 4 5 5 7 7 8 9 9 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	478805	384350 42	890: 40000 - 20000 - 0 - ga ^{gebh}	Harder	Transfer Street	and the second	A robat	sauton	Corolla						Drag fields between T FILTERS AXIS (CATEG dayiofweek •	areas below: III LEGEND (SERI model

Figure 3-16. Pivot Chart for sales by day of week

Using Slicers

Slicers can be used to filter data and to zoom in on specific field values in the Pivot Table. To set up a slicer, do the following:

- 1. Go to the Pivot Table Tools Analyze tab and select Insert Slicer.
- 2. From the Insert Slicers pop-up window, select the fields that you want to slice, such as salesperson and date, and click OK.
- **3.** The result is shown in Figure 3-17, which displays the sales by salesperson for each model and date...A specific salesperson can be selected, and the slicer menu can be selected to view sales for that salesperson.

E) Fi	E HOME	INSERT PAG	SE LAYO	UT FC	ORMULAS	firstpriv DAT/	ottable	sept3 REVIEV	- Excel W VIEW	/ PO	WERPIVO	т	Te	sam	ICER TOOL	S		li d					? 0	- 8 Neil •
Slice sale	r Caption: sperson Slicer Settings Slicer	Report				alicer Styles					Bring] g rd - B	Se	end Se ward -	lection Ali Pane	ign Group	Rotate	Colur	mns: 1 nt: 0.26" n: 1.81" Buttons	:	Height: 2.7 Width: 2* Size	6" () () ()		
sal	esperson *	XVJ	fx																					
	A	В		С	D	E	F		G	н	1	J	J	K	L	М	N	0	Р	Q	R	s	т	U
1																								
2																								
3	Sum of price	Column Lab	els 👻													Eab								
5	Row Labels	2-Jan		5-Jan	8-Jan	9-Jan	11-	sale	esperson		~	×	In	22-Jar	24-Jan	4-Feb	6-Feb	9-Feb	12-Feb	15-Feb	18-Feb	21-Feb	24-Feb	27-Feb
6	Camry	1	26789	24654	26756		276	an	nderson			Ť.												
8	Corolla					19345		jo	nes				10	1943	25431	26543	25678	26543	23432	24898	24532	25432	23456	25434
9	Grand Total		26789	24654	26756	19345	276	(en	nith			1	76	1943	2 25431	26543	25678	26543	23432	24898	24532	25432	23456	25434
10								-				-												
11													9											
12																								
13																								
14													E											
15													E											
16																								
17																								
18																								
19																								

Figure 3-17. Filtering results using salesperson slicer

Adding a Time Line

A time line can be added to slice by time intervals. To add a time line, do the following:

- 1. From the Pivot Table Tool Analyze tab, select Insert Timeline. Select sell date in the Insert Timeline window and click OK. The time period for the time line can be selected by clicking the down arrow at the upper right of the Timeline window to select years, quarters, months, or days.
- 2. As seen in Figure 3-18, Timeline and Slicer work together. A specific salesperson and date can be selected to view sales by that salesperson on that date.

X		⇒ - 🛄 ∓ INSERT PAI	SE LAYOUT	FORMU	firstpr LAS DA	ivottablesep TA REV	pt3 - Exce /IEW	VIEW	POWE	RPIVOT	Tean	TIM	ELINE TO	DOLS								? @	I - E
Pa	Ste Sormat F	Painter B I U	- - - ⊞- ⊉	A A	= = ;	= 秒 • ≡ €≣ +3	■ ■ M	/rap Text lerge & C	enter ~	Gener \$ •	al 96 9	+.0 .00 -00 +0	Conc	ditional satting =	Format a Table -	s Cell Styles +	e E Inser ∙	t Delete	Format	∑ Auto ↓ Fill - ℓ Clear	Sum =	Arr Arr Sort & Find & Filter * Select *	
	Clipboard	6	Font		*	Alig	priment		9		Number	-0			Styles			Cells			Edit	ing	
54	ell date *	XVJ	Îx																				
1	A	В	c	D	E	F G	н	1	J	K	L	М	N	0	P	Q	R	S	T	U	V	W	х
1 2																							
3	Sum of price	Column Labels	r)																				
4	Row Labels *	∋Jan 10-Jan	Grand Total			salespe	rson	-	*														
6	Corolla	19434	19434			ander	500											100					
7	Grand Total	19434	1 19434			anuer	5011				sell d	ate						*					
8						jones					Jan 10	,2014					0/	IYS -					
9						smith					JAN 20	14											
10						-					7 8	8 9	10	11 12	13	14 15	16 1	17 :					
12						-			_														
13																		•					
14						1																	
15																							
16																							
17							_		_														
18																							
19																							
20																							

Figure 3-18. Filtering data with a time line plus salesperson slicer

Importing Pivot Table Data from the Azure Marketplace

Data can be imported into a Pivot Table from the Azure Marketplace, which includes a wide variety of data, such as demographic, environmental, financial, retail, and sports. Some of the data is free, and some is not. The following example shows how to access free data.

1. Go to the Azure Marketplace web site at datamarket.azure.com/browse/data. The screen shown in Figure 3-19 will appear as of August 2015.



Figure 3-19. Azure data market browse screen

- 2. The first time you use the site, you may be asked to fill in account information and to create a free account ID. If you already have a free Microsoft account, such as a Hotmail.com or Outlook.com account, you can use that.
- 3. Click Data at the top of the screen. The screen shown in Figure 3-20 will appear.



Figure 3-20. Azure Marketplace

- 4. Scroll through the data sources. You may have to go through an additional signup process to use free data.
- **5.** For this example, we will use Real GDP Per Capita By State (1997–2011), as shown in Figure 3-21. If prompted, enter your Microsoft ID and agree to the terms of use.



Figure 3-21. Real GDP all industries per state database

- 6. Click Explore this Dataset in the middle of the screen.
- 7. The data will be displayed, as shown in Figure 3-22. Note, in the upper-right corner of the screen, the Download Options of Excel (CSV), PowerPivot 2010, and PowerPivot 2013. Select PowerPivot 2013.

le Ed	it <u>V</u> iew H Real GDP Per	igtory <u>B</u> ookmi Capita By Sta	arks <u>I</u> ools	Help Service Explo	rer Microso	ft ×	+											-	0
•	https://da	tamarket azure .	com/datase	t/explore/d11	a9c01-19a1	-4196-985	8-a664a95a	4347				Ø ∀ C	Q, Search		☆自	۵	÷	ŵ	\$
	Ś		Real GDI Real GDP I Primary Ad	P Per Capita Per Capita B ccount Key S	ta By Stat y State (19 how	te (1997 97-2011)	-2011)							Download Options Excel (CSV) PowerPivot 2010 PowerPivot 2013	 Micro	soft A	zure 1	Marke	tp
JRL fo http Displa ID	ving 100 of 9	pressed query: atamarket.az 200 rows Page AREA * United States	1 > VEAR -	BEA/RealGD	PPerCapit rCapital 💌	aByState	/v1/Real	GDPSyStat	ePerCapit	a 1									
2	0	United States	1998	37251															
3	0	United States	1999	38587															
4	0	United States	2000	39775															
5	0	United States	2001	39879															
6	0	United States	2002	40192															
7	0	United States	2003	40701															
8	0	United States	2004	41709															
9	0	United States	2005	42483															
10	0	United States	2006	43220															
11	0	United States	2007	43633															
12	0	United States	2008	43079															
13	0	United States	2009	41640															
14	0	United States	2010	41764															
15	0	United States	2011	42070															
16	1000	Alabama	1997	27996															
17	1000	Alabama	1998	28642		_	-		1	I III III IIII	_					24.7	1.00	- 40	
	P			1	hn			XI	w	Ju.							18 .	d 40	

Figure 3-22. Display of real GDP per capita records

- 8. You will be prompted to save the file. Click OK. The file will appear in the download section of your web browser. What you will see now depends on which web browser you are using. To save the file and open in Excel, click whatever prompts appear. If using Firefox, double-click the file name to start the process. If you are prompted to select the program to use to open the file, choose Excel.
- **9.** A pop-up window will appear, as shown in Figure 3-23, asking for how you want to view the data. Select Pivot Table Report, click OK, and wait for the data to download.



Figure 3-23. Import Data dialog

- **10.** If a pop-up appears asking for your account key, enter your Microsoft username (e.g., a Hotmail or Outlook.com e-mail address).
- 11. Start designing the Pivot Table. Drag year to the Columns box, drag area to the Row box, and Real GDP to Values, as shown in Figure 3-24.

Image: Solution of the	FORMULAS DAT	Book1 - Excel A REVIEW VIEV	W POWERPIVOT T	PIVOTTABLE TOOLS earn ANALYZE DESIG	, ilili vi		? • - 8 Neil - C
Cut Calibri - 11	- A' A' = = =	- ₩rap	Text General	· 🔣 🛛	B 🗩 🗒	- 🖹 🏹 🛛	AutoSum * A
Paste Copy *			8. C. N.	€ 0 00 Conditional For	mat as Cell Inse	ert Delete Format	Fill * Sort & Find &
- S Format Painter	·· A · = = -	= th th Emiliar Merge	e ol Center * \$ * %	Formatting * Ta	able * Styles * *	4	Clear * Filter - Select *
Clipboard 🕫 Font	r <u>a</u>	Alignment	rs Num	ier 🕫 Styli	es	Cells	Editing
A1 * : $\times \checkmark f_X$ Sum of	RealGDPPerCapital						
A B	C D	E F	G H I	J K L	M N	0 P -	
1 Sum of RealGDPPerCapital Column Labels							PivotTable Fields * 3
2 Row Labels 💌 199	7 1998 1999	2000 2001	2002 2003 200	4 2005 2006 200	7 2008 200	2010 20:	ACTIVE ALL
3 Alabama 2799	5 28642 29518	29792 29955	30584 31079 3245	2 33064 33240 3334	9 32995 3174	5 31655 3130	Change Ealth In addite second
4 Alaska 6058	9 56896 55892	54362 56322	57692 55949 5795	5 56713 59044 5990	9 59057 6325	2 61040 6185	Choose fields to add to report.
5 Arizona 2991	7 31745 33505	34736 35267	35282 36315 3672	5 38186 39643 3970	8 38305 3575	4 34874 350	A RealGDPByStatePerCapital
6 Arkansas 2740	5 27569 28784	28845 29058	29674 30360 3119	4 31724 32107 3201	4 31701 3097	31211 3114	OI D
7 California 3656	7 38453 40832	43262 42702	43066 43936 4559	4 47225 48522 4878	9 48259 4602	45581 4604	FIPS
8 Colorado 3920	2 40576 42545	45101 45590	45525 45298 4577	8 46938 47261 4759	9 47659 4662	45551 4575	AREA
9 Connecticut 4916	2 50316 51073	54302 54370	53094 53055 5565	9 56190 58048 5961	3 58874 5755	55223 5624	VEAR
10 Delaware 5625	4 56797 57793	59598 61663	59318 61751 6356	8 64780 64319 6547	6 61460 6220	7 62679 6315	RealGDPPerCapital
11 District of Columbia 11616	9 117449 122148	122096 128456	132630 136964 14253	5 146063 147172 14916	8 152167 14793	9 148605 14825	
12 Far West 3706	3 38707 40907	42844 42185	42535 43293 4485	7 46400 47747 4824	0 47773 4560	45479 459	
13 Florida 3142	0 32298 33313	34198 34414	34912 35848 3680	5 38128 38895 3887	5 37237 3560	3 34913 3468	
14 Georgia 3652	0 37634 39329	40076 39875	39705 39793 4022	9 40685 40428 4050	5 39503 3712	37025 372	Drag fields between areas below:
15 Great Lakes 3606	3 37069 37996	38894 38333	38910 39497 4003	2 40219 40296 4061	7 39786 3825	3 38254 387.	T FILTERS III COLUMNS
16 Hawaii 3970	1 38778 39148	40143 39385	39966 41139 4254	2 43992 44992 4534	8 45163 4351	5 42620 4211	VEAR .
17 Idaho 2520	0 25806 27897	30332 29768	30028 30267 3213	8 34080 33680 3422	4 33553 3198	32559 324(1001
18 Illinois 4002	8 40963 41961	43198 43145	43171 43941 4490	5 45166 46105 4659	3 45970 4456	5 44730 452	
19 Indiana 3325	7 34697 35395	36427 35610	36441 37549 3830	5 38157 38252 3902	4 38082 3634	3 36699 369	E ROWS E VALUES
20 Iowa 3344	4 33595 34288	35946 35186	36151 37420 4003	7 40567 40698 4243	9 41451 4083	41365 4195	APEA T
21 Kansas 3359	5 34593 35344	36347 36717	36908 37734 3764	5 38307 39390 4095	8 41059 3944	39438 3948	Sum of Resid *
22 Kentucky 3125	7 31856 32588	31690 31580	32239 32412 3277	7 33312 33777 3341	9 33293 3244	5 32319 323	
23 Louisiana 3736	38855 39087	37573 38315	38625 40154 4188	7 43080 44823 4236	4 41197 4233	45061 4500	
Real GDP Per Capita By State (1	+		i (3 AFC3 AAA7C AAAA	0 34640 3304	• • • • • • • • • • • • • • • • • • •	Defer Layout Update UPDATE

Figure 3-24. Real GDP data imported into Excel

12. To use a slicer to filter the data, click Insert Slicer on the Pivot Table Tools Analyze tab. Select the Area field and click OK. On the Slicer menu, select a state, such as California, to view its real GDP per capita, as shown in Figure 3-25.

X	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	insert	PAGE LAYOUT	FORML	ILAS	DATA	REV	IEW	VIEW	PO	AzureRea WWERPIV	IGDP do	wnload Team	- Excel					N.					? •
Pa	A Cut Copy - aste Format Painte Clipboard	Calibr	ni • 11 • 7 U • ⊞ • ⊘ Font] A* A - <u>A</u>		H H	8⁄- €≣ 4≣ Alig	nment	Vrap Te Merge &	xt : Center		eneral 5 - % Num	> 2	•	Conditi Formati	ional Fo	ormat as Table =	Cell Styles -	Inser	t Delete	Format	∑ AutoSum ↓ Fill + ≪ Clear + Ed	Sort & Filter	Find & Select •
a	2 * :	Xv	fx																					
	A		B	с	D	E	F	G	н	1	1	к	L	м	N	0	Р	Q		R	s	Formula	lar U	v
1	Sum of RealGDPPe	erCapital	Column Labels •	1998	1999	2000	2001	2002	2003	2004	1 2005	2006	2007	2008	2009	2010	2011							
3	California	100	36567	38453	40832	43262	42702	43066	43936	45594	47225	48522	48789	48259	46029	45581	46041		-					
4	Grand Total		36567	38453	40832	43262	42702	43066	43936	45594	47225	48522	48789	48259	46029	45581	46041							
5																								
6							AREA			-	*													
8							Alah	ama			^													
9																								
10	i						Alas	ка																
11	í.						Ariz	ona																
12							Arka	insas																
13	1						Call	fornia																
14	1						Call	UIIIIa		_														
15							Colo	rado																
16							Con	necticu	at															
17							Dela	ware																
18	1						Dele	ware			~													
19																								
20																								

Figure 3-25. Real GDP data filtered by slicer

Summary

This chapter showed how to use Pivot Tables and Pivot Charts based on a single table and how to import data from the Microsoft Azure Marketplace. Chapter 4 will show how to use the Excel Data Model to extract data from multiple tables.

CHAPTER 4

Building a Data Model

This chapter will cover the Excel Data Model also known as the PowerPivot Data Model. PowerPivot for Excel 2013 is based on the tabular engine of SQL Server Analysis Services 2012 running inside Excel. The Data Model allows relating of multiple tables. It runs completely in memory and compresses data so that it can process millions of rows.

The Excel Data Model is a collection of tables, including the relationships between them. The Data Model provides the structure for Power BI. This chapter will demonstrate how to use Pivot Tables to extract summary information from multiple tables in a Data Model. The examples used will progress from the simple to the more complex.

Enabling PowerPivot

PowerPivot is included with Excel 2013, but it is not enabled by default. To enable PowerPivot, follow these steps:

- 1. Click the File tab to see the Backstage view.
- 2. Click Options at the bottom of the left pane.
- 3. Select Add-Ins on the left pane.
- **4.** At the bottom of the window, pull down the arrow after Manage, select COM Add-ins, and click Go, as shown in Figure 4-1.

CHAPTER 4 BUILDING A DATA MODEL

	Excel O	ptions	?
General	View and manage Microsoft Office	Add-ins.	
Proofing	Add-ins		
ave			1
uve	Name 🗠	Location	Туре
anguage	Active Application Add-ins		
dvanced	Microsoft Office PowerPivot for Excel 2013	C:\dd-in\PowerPivotExcelClientAddIn.dll	COM Add-in
aroneea	Microsoft Power Map for Excel	C:\ Excel Add-in\EXCELPLUGINSHELL.DLL	COM Add-in
ustomize Ribbon	Microsoft Power Query for Excel	C:\Windows\System32\mscoree.dll	COM Add-in
uick Access Toolbar	Power View	C:\Add-in\AdHockeportingExcelClient.dll	COM Add-in
dick Access Toolbal	Team Foundation Add-In	C:\erver(11.0/amd64(1FSOfficeAdd-in.dil	COM Add-In
dd-Ins	Inactive Application Add-ins		
rust Center	Analysis ToolPak	C:\fice15\Library\Analysis\ANALYS32.XLL	Excel Add-in
lust center	Analysis ToolPak - VBA	C:\e15\Library\Analysis\ATPVBAEN.XLAM	Excel Add-in
	Date (XML)	C:\icrosoft shared\Smart Tag\MOFL.DLL	Action
	Euro Currency Tools	C:\ice\Office15\Library\EUROTOOL.XLAM	Excel Add-in
	Microsoft Actions Pane 3		XML Expansion Pack
	Solver Add-in	C:\ffice15\Library\SOLVER\SOLVER.XLAM	Excel Add-in
	Desument Polated Add ins		
	No Document Related Add-ins		
	No Document Netated Add-bis		
	Add-in: Microsoft Office PowerPiv	ot for Excel 2013	
	Publisher: Microsoft Corporation		
	Compatibility: No compatibility informat	ion available	
	Location: C:\Program Files\Microso PowerPiyotExcelClientAdo	ft Office\Office15\ADDINS\PowerPivot Excel Ad	Id-in\
	Description: Microsoft Office PowerPiv	rot for Excel 2013	
	M <u>a</u> nage: COM Add-ins 🗸 G	0	
] []			OK Cance

Figure 4-1. Excel Options screen

5. Check Microsoft Office PowerPivot for Excel 2013, Microsoft Power Map for Excel, and Power View, as shown in Figure 4-2, and click OK.



Figure 4-2. COM Add-Ins dialog

Relational Databases

Until a few years ago, relational databases were the most widely used type of database. In the last few years, NoSQL (or nonrelational) databases have become popular for log data and other types of big data. NoSQL databases will be covered more extensively in future chapters. Still, today, the majority of transactional business data is stored in relational databases.

One of the principles of relational database design is avoiding repeating fields. If repeating fields are encountered, the data is normalized by creating other tables, as described in Chapter 1. For instance, in a personnel example, the primary key of the employee table would be the employee Social Security number, because it is unique, and it is required for financial reporting. The employee would be assigned to a department, but only a department code would be stored in the employee table. The department code would be a foreign key that would point to the appropriate record in the department table, which would have other information, such as the department office, department manager, and phone number. Thus, the department information would not have to be repeated for each employee record.

Database Terminology

The following terms are commonly used to describe relational databases and the relationships between tables within a database.

- A *field* is the smallest unit of data in a table. Fields are stored in columns in Excel tables.
- A *record* is all of the data or information about one person or one thing. In Excel, a record is a row in a table.
- A *table* is a collection of related data stored in a structured format within a database. It consists of fields (columns) and records (rows).
- A *database* is a collection of tables.
CHAPTER 4 🔳 BUILDING A DATA MODEL

- The *primary key* of a relational table is a field (or fields) that uniquely identifies each record in the table. For data involving people, especially data that requires tax reporting, the most common primary key is a Social Security number. It is, of course, important to have proper security when dealing with personal data such as Social Security numbers.
- A *foreign key* is a field or fields in one table that uniquely identifies a row of another table.
- A *candidate key* is a field or fields in a table that can uniquely identify any record without referring to any other data. Each table may have one or more candidate keys, but one candidate key is selected as the primary key. This is usually the best among the candidate keys.
- When a key is composed of more than one column, it is known as a *composite key*.
- A relationship allows accessing data from multiple tables, based on a key field.
- *Normalization* is the process of organizing the tables in a relational database to minimize redundancy. Normalization often involves dividing large tables into smaller tables and defining relationships between them.
- *Third Normal Form (3NF)*, defined by relational database model inventor E. F. Codd, eliminates fields not dependent on the primary key. To be in Third Normal Form, each field in a table must depend on the primary key, and only the primary key, of that table. This is called eliminating **transitive tendencies**, where one field depends on another field that is not the primary key.
- A *source table* is the table wherein the relationship starts. In the example in the following section, the employee table is the source table.
- The *related table* is the table that contains the values to look up. In the following example, the department table is the related table.

Creating a Data Model from Excel Tables

This section will start with a simple example, to illustrate the principles of relating tables. In this example, a Data Model will be created from two small tables: employee and department. The tables are normalized, meaning that repeating fields have been eliminated. Instead of recording all the information about a single department in the employee record, only a department code is used. Then a relationship is defined between the employee and department table, so that the department information can be extracted from the department table.

A Pivot Table can be used to extract meaningful information from a Data Model. To create a Pivot Table that draws on data from both tables to show average salary by department, do the following:

- 1. Define each group of data as a table by clicking inside the table and pressing Ctrl+T.
- 2. Click anywhere in the table and click the Table Tools Design tab. On the left side of the ribbon, type in a table name, in this case, *employees*, as shown in Figure 4-3. Also, rename the departments.

CHAPTER 4 🔳 BUILDING A DATA MODEL

X		• 🔿	- 💷 +		first	excelrelationalpviot - Ex	cel		TA	BLE TOOLS							?
Tak err	ILE I le Name: ployees Resize Tab Properties	IOME	INSERT PAGE LAY Summarize with PivotTal Remove Duplicates Convert to Range Tools	VOUT FC ble Insert Slicer	RMULAS Export Refre	DATA REVIEW		W POWERP Header Row Total Row Banded Rows	IVOT Team First Column Last Column Banded Columns Table Style Options	DESIGN				Tab	le Styles		
E	A	▼ : 8	$\times \checkmark f_x$	98000 E	Export Export the SharePoint	data in this table to a list or Visio diagram.	-	j	ĸ	L	м	N	0	p	Q	R	s
1	Iname - Smith	fname John	 hiredate dept 1/1/2010 IT 	salary •	9.88E+08												
3	Anderson	Mary	3/1/2009 HR	75000	8.77E+08												
4	Johnson	Karen	6/1/2011 IT	87000	7.65E+08												
5	Franklin	Ken	11/1/2008 HR	74000	6.54E+08			deptcode 💌	deptname	▼ office ▼							
6								HR	Human Resources	312							
7								IT	Information Technolo	ogy 413							
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	

Figure 4-3. Employee and department tables

3. Highlight the employee table. Go to the Insert tab and select Pivot Table. In the Create PivotTable dialog, select *Add this data to the Data Model* and click OK, as shown in Figure 4-4.

X		5.0.				firs	texcelrelati	ionalpviot	- Excel			TAI	BLE TOOLS							11/2
		HOME	INSERT	PAGE LAYOU	T FOI	RMULAS	DATA	REVIEW	VIE	W POWERPIVO	T Tea	m	DESIGN							
P'n	otTable F	ecommend PivotTable Tables	ded Table	Pictures Onli Pictu	ine ires A+ -	🛍 Store 🎝 My Ap App	ps - 🗹 25	Recomm	nended irts	前 - 書 - 索 - 然 - 益 - 前 - _{Pi} ● - ⊵ - Charts	ivotChart	Power View Reports	Line Column Sparklin	Win/ Loss	Slicer	Timeline	Hyperlink Links	Text Box 8	Header k Footer	- π Ec - Ω Sy Sy
Ε	2	* 1	Xv	fx 98	000															
	A	В	с	D	E	F	G	н	T.	J.		ĸ	L	м	N	0	P	Q	R	s
1	Iname Smith	John	hiredate 1/1/20	✓ dept ▼ s LO IT	alary - 98000	9.88E+08					Cr	eate Piv	otTable	?	×					
3 4	Anderso Johnson	n Mary Karen	3/1/20 6/1/20	29 HR 11 IT	75000 87000	8.77E+08 7.65E+08				Choose the data	that you w	ant to ana	alyze							
5	Franklin	Ken	11/1/20	08 HR	74000	6.54E+08				Select a tabl <u>Table/Ra</u>	e or range inge: emp	ployees			18					
7										O Use an exter	mal data so	ource								
9 10										Connect	ion name:									
11										Choose where yo	u want the	PivotTabl	le report to be pla	ced	_					
13										O Existing Wor	rksheet									
14	-	-								Location	ю				18	-				
16	-									Add this dat	you want t ta to the Da	o analyze sta <u>Model</u>	multiple tables							
18													ОК	Car	ncel					
19										-										
21																				

Figure 4-4. Create PivotTable dialog

4. In the PivotTable Fields pane, click All (at the upper right), to see both tables, as shown in Figure 4-5. Drag the deptname field from the department table to the columns box. Drag Sum of salary from the employees table to values. Note the prompt "Relationships between tables may be needed." Click Create.

CHAPTER 4 BUILDING A DATA MODEL

tTable Name: // xtTable2 // Options = PivotTable	Active Field: Sum of salary Field Settings Down Active Field	↑ 1 → Group Drill 1 → Group Up - → Group Group	Selection Ind up Ind Field Sti pup	ert Insert cer Timelir Fil	Filter ne Connections ter	Refresh Change Data * Source * Data	Clear * Clear * Select * Move PrvotTable Actions	禄 Fields, Items, & Set 「衣 OLAP Tools * 中音 Relationships Calculations	s* PivotChart Re	Commended PivotTables pls	+/- Field Buttons Headers Show
*	$X \checkmark f_x$										
A	8 Column Labels 💌	C formation Technology	D Grand Total	E	F	G H	I J	K L	M N	PivotTable ACTIVE ALL Choose fields to at	Fields 🔹
Sum of salary	334000	33400	0 334000							Relationships between tables may be needed.	
										Employees Iname fname	5
										Drag fields betwee	en areas below:
										▼ FILTERS	III COLUMNS deptname
										= ROWS	Σ VALUES

Figure 4-5. PivotTable showing total salaries by department

5. As shown in Figure 4-6, select employees as the first table and dept as the foreign key. Select departments as the second table and deptcode as the primary key. Click OK. Note that the field names, as defined by the column headings, do not have to be the same.

X	1 - 5 ·			Book2 -	Excel	VIEW DOW	VERDIVOT Term	PIVOT	TABLE TOOLS	N				?	9 -/5/ Nation
Piv Ph	otTable Names votTable2 Options * PivotTable	Active Field:	↑ 13 Drill 13 Up - 13 Group 1 Up - 13 Croup 1 Croup 1 Crou	Selection p Field Sik	ert Insert Filt	Filter e Connections	Refresh Change Data * Source * Data	Clear Clear Select Move	* t * e PivotTable	院 Field	ds, litems, & Sets = IP Tools = itionships Calculations	PivotChart	Recommended PivotTables	Field + List But	/- Field tons Headers
A	3 *	$ X \lor f_x$													
1	A	В	c	D	E	F	G H	1	J	к	LI	M N	Pivot	able Fie	elds 👻
2 3 4		Column Labels 💌	nformation Technology	Grand Total									ACTIVE Choose fi	ALL elds to add to	report: 🔯 🔻
5	Sum of salary	y 334000	33400	334000			Create	alationsh				? ×	Relations	hips tables may	CREATE X
8 9				Pick the tab	Create Relationship ?						P III dej				
10	-			Enployed Related T	ts able:			dept Related	(Poreign): Column (Prim	ary):		v	A memployees		
12				departme	ents lationships l	hetween tables i	is necessary to show rel	deptcos	de m different t	ables on t	ha same report	~		fname	
14				Creating re			o necessary to show to	iccu upto int		auto on c			Drag field	s between are	ras below:
16				Manage R	elationships	has				_	OK	Cancel	Y FILTE	RS	deptname
17															
19													= ROW	s	Σ VALUES
21 22														(Sum of salary 💌
23		Sheet2 Sheet1	(+)				: [4]						• Defer	Layout Updat	e UPDATE

Figure 4-6. Create Relationship dialog

6. Drag lname to the Rows box. The resulting Pivot Table, as shown in Figure 4-7, will display the sum of salaries by department.

FIL	E HOME	- 📰 = INSERT PAGE L	AYOUT FORMULAS	Book2 - Ex DATA REV	cel NEW V	IEW POW	ERPIVO	T Team	PIN	VOTTABLE TOOL	s GN					? 05 - 0 Neil - 0
Past	Cut Copy ~ Format Pair Clipboard	Calibri B I U +	$ \begin{array}{c c} \cdot & 11 & \cdot & A^* & A^* \\ \hline \hline & \cdot & \Delta^* & \cdot & \Delta^* \\ \hline \hline & \cdot & \Delta^* & \cdot & \Delta^* \\ \hline \\ \hline \\ \text{ont} & & r_2 \end{array} $	=	E Men	ip Text rge & Center	Ger Ş	• % • •	• • • • • • • • •	Conditional Formatting *	ormat as Table - St yles	Cell Ir yles -	nsert Delete Fo	ormat	AutoSum * A Fill * Z Clear * Filter - S Editing	Find & jelect *
A3	•	$\times \checkmark f_x$	Sum of salary													
1 2 3 5	A Sum of salary	B Column Labels 💌	C .	D	E	F	G	н	1	J	К	L	M	N 4	PivotTable F ACTIVE ALL Choose fields to add	ields 🔹
5 4 6 F 7 J	Anderson Franklin Johnson	75000 74000	87000	75000 74000 87000											✓ employees ✓ Iname	•
9 (10 11 12	Grand Total	149000	185000	334000											☐ hiredate ☐ dept ☑ salary	
13 14 15															Drag fields between	areas below:
16 17 18																deptname
20 21 22															E ROWS	∑ VALUES Sum of salary ▼
23	Sh	eet2 Sheet1	(+)					: (4)							Defer Layout Upo	Jate UPDATE

Figure 4-7. PivotTable showing employee salary by department

7. To see instead the average salary by department, first remove the lname field from the Rows box and then in the PivotTable Fields pane, click the drop-down arrow next to Sum of salary in the Values box. Select value field setting and then average to see the average salary by department, as shown in Figure 4-8.

CHAPTER 4 BUILDING A DATA MODEL

E FILE	5 · ♂ HOME	- 🔝 = INSERT	PAGE LAYOUT	firstexcelrelation FORMULAS DATA	nalpviot - Excel REVIEW VIEW	POWEF	PIVOT Te	Piv am ANA	OTTABLE TOOLS	րդդ	20				? 🗉 – 🗗 > Neil - 🖸
Manage Data Mode	Calculate Fields * Calcul	d KPls	Align Vertically	Add to Update De Data Model All Tables Relati	etect Settings										
C3	•	×	f _x Column I	Labels											,
a A 1		B	c	D	E	F	G	н	L J	к	L	м	-	PivotTable	Fields - ×
2 3 4	-		Column Labels	Information Technolog	y Grand Total									ACTIVE ALL Choose fields to add to report:	
5 6 7 8 9	Aver	age of salar	y 74500	925	00 83500									minimum employees iname fname hiredate dept	•
11 12 13														Drag fields between	• areas below:
15 16 17														▼ FILTERS	III COLUMNS deptname •
18 19 20														■ ROWS	∑ VALUES Average of sal ▼
21	SI	neet2 Sh	eet3 Sheet1	(+)			1.4						•	Defer Layout Up	date UPDATE

Figure 4-8. PivotTable showing average salary by department

8. This example showed how to relate two tables, employees and departments, using department code. dept is a foreign key in the employees table, and deptcode is the primary key for the departments table. The Pivot Table relates the two tables.

Loading Data Directly into the Data Model

In this example, data will be loaded directly into the Data Model using an example from Microsoft's Northwind Access Database. This example demonstrates running PowerPivot on top of Excel.

To load data directly into the Data Model, do the following:

- 1. Click the PowerPivot tab to open the ribbon.
- 2. Click the Manage button at the top-left end of the ribbon.
- **3.** Once the PowerPivot window is open, you can load data by clicking *From Database* in the Get External Data group. This provides a choice of database sources, as shown in Figure 4-9. For this example, *From Access* will be selected.

CHAPTER 4 🔳 BUILDING A DATA MODEL



Figure 4-9. PowerPivot Get External Data group

- 4. In the Table Import Wizard dialog, click the Browse button to find the Access file you want to import. For this example, Database1 in the documents folder will be selected. This file is available from the Apress web site.
- 5. Click Test Connection. If the test is successful, click Next.
- **6.** PowerPivot opens the Table Import Wizard, as shown in Figure 4-10, to step you through the process of loading data. Choose *Select from a list of tables and views to choose the data to import.* Click Next.

Table Import Wizard	?	×
Choose How to Import the Data		
You can either import all of the data from tables or views that you specify, or you can write a query using SQL that specifies the data to import.		
Select from a list of tables and views to choose the data to import		
○ <u>W</u> rite a query that will specify the data to import		
< <u>B</u> ack <u>N</u> ext > <u>F</u> inish (Cance	

Figure 4-10. Table Import Wizard

7. Click the top-left box to select all tables, as shown in Figure 4-11. Note that in the lower right corner of the dialog, there is an option to *Preview & Filter*. This option can be used to select a subset of data from very large datasets. In this example, all rows in all tables will be imported. Click Finish to complete the import.

ect	the ta	bles and views that you wa	nt to import data from.		
Dat <u>T</u> ab	abas les a	se: C:\Users\Neil\Documen nd Views:	ts\Database1.accdb		
•		Source Table	Friendly Name	Filter Details	^
•		Categories	Categories		
✓		Customers	Customers		
•		Employees	Employees		
•		Inventory	Inventory		
✓		Invoices	Invoices		
✓		OrderDetails	OrderDetails		
✓		OrderDetailsStatus	OrderDetailsStatus		
✓		Orders	Orders		
✓		OrderStatus	OrderStatus		
✓		Products	Products		
◄		PurchaseOrderDetails	PurchaseOrderDetails		
◄		PurchaseOrders	PurchaseOrders		
✓		SalesReports	SalesReports		
✓		Shippers	Shippers		
		C 1			•
			Select Rel	ated Tables Preview	& Filter

Figure 4-11. Table Import Wizard—Select Tables and Views

Relationships are automatically loaded into the data model, if they are present in the imported data. Note that PowerPivot tables are read-only, so that the data cannot be changed.

To view relationships, go to the PowerPivot tab and select Manage to open the PowerPivot window. Then click the PowerPivot Diagram View at the right end of the ribbon to see the relationships between tables, and click Finish to complete the import, as shown in Figure 4-12.

CHAPTER 4 🔳 BUILDING A DATA MODEL



Figure 4-12. Diagram View

Creating a Pivot Table from Two Tables

In this example, a Pivot Table will be created, based on the inventory and products tables. Notice in Figure 4-12 that the two tables are linked, based on product code. The Inventory table includes the on-hand data for each item in inventory, with a primary key of ProductCode. To find the full product name, a lookup has to be performed in the product table, based on product ID. To create a Pivot Table based on these two tables, do the following:

- 1. Click Pivot Table on the File tab in the PowerPivot window and select PivotTable from the menu.
- 2. Click OK to put the Pivot Table in a new worksheet.
- 3. Click the arrow next to the products table in the Pivot Table Fields pane and drag the ProductName field to the Rows box. Then scroll up to go to the inventory table, click the left arrow to show the fields, and drag OnHand to the Values box. The results are shown in Figure 4-13, which indicates the sum of the on-hand values for each product.

FLE	ち・ c [⇒] ・ □ = HOME INSERT PAGE LAYOUT FORMU	importnort	twindsept 23 last - Excel A REVIEW VIEW	POWERPIVOT Team	PIVOTTABLE TOOLS ANALYZE DESIGN	ada ya 🖬		0	? 🗉 – 🗗 🗙 Neil - 🖸
Manage Data Model	Calculated KPIs Fields - Calculations Slicer Alignment	to Update fodel All Tables	Detect Settings						~
83	\star : $ imes$ \checkmark $f_{\!X}$ Row Labels								~
A	B	с	D	E		F	-		
1								Pivot lable F	ields **
2	Reve Labole	Colland						ACTIVE ALL	
4	Northwind Traders Almonds	100						Choose fields to add	to report: 🛛 🕈 🔻
5	Northwind Traders Beer	100							
6	Northwind Traders Boysenberry Spread	150						U Waste	
7	Northwind Traders Brownie Mix	0						Shipped	- n
8	Northwind Traders Cajun Seasoning	400						Allocated	
9	Northwind Traders Cake Mix	0						BackOrdere	d
10	Northwind Traders Chai	200						□ InitialLevel	
11	Northwind Traders Cherry Pie Filling	55						OnHand	w
12	Northwind Traders Chicken Soup	500							
13	Northwind Traders Chocolate	150					_	Drag fields between a	areas below:
14	Northwind Traders Chocolate Biscuits Mix	100						T FU TERS	III COLUMNS
15	Northwind Traders Clam Chowder	55						T HEIERS	E COLONICS
16	Northwind Traders Coffee	150							
17	Northwind Traders Corn	100							
18	Northwind Traders Crab Meat	200							
19	Northwind Traders Dried Apples	100						E ROWS	Σ VALUES
20	Northwind Traders Dried Apples	100						ProductName •	OnHand •
22	Northwind Traders Dried Plans	100							
23	Northwind Traders Fruit Cocktail	50					1		
24	aparter of graden Darabi								LIDDATE
4 1	Sheet2 Sheet1 (+)			1			Þ	Derer Layout Opd	UPDATE

Figure 4-13. PivotTable showing product name and on-hand quantity

Creating a Pivot Table from Multiple Tables

This example will demonstrate how to create Pivot Tables from more than two tables, based on the relationships between the tables. It will use the PowerPivotTutorialSample spreadsheet from the Contoso database, which can be downloaded from the following URL: Tinyurl.com/PowerPivotSamples.

A Pivot Table accessing fields from multiple related tables can be created by following these steps:

- 1. Download the PowerPivotTutorialSample spreadsheet from the preceding URL.
- 2. Click the Manage icon on the PowerPivot tab to open the PowerPivot window. Then click Diagram View to see the diagram of relationships, as shown in Figure 4-14. Note that the arrows show relations in a somewhat imprecise manner: using the scrollbar to scroll through the field names does not change the position of the arrow. A clearer way to see relationships is to minimize the PowerPivot window and click Relationships on the Data tab, to see the actual field names that are linked in the Manage Relationships screen, as shown in Figure 4-15. Note that you can also see a similar view by clicking Manage Relationships on the Design tab of the PowerPivot window.

CHAPTER 4 BUILDING A DATA MODEL



Figure 4-14. Diagram view of relationships

	Manage I	Relationships	? ×						
Status	Table 🔶	Related Lookup Table	<u>N</u> ew						
Active	dbo_DimProduct (ProductSubcategoryKey)	dbo_DimProductSubcategory (ProductSubcate	Edit						
Active	dbo_DimProductSubcategory (ProductCategory	ProductCategory (ProductCategoryKey)	Editin						
Active	dbo_FactSales (channelKey)	dbo_DimChannel (ChannelKey)	Activate						
Active	dbo_FactSales (DateKey)	dbo_DimDate (Datekey)	Deactivate						
Active	Active dbo_FactSales (StoreKey) Stores (StoreId)								
Active	Active Stores (EntityId) dbo_DimEntity (EntityKey)								
			Close						

Figure 4-15. Manage Relationships screen

- 3. To create a Pivot Table based on related fields, select the PowerPivot tab and Click Manage. If the Diagram View is still displayed, click Data View. Notice the tabs at the bottom of the screen that show the tables included in this example. A PivotTable will be created that links the dbo_FactSales table with the dbo_DimProduct table and the dbo_DimDate table. The purpose of the Pivot Table is to analyze product sales by date.
- **4.** Click Pivot Table and select Pivot Table from the drop-down. In the Create PivotTable dialog, accept the default of New Worksheet by clicking OK.
- 5. In the PivotTable Fields pane, the tables will be shown preceded by a right-facing arrow. Click the right-facing arrow to see the fields in that table. Scroll to the dbo_DimProduct table and click the arrow. Drag the ProductName field to the Rows box. Click the arrow to the left of the dbo_DimProduct table to suppress the fields display. Scroll to the dbo_DimDate table and click the arrow, to see the fields. Drag the FullDateLabel field to the Columns box. Click the arrow to suppress the fields. Scroll to the dbo_FactSales table and click the arrow, to see the fields. Drag the SalesQuantity field to the Values box. The results will be shown in Figure 4-16. It is possible to link fields from multiple tables because of the relationships shown in Figure 4-15.

FILE	ном	C [®] - □ = Copy of PowerPivotTuto E INSERT PAGE LAYOUT FORMULAS DATA	rialSamplesept25 - Excel REVIEW VIEW PC	WERPIVOT	Pivi Team ANAL	OTTABLE TOOLS YZE DESIGN	րդի			? 🗉 – 5 Neil -
1	& Cut	Calibri - 11 - A A = = = - >	- Wrap Text	General				# > m	∑ AutoSum + A	46
Paste	Copy		an Maran Ri Canta	5 . 04		Conditional Forma	at as Cell	Insert Delete Format	Fill • Sort 8	Find &
*	💞 Format	Painter	Herge & Center	* \$*70	, 00 +0 F	ormatting * Tabl	le * Styles *		Clear * Filter	Select *
	Clipboard	rs Font rs	Alignment	rs Nur	iber 15	Styles		Cells	Editing	
84	*	$f_x \sim f_x$ Row Labels								
al.	А	В	с	D	E	F	G	н		etala e
2									Pivot lable	Fields
3		Sum of SalesQuantity	_ Column Labels 💌						ACTIVE ALL	
4		Row Labels	2007-01-01	2007-01-02	2007-01-03	2007-01-04	2007-01-05	5 2007-01-06 20	Choose fields to as	id to report:
5		A. Datum Advanced Digital Camera M300 Azure								
6		A. Datum Advanced Digital Camera M300 Black	54	70	18	8	7:	2 27	UnitCos	
7		A. Datum Advanced Digital Camera M300 Green	18	60	51	63	3	5 9	UnitPric	•
8		A. Datum Advanced Digital Camera M300 Grey	34	9	27	39	9	9 31	SalesQu	antity
9		A. Datum Advanced Digital Camera M300 Orange	31	27	9	34	50	0 63	ReturnQ	uantity
10		A. Datum Advanced Digital Camera M300 Pink	42	18	29	18	39	9 22	Dran fieldr hebure	in arear below
11		A. Datum Advanced Digital Camera M300 Silver	16	29	27	79	17	7 26	ung neur cente	
12		A. Datum All in One Digital Camera M200 Azure							▼ FILTERS	III COLUMNS
13		A. Datum All in One Digital Camera M200 Black	9		35	13	1	7 34		FullDateLabel
14		A. Datum All in One Digital Camera M200 Green	36	38	42	44	34	4		
15		A. Datum All in One Digital Camera M200 Grey	8	18	35		2	5 17		
16		A. Datum All in One Digital Camera M200 Orange							I ROWS	Σ VALUES
17		A. Datum All in One Digital Camera M200 Pink	116	26	44	83	34	6 35	ProductName *	Sum of Sales •
18		A. Datum All in One Digital Camera M200 Silver	48	17	25	62	2	7 27		
19		A. Datum Bridge Digital Camera M300 Azure	35	18	31	9	2	5 34		
20	•	Sheet2 Sheet1 Sales by Channel Profit by Categ	ory ContosoStores	💮 ! 💽				3	Defer Layout U	UPDAT

Figure 4-16. PivotTable showing sales by product and date

6. To further modify the Pivot Table by viewing sales by product by year, remove FullDateLabel from the Columns box and replace it with CalenderYear from the DimDate table. The results are shown in Figure 4-17.

CHAPTER 4 BUILDING A DATA MODEL

E FILE	5 - ном	C [®] - □ = Copy of PowerPivotTutor E INSERT PAGE LAYOUT FORMULAS DATA I	ialSamplesept25 - Excel REVIEW VIEW POWI	RPIVOT 1	eam AM	PIVOTTABLE TOOLS IALYZE DESIGN					? E - 8 > Nei - 0
The a	K Cut	Calibri - 11 - A A = = = 🖉 🗞	- 🐺 Wrap Text	General				# *		AutoSum • A	44
Paste	Copy	в / u · 🗄 · 🍐 · д · 🚍 = 🚝	1 Merge & Center	5 - %	,	Conditional Form	nat as Cell	Insert Delete	Format	Fill + Sort & F	ind &
	Pormat	Painter				Formatting * Ta	ble * Styles *	• •		Clear Filter S	elect -
84	*	$\therefore \sqrt{f_x}$ Row Labels	ung marcin.			July				Loung	
ai.	A	В	с	D	E	F	G	н	1.		
2			2							PivotTable F	ields •×
3		Sum of SalesQuantity	Column Labels 🕞							ACTIVE ALL	
4		Row Labels	2007	2008	2009	Grand Total				Choose fields to add	to report:
5		A. Datum Advanced Digital Camera M300 Azure	8999	6097	3846	18942					
6		A. Datum Advanced Digital Camera M300 Black	12087	5971	4032	22090				FullDateLat	xel 🗎
7		A. Datum Advanced Digital Camera M300 Green	11400	5922	4958	22280				DateDescrip	ption
8		A. Datum Advanced Digital Camera M300 Grey	10713	6125	3729	20567				CalendarYe	tar
9		A. Datum Advanced Digital Camera M300 Orange	12181	6193	4354	22728				CalendarYe	arLabel
10		A. Datum Advanced Digital Camera M300 Pink	11121	5648	4507	21276				Drag fields between	seast halour
11		A. Datum Advanced Digital Camera M300 Silver	10776	5912	3833	20521				engineas etineen	
12		A. Datum All in One Digital Camera M200 Azure			4765	4765					III COLUMNS
13		A. Datum All in One Digital Camera M200 Black	11411	5867	4514	21792					CalendarYear 👻
14		A. Datum All in One Digital Camera M200 Green	11280	6178	4368	21826					
15		A. Datum All in One Digital Camera M200 Grey	11638	6347	4101	22086				-	
16		A. Datum All in One Digital Camera M200 Orange	8797	6137	4265	19199			_	E ROWS	Σ VALUES
17		A. Datum All in One Digital Camera M200 Pink	11571	5901	4811	22283			_	ProductName •	Sum of Sales •
18		A. Datum All in One Digital Camera M200 Silver	11820	6035	3886	21741					
19		A. Datum Bridge Digital Camera M300 Azure	10869	5722	3932	20523					
20		A Datum Bridge Digital Camera M200 Black	N200	A 102	4477	20240				Defer Layout Upd	uppate

Figure 4-17. PivotTable showing sales by product and year

Adding Calculated Columns

Calculated columns can be added to a table in PowerPivot, using Data Analysis Expressions (DAX). Regular Excel formulas are based on cells. By contrast, DAX formulas are based on columns. DAX will be covered more extensively in future chapters. This example will show how to set up a calculated column. To add a calculated column, take the following steps:

- 1. Select the PowerPivot tab and click Manage.
- 2. Select the dbo_FactSales table by clicking the tab at the bottom of the display.
- **3.** Scroll to the right to the TotalProfit column. Note that to the right, the column heading is Add Column.
- 4. Select any cell in that column and type = to start a formula.
- 5. Create an extended price column that is the product of unit price times quantity. To do this, after typing =, click the SalesQuantity column, type *, and then click the unit price quantity. The formula is shown in Figure 4-18, in the formula bar: =[SalesQuantity]*[UnitPrice]

	5.	C · 🗐 +		Copy of PowerPivotTutori	alSamplesept25 - Excel		PW	OTTABLE TO	POLS		-		e el 74	15/14
FILE	но	(a) (a) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	* ÷ ≑	PowerPivot	for Excel - Copy of Pow	werPivotT	utorialSample	sept25.xls	sx		×		Nei	i- [0]
F	0	File Home	Design Ad	wanced	AI		States		0.01.01.01	^	0			
Manage	Cale		3	Eormat : General :	24		Create KRI		G Chagram View					
Data Mod	Fi Fi	Paste Get Ext	ernal Refresh	PivotTable 5 - % 2 3	Clear All Sort by	Find	a create avr	Data D	A Calculation Area					^
		Clipboard		Formatting	Sort and Filter	Find	Calculations		View					
84	_		X √ f _x =	[SalesQuantity]*[UnitPrice	2]						¥			^
		- UnitCost	UnitPrice	SalesQuantity 🖬 Return	Quantity 💽 ReturnAm	iount 💌	DiscountQuant	ity 🖬 D	AscountAmount	TotalCost 💌	1			
		\$48.43	\$94.99	9	0	\$0		0	\$0	\$435.87				
- Al-	A	\$134.74	\$293	9	0	\$0		0	\$0	\$1,212.66	III P	DivotTable E	olds	- ×
2	_	\$167.73	\$329	9	0	\$0		0	\$0	\$1,509.57		rivotrable i	leius	
3		\$96.08	\$289.99	9	0	\$0		0	\$0	\$864.72		ACTIVE ALL	3	
4		\$261.66	\$569	9	0	\$0		0	\$0	\$2,354.94		Choose fields to add t	o report:	Q٠ •
5		\$141.47	\$427	9	0	\$0		0	\$0	\$1,273.23				
6		\$7.59	\$16.5	9	0	\$0		0	\$0	\$68.31		FullDateLab	el	÷.
7		\$5.09	\$9.99	9	0	\$0		0	\$0	\$45.81		DateDescrip	tion	_
8		\$73.12	\$159	9	0	\$0		0	\$0	\$658.08		CalendarYe	ar	
9		\$118.65	\$258	9	0	\$0		0	\$0	\$1,067.85		CalendarYe	arLabel	Ŧ
10		\$75.96	\$149	9	0	\$0		0	\$0	\$683.64		Drag fields hetware a	rear helows	
11		\$6.39	\$13.89	9	0	\$0		0	\$0	\$57.51		ong news sections	itas vervin.	
12	_	\$71.37	\$139.99	9	0	\$0		0	\$0	\$642.33		T FILTERS	III COLUMN	45
13		\$21.92	\$43	9	0	\$0		0	\$0	\$197.28			CalendarYea	ar •
14		\$152.44	\$299	9	0	\$0		0	\$0	\$1,371.96				
15		\$827.97	\$2,499	9	0	\$0		0	\$0	\$7,451.73	~			
16											^	≡ ROWS	Σ VALUES	
17												ProductName *	Sum of Sale	s •
18														
19											~			
20		<	na nintra la				forefore Decide	-	lan companya lan	; 	2	Defer Layout Upd	te U	JPDATE
	·	l doo_umchannel d	a des are		ct doo_uimefoductSubcate	gory 600	Factoales Produ	ctuategory	Be contosostores Sto	ies		FEB 1001		1

Figure 4-18. Adding calculated column

- 6. Then press Enter to complete the formula.
- 7. To rename the column, right-click the column heading, select Rename Column, and name it "extended price." The results are shown in Figure 4-19. Note that the values in the new column are the same as the TotalSales but are based on a formula.

B B Clip DeverPhoot for Eacel - Copy of PowerPhootTutorialSamplesept25.xix Clip Name		5.	e.	-					Copy of Power	PivetTut	orialSampleseot25 -	Excel	3 III IIIII/	- 6		- 181 -2	1/6/
Image of the price Image of the price Advanced Image of the price Advanced Image of the price	FILE	но	1 a	1 🖬 🗇 🖑 -	= 1		Power	Pivot for Exc	cel - Copy of Pow	erPivot	TutorialSamples	ept25.xlsx		× 🔍		N	leil - K
Marge Cr Parts Data Type:: Data Type::		1	File	Home D	esign A	Advanced							^	0			
A1 Constants Vers A Constants Vers A Constants Constants Vers A Constants Constants <thconstants< th=""> Constants Const</thconstants<>	Manage Data Mod	Calc Fie	Paste	Get Externa Data •	Refresh	PivotTable	Data Type : Format : ~ \$ ~ % >	- 2 X X8 +33 2	Clear All Sort by Filters Column	Find	∑ AutoSum -	Data View Diagram View Show Hidden Data					
A CostaCost TotalSust TotalProfit extended price Add Column 9 5435.87 \$\$45.87 \$\$1,212.66 \$\$2,637 6 1/1/2008 12.0 \$\$1,428.43 \$\$2,637.00 \$\$1,212.66 \$\$2,637 6 1/1/2008 12.0 \$\$1,428.43 \$\$2,637.00 \$\$1,51.50 \$\$2,535.00 \$\$1,212.66 \$\$2,637.00 \$\$1,212.66 \$\$2,637.80 \$\$1,212.66 \$\$2,637.80 \$\$1,122.50	A1		Cipoc	vara 🗸	fx		Pormas	ung	Sort and Pitter	Find	Calculations	view		8			
A Construction Construct				IntalCost 🔽 T	atalSalar		UpadiD 🔽	LoadDate	UndateDate	Tota	Profit	avtended price	Add Column				
A b \$1,212.66 \$2,637 6 1/1/2008 12 1/1/2008 12 \$1,424.34 \$2,637.00 PivotTable Fields 2 b) \$1,505.57 \$2,261 6 1/1/2008 12 \$1,1/1/2008 12 \$1,1/1/2008 12 \$1,1/1/2008 12 \$1,1/1/2008 12 \$1,1/1/2008 12 \$1,775.5 \$1,253.43 \$5,210 \$1,775.3 \$1,277.23 \$3,3483 \$1/2/2008 12 \$2,2660.77 \$3,3483.00 \$1 \$1 \$1/2/2008 12 \$2,766.06 \$5,121.00 \$1 \$1 \$1 \$1,272.23 \$3,3483 \$1/2/2008 12				\$435.87	Ś	854.91	6	1/1/2008 12	1/1/2008 12:0.		\$419.0	4 \$854.91	Aug Commit				
2 > \$1,599.57 \$2,961 6 1/1/2006 12.0 \$1,451.43 \$2,961.00 PMORTaBDE FIEldS 3 > \$264.72 \$2,2609.91 6 1/1/2006 12.0 \$1,745.19 \$2,2609.91 ACTVE ALL 5 > \$54,592.45 \$51.21 1/1/2008 12.0 \$2,569.77 \$3,843.00 > Choose Fields to add to report Choose Fields to add to rep	- 54	Α)	\$1,212.66		\$2,637	6	1/1/2008 12	1/1/2008 12:0.		\$1,424.3	4 \$2,637.00		1	Table F		~
3 > 5 \$864.72 \$2,2609.91 6 1/1/2008 12 1/1/2008 12 \$1,745.19 \$2,2609.91 ACTW ALL 4 > 5 \$2,354.94 \$5,121 6 1/1/2008 12 1/1/2008 12 \$2,766.06 \$5,121.00 Choose fields to add to report. Image: Choose fields to add to report. Imag	2)	\$1,509.57		\$2,961	6	1/1/2008 12	1/1/2008 12:0.		\$1,451.4	3 \$2,961.00			ot able F	ielas	1
4 → 52,354.34 \$5,121 6 1/1/2008 12 1/1/2008 12 \$2,766.06 \$5,121.00 → Choose fields to add to report. ⊕ 5 → 51,272.23 \$3,843 6 1/2/2008 12 \$2,766.06 \$5,121.00 → ⊕	3)	\$864.72	\$2,	609.91	6	1/1/2008 12	1/1/2008 12:0.		\$1,745.1	9 \$2,609.91		ACT	IVE ALL		
5 1 \$1,273.23 \$3,843 6 1/2/2006 12:0 \$2,569.77 \$3,843.00 Image: Solid Soli	4)	\$2,354.94		\$5,121	6	1/1/2008 12	1/1/2008 12:0.		\$2,766.0	6 \$5,121.00	6	Cho	ose fields to add t	o report:	4
6 D \$583.31 \$148.5 6 1/2/2008 12 1/2/2008 12 \$580.19 \$148.50 P 0 Controstores P 0 Controstores P 0 <td>5</td> <td></td> <td>3</td> <td>\$1,273.23</td> <td></td> <td>\$3,843</td> <td>6</td> <td>1/2/2008 12</td> <td> 1/2/2008 12:0.</td> <td></td> <td>\$2,569.7</td> <td>7 \$3,843.00</td> <td></td> <td></td> <td></td> <td></td> <td></td>	5		3	\$1,273.23		\$3,843	6	1/2/2008 12	1/2/2008 12:0.		\$2,569.7	7 \$3,843.00					
7 1 545.81 \$89.91 6 1/2/2008 12 1/2/2008 12 544.1 \$89.91 0 0 0 0 0 5658.08 \$1.431 6 1/2/2008 12 1/2/2008 12 5772.92 \$1.431.00 0 0 0 0 0 5668.64 \$1.341 6 1/2/2008 12 \$1/2/2008 12	6)	\$68.31		\$148.5	6	1/2/2008 12	1/2/2008 12:0.		\$80.1	9 \$148.50		• E	ContosoStores		-
8 □ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	7		3	\$45.81		\$89.91	6	1/2/2008 12	1/2/2008 12:0.		\$44.	1 \$89.91		> E	dbo_DimChan	nel	
9 D 51,067.85 52,322 6 1/2/2006 12 1/2/2006 12 51,254.15 52,322.00 10 D 5683.64 51,341 6 1/3/2006 12 1/3/2006 12 567.53 51,1341.00 Drag fields between areas below: 11 D 557.51 512.25.01 1/3/2008 12 1/3/2008 12 5687.5 512.25.01 Drag fields between areas below: 12 D 5642.33 51,259.91 6 1/3/2008 12 1/3/2008 12 510.75.8 51.259.91 Tell	8)	\$658.08		\$1,431	6	1/2/2008 12	1/2/2008 12:0.		\$772.9	2 \$1,431.00		5.0	dho DimDate		
10 10 5683.64 \$1,341 6 1/3/2008 12:0 \$657.36 \$1,341.00 Drag fields between areas below: 11 10 557.51 \$125.01 6 1/3/2008 12:0 \$567.36 \$1,341.00 Drag fields between areas below: 12 0 \$564.23 \$1,25.01 6 1/3/2008 12:0 \$567.36 \$1,25.01 The provide the providet the providet the provide t	9		3	\$1,067.85		\$2,322	6	1/2/2008 12	1/2/2008 12:0.		\$1,254.1	5 \$2,322.00					- 6
11 12 \$57,51 \$125,01 6 1/3/2008 12: 1/3/2008 12: \$67,5 \$125,01 Drag fields between areas below: 12 12 5 \$542,33 \$1,259,91 6 1/3/2008 12: \$617,56 \$1,259,91 T T T T \$13/2008 12: \$1/3/2008 12:	10)	\$683.64		\$1,341	6	1/3/2008 12	1/3/2008 12:0.	.,	\$657.3	6 \$1,341.00					
12 D \$642.33 \$1,259.91 6 1/3/200812 1/3/200812 \$617.58 \$1,259.91 \$1	11		3	\$57.51	\$	125.01	6	1/3/2008 12	1/3/2008 12:0.		\$67.	5 \$125.01		Drag	g fields between a	reas below:	
13 1 5197.28 5387 6 1/4/2008 12: 5189.72 5387.00 14 2 51,371.96 52,691 6 1/4/2008 12: 1/4/2008 12: 51,319.04 52,691.00 15 2 57,451.73 522,491 6 1/4/2008 12: 1/4/2008 12: 515,039.27 522,491.00	12)	\$642.33	\$1,	259.91	6	1/3/2008 12	1/3/2008 12:0.		\$617.5	8 \$1,259.91		T	FILTERS	III COLUN	INS
14 15 51,371,96 52,691 6 1/4/2008 12:0 51,319,04 52,691.00 15 1 57,451,73 522,491 6 1/4/2008 12:0 515,039.27 522,491.00 16 1 1 1/4/2008 12:0 1/4/2008 12:0 515,039.27 522,491.00 16 1 1 1/4/2008 12:0 1/4/2008 12:0 515,039.27 522,491.00 17 1 1 1 1 1 <td>13</td> <td>_</td> <td>3</td> <td>\$197.28</td> <td></td> <td>\$387</td> <td>6</td> <td>1/4/2008 12</td> <td> 1/4/2008 12:0.</td> <td></td> <td>\$189.7</td> <td>2 \$387.00</td> <td>6</td> <td></td> <td></td> <td>CalendarY</td> <td>rear 🔻</td>	13	_	3	\$197.28		\$387	6	1/4/2008 12	1/4/2008 12:0.		\$189.7	2 \$387.00	6			CalendarY	rear 🔻
b 57,451.73 \$22,491 6 1/4/2008 12:0 \$15,039.27 \$22,491.00 Image: Control of the contro	14	-	3	\$1,371.96		\$2,691	6	1/4/2008 12	1/4/2008 12:0.		\$1,319.0	4 \$2,691.00	i				
16 I I I I I I I I I I I I I I I I I I I	15	_)	\$7,451.73	\$	22,491	6	1/4/2008 12	1/4/2008 12:0.		\$15,039.2	\$22,491.00		~			
17 18 19 19 19 19 19 19 19 19 19 19 19 19 19	16									0				= =	ROWS	∑ VALUES	5
18 19 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	17	_												Pre	oductName •	Sum of Sal	les •
19 Contraction of the contractio	18	_															
Sn. C Defer Layout Update UPDAT	19													~			
dbo_DimChannel dbo_DimEntity dbo_DimProduct dbo_DimProductSubcategory abo_FactSales ProductCategory as ContosStores Stores	20	_	<								100		>	2 P	Net at 1 and 1 and	100	UDDAT
	4	•	dbo_D	imChannel dbo	DimDate	dbo_DimEnt	ity dbo_Dim	Product dbo	DimProductSubcateg	ory dbo	FactSales Produc	tCategory an ContosoStores	Stores		verer cayout Upd	ste	UPDAT

Figure 4-19. Column renamed "extended price"

When using PowerPivot calculated columns, all cells share the same formulas. It is not possible to use formulas involving individual cells. Aggregation functions such as SUM(), Average(), and COUNT() can apply to an entire column, to extract summary data.

Adding Calculated Fields to the Data Model

This example will show how to design a Pivot Table that shows sales quantity by product by month, using calculated fields. To do this follows these steps:

- 1. Click the Pivot Table icon and accept the default of a new sheet.
- **2.** From the dbo_DimDate table, drag the CalenderMonthLabel field to the Rows box.
- **3.** From the ProductCategory table, drag theProductCategoryName field to the Columns box.
- 4. From the dbo_FactSales table, drag the SalesQuantity field to the Values box.

To add a calculated field to the data model, take the following steps:

- 1. On the PowerPivot tab, click Calculated Field.
- 2. Click new Calculated Field.
- **3.** From the drop-down, select the table on which the calculated field will be based, in this case, dbo_FactSales.
- 4. Name the calculated field "profitability."
- 5. Click the Fx icon, to see the functions. Many of the functions are similar to basic Excel functions.
- 6. Scroll down to SUM(), as shown in Figure 4-20, and click OK.

	Calculated Field	?	×
Table name:	dbo_DimChannel		~
Calculated field name:	profitability		
Description:			
Formula: fx Che	sk formula		
	me)]		^
Formatting Options			~
Category: General Number Currency Date TRUE/FALSE			
	ОК	Cano	el

Figure 4-20. Calculated Field dialog

- 7. Following the left parenthesis, start typing the name of the table to use, in this case, dbo_FactSales. The fields for that table will then be displayed. Scroll down to dbo_FactSales[totalprofit] and double-click. Type the closing parenthesis. Then type / and sum(dbo_FactSales[totalsales] and close the parenthesis.
- 8. Then Click *check formula*. If the formula tests OK, click OK.

The calculated field column will appear as shown in Figure 4-21.

CHAPTER 4 🔲 BUILDING A DATA MODEL

YOUT	FORMULAS	DATA	REVIEW	VIEW	POWERPIVO	T Team	ANALYZE	DESIGN				N	leil - K
rtically prizontally gnment	Add to Data Model Tables	Ipdate All	Detect Relationships	Settings									
A	В		с		D		E	F	G	-	DivetTable F	iolda	
2											PIVOLI ADIE F	leius	
3		(olumn Labe	ls 👻							ACTIVE		-
4		,	UDIO			CAMERAS &	CAMCORDERS		CELL PHONES		Choose fields to add t	o report:	\$
5	Row Lab	els 💌 S	um of Sales	Quantity	profitability	Sum of Sales	Quantity	profitability	Sum of SalesQuantity	profi			
6	April			105792	0.564066942		553823	0.598564699	1098835	0.55	Relationships	CREATE	
7	August			126271	0.570772717		559430	0.595386325	1195289	0.55	be needed.	Grievine	
8	Decemb	er		136120	0.57209954		629430	0.593788902	1282760	0.54	▷ □ ContosoStores		-
9	February	1		82079	0.563759794		457396	0.59519889	918441	0.54	_		- 1
0	January			81552	0.569484401		463366	0.596732292	958397	0.55	DimChannel	nel	
1	July			129378	0.567648063		586436	0.592118875	1255770	0.5			
2	June			114858	0.573078599		560415	0.596352499	1143624	0.55	Drag fields between a	reas below:	
5	March			85/3/	0.562312927		4481/8	0.592830026	8/3839	0.55			
6	Novomb	or		126521	0.502520032		531184	0.590435425	1204739	0.55	T FILTERS	III COLUN	INS
6	October	ei		121222	0.505855585		549012	0.53501500	1120207	0.55		ProductCa	a •
7	Sentemi	or		121566	0.573613501		541549	0.595262881	1121090	0.55		Σ Values	•
18	Grand To	otal		1353298	0.569514254		6551189	0.595910831	13459636	0.55	II ROWS	Σ VALUES	s
19											CalendarMo	Sum of Cal	-
20											Calendativio	profitabilit	h
21												II Promabilit	9

Figure 4-21. PivotTable showing profitability and sales by month

Summary

This chapter provided an introduction to building a Data Model in PowerPivot. In future chapters, more advanced features of the Data Model will be explored. Figure 4-22 shows the number of rows downloaded for each table. Notice that the dbo_FactSales table has almost 4 million rows. That number of rows is supported in PowerPivot but not in Excel.

	Т	able Import Wizard	? ×						
Importin	g								
The import operation might take several minutes to complete. To stop the import operation, click the Stop Import button.									
	Success	Total: 6	Cancelled: 0						
Deta	aile:	Success: 6	Error: 0						
	Work Item	Status	Message						
	dbo_DimChannel	Success. 4 rows transferred.							
\bigcirc	dbo_DimDate	Success. 2,556 rows transferred.							
\bigcirc	dbo_DimEntity	Success. 423 rows transferred.							
\bigcirc	dbo_DimProduct	Success. 2,517 rows transferred.							
\bigcirc	dbo_DimProductSubcategory	Success. 44 rows transferred.							
	dbo_FactSales	Success. 3,962,476 rows transferred.							
		Stop Import	Close						

Figure 4-22. Table Import Wizard showing number of rows imported

CHAPTER 5

Using SQL in Excel

This chapter covers using SQL (Structured Query Language) in Excel to import and manipulate subsets of large data files and to extract summary statistics.

History of SQL

SQL was first developed by IBM in the early 1970s and was called SEQUEL (Structured English Query Language). It was used to manipulate and retrieve data from early IBM relational database management systems (RDBMS). Because of trademark conflicts, SEQUEL was shortened to SQL. It was later implemented in other relational database management systems by Oracle and later Microsoft. In 1986, the American National Standards Institute (ANSI) issued an SQL standard. In my opinion, the correct pronunciation is S-Q-L, but some people still pronounce it Sequel.

SQL is a declarative language, meaning that it specifies what results are wanted without defining the how. SQL consists of a Data Definition Language and a Data Manipulation Language. In Excel, mostly the query feature of the Data Manipulation Language is used, because the data definition is performed through Excel, by importing tables from other sources. SQL was originally based on the relational model, meaning that data is stored in multiple tables, based on a fixed structure.

NoSQL

As discussed in Chapter 1, the term *NoSQL* is a misnomer, because some NoSQL systems have an SQL interface. A more accurate term would be *non-relational*, or *not only SQL*. Relational databases have a fixed schema that must be defined before any data can be entered. By contrast, NoSQL databases can access data with different schemas. This is sometimes called schema on read.

NewSQL

NewSQL tries to be a compromise between relational and NoSQL systems. These systems use a relational architecture and have SQL as their primary interface but aspire to have the same scalability and flexibility as NoSQL systems.

SQL++

SQL++ is a new form of the language, which enhances traditional SQL and is based on a semi-structured data model that can access JSON data as well as traditional relational data. It is generally backward compatible with traditional SQL.

One implementation of SQL++ is N1QL (pronounced "Nickel"), which is being developed by Couchbase based on a 2014 paper by Kian Win Ong, Yannis Papakonstantinou, and Romain Vernoux entitled "The SQL++ Unifying Semi-structured Query Language, and an Expressiveness Benchmark of SQL-on-Hadoop, NoSQL and NewSQL Databases."

SQL Syntax

The mainstay of the SQL query language is the SELECT command. The syntax of that command is shown following. Note that clauses enclosed in square brackets are optional.

```
SELECT [DISTINCT] <columns list>
[AS <optional alias name for returned columns>]
FROM 
[WHERE <condition>]
[GROUP BY <column expression for calculating aggregates>]
[HAVING <group level filter condition>]
[ORDER BY <sort order for the results table>]
```

The SELECT clause names the columns or column expressions to be shown. An asterisk can be used to specify all columns. The FROM clause lists the table or tables to be accessed. The WHERE clause is used to provide a criteria for relating multiple tables and/or provides a row-level filter, which is used to eliminate rows that do not meet the desired condition. GROUP BY is used to arrange the results table in groups for data aggregation such as subtotaling. The HAVING clause provides a group-level filter for the results set. The ORDER BY clause is used to arrange the results table in the specified order. The methodology for doing this in PowerPivot and Excel will be demonstrated in this chapter.

When using the SELECT statement, the following questions should be addressed:

SELECT: What are the desired columns or columns expressions to be displayed?

FROM: In what tables are these columns?

WHERE: What are the conditions that link the tables? What are the selection criteria to eliminate unwanted rows?

GROUP BY: What values are to be aggregated for such purposes as subtotals?

HAVING: What conditions must each record meet to be included in the result table when using aggregation?

ORDER BY: In what order should the results be presented?

For example, the following SQL SELECT statement extracts the name and address fields from the addrlist table in which the zip code is equal to "94704."

SELECT name, address FROM addrlist WHERE zip="94704"

SQL Aggregate Functions

SQL includes various aggregate functions to allow calculating subtotals, subaverages, subcounts, etc., based on the condition in the WHERE clause.

These functions include:

COUNT(): Returns the number of non-empty rows for the conditions specified

SUM(): Returns the sum of all values in the specified columns

MIN(): Returns the lowest value in the specified column

MAX(): Returns the largest value in the specified column

AVG(): Returns the average value for the specified column

The functions can be applied to the entire table or subsets of the table, as defined by the WHERE conditions. Look at these statements, for example:

SELECT COUNT(*) FROM offices: Would return the number of rows in the offices table

SELECT COUNT(*) FROM office WHERE region="CH": Would return the number of offices in the Chicago region

 ${\tt SELECT}\ {\tt SUM}({\tt sales})\ {\tt FROM}\ {\tt offices}: {\tt Would}\ {\tt return}\ {\tt the}\ {\tt sum}\ {\tt of}\ {\tt sales}\ {\tt for}\ {\tt all}\ {\tt rows}\ {\tt in}\ {\tt the}\ {\tt offices}\ {\tt table}$

SELECT SUM(sales) FROM offices where region="CH": Would return the sum of the sales field for all offices in the Chicago region

Subtotals

It would be possible to display a count of all offices in each region, using the following statement:

```
SELECT region, COUNT(*) FROM offices GROUP BY region
```

Subtotals of sales for all regions could be generated using the following statement:

```
SELECT region, SUM(sales) FROM offices GROUP BY region
```

Joining Tables

When querying multiple tables, it is best to use a WHERE clause to specify the condition for joining the tables. This is called an equijoin, because it is based on an equality test. If the WHERE clause is omitted, every row in the first table will be paired with every row in the second table. This is called the Cartesian product, or outer join of the two tables.

Importing an External Database

This example will use the Access Northwind sample database. To execute this sample, take the following steps:

- 1. Click the PowerPivot tab to open the ribbon.
- 2. Click the Manage button.
- **3.** On the PowerPivot tab, click "Get External Data" and "From Database." Select "From Access."
- 4. Click Browse and select Database2, which was previously downloaded from the Northwind Database site, as shown in Figure 5-1. This file is also available for download from the Source Code/Downloads tab on the Apress book page at www.apress.com/9781484205303.

	Table Import Wizard	? ×
Connect to a Microsoft Acce	ss Database	
Enter the information required	to connect to the Microsoft Access database.	
Friendly connection name:	Access Database2	
Database name:	C:\ leare\Neil\Documents\Database2 accdh	Browse
Database name.	C. (USEIS Wein DOCUMENTS (Databasez.accub	DIO <u>W</u> 36
Log on to the database		
User name:		
Password:		
Save my	password	
	Advanced	Test Connection
	_	
	< <u>B</u> ack <u>N</u> ext > <u>F</u> inish	Cancel

Figure 5-1. Importing database with Table Import Wizard

- 5. Click Test Connection. If the connection tests OK, click Next.
- 6. Select the second option, "Write a query that will specify the data to import," as shown in Figure 5-2. Click Next.

Table Import Wizard	?	×
Choose How to Import the Data		
You can either import all of the data from tables or views that you specify, or you can write a query using SQL that specifies the data to import.		
 Select from a list of tables and views to choose the data to import Write a query that will specify the data to import 		
< <u>B</u> ack <u>N</u> ext > <u>F</u> inish (ance	el

Figure 5-2. Select "Write a query that will specify the data to import"

Before writing a query, it is helpful to know the structure and relationships defined in the database. When importing an Access database, the relationships defined in Access are imported, as shown in Figure 5-3.

CHAPTER 5 USING SQL IN EXCEL

	Ма	nage Relationships	? ×
Status	Table 🔺	Related Lookup Table	<u>N</u> ew
Active	Employee Privileges (Employee ID)	Employees (ID)	Edit
Active	Invoices (Order ID)	Orders (Order ID)	<u>c</u> on
Active	Order Details (Order ID)	Orders (Order ID)	Activate
Active	Order Details (Product ID)	Products (ID)	
Active	Order Details (Status ID)	Order Details Status (Status ID)	Deactivate
Active	Orders (Employee ID)	Employees (ID)	Delete
			Close

Figure 5-3. Existing relationships

7. The first query will import all fields from the orders and employees tables, based on the relationship shown in Figure 5-3, where the EmployeeId field in the Orders table is linked to the ID field in the Employees table. In the SQL Statement window, enter the following SQL Statement:

```
SELECT * from orders, employees
```

8. Then click Validate. The window will appear as shown in Figure 5-4, if the SQL is entered correctly.

Table Import Wizard ? 🗙							
Specify a SQL Query							
Type or paste a SQL query to select data to import from the source database.							
Friendly Query Name: Query							
SQL Statement:							
Select * from orders, employees							
Validate Design							
The SQL statement is valid.							
< <u>B</u> ack <u>N</u> ext > <u>Finish</u> Cancel							

Figure 5-4. SQL select statement

9. Click Finish to see the results, which are shown in Figure 5-5. Click Close to access the data.

			Table Import Wizard	? ×
Imp	ortin	g		
The click	impo the	rt operation might take Stop Import button.	several minutes to complete. To stop the import opera	ation,
	•	Success	Total: 1 Success: 1	Cancelled: 0 Error: 0
	Deta	ils:		
		Work Item	Status	Message
	\bigcirc	Query	Success. 432 rows transferred.	
			<u>Stop</u> Import	Close

Figure 5-5. Success screen showing 432 rows imported

The actual data retrieved after clicking Close is shown in Figure 5-6. Note that 432 rows were imported, because no relationship was specified between the tables, so every row in the first table was paired with every row in the second table. Note that 432 is the product of 9 rows in the employees table and 48 rows in the orders table.

CHAPTER 5 USING SQL IN EXCEL

(1回) 등 5· · · · · · · · · · · · · · · · · ·		PowerPivot for Excel - B	ook13	- 🗆 🗙
File Home Design Ad	lvanced			^ 📀
Paste B Clipboard	PivotTable Format : Text • \$ - % 2 300 400 Formatting	2↓	∑ AutoSum • Create KPI Calculations Calculations Calculations	w n Area
[Ship Address] - 7	89 29th Street			*
🚄 Order ID 💽 Employee ID	💌 Customer ID 💌 Order D	ate 💌 Shipped Date 💌 Sh	nipper ID 💌 Ship Name 💌 Ship A	ddress 💌 Ship City 🔽 ^
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
61	9 4 4/7/200	6 12:0 4/7/2006 12:00:	3 Christina Lee 123 4th	Street New York
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
62	3 29 4/12/20	06 12: 4/12/2006 12:00	2 Soo Jung Lee 789 29	th Street Denver
64	8 6 5/9/200	6 12·0 5/9/2006 12·00·	2 Francisco Dár 123.6th	Street Milwaukee Y
				^
<				, [*]
Query				
Record: I 123 of 432	н			:

Figure 5-6. Results table from SQL query

Specifying a JOIN Condition and Selected Fields

In this example, a join condition will be specified, so that only one order is imported for each employee. This example will show how to import only selected fields. Note that fields with spaces in them must to be enclosed in square brackets. When fields from multiple tables are imported, the field names are specified as tablename.fieldname.

For this example, follow steps 1-6 from the example in the preceding "Importing an External Database" section and then complete the following steps:

Enter the SQL statement, as shown in Figure 5-7. Note that the WHERE condition is specified, linking the two tables based on employees.id=orders.[employee id]. The employee last name will be displayed from the employees table. The employee ID will be displayed from both the employees and orders tables, to demonstrate that they are the same.

Select employees.id, employees.[last name],orders.[employee id] from employees, orders where employees.id=orders.[employee id]

CHAPTER 5 🔲 USING SQL IN EXCEL

	Table Imp	ort Wizard		?	×
Specify a SQL Query					
Type or paste a SQL query	to select data to import fro	om the source da	tabase.		
Friendly <u>Q</u> uery Name:	Query				
SQL Statement:					
select employees.id.emp employees.id=orders.[en	oyees.[last name],orders.[e	nployee id] from e	mployees, orders w	/here	< >
			<u>V</u> alidate	<u>D</u> esign	
The SQL statement is va	lid.				
	< <u>B</u> ack	<u>N</u> ext >	<u>F</u> inish	Cance	

Figure 5-7. SQL query

- 2. Click Validate to ensure that the query is valid, as also shown in Figure 5-7.
- **3.** Now, click Finish. Note that because the condition is specified, only 48 rows (the number of records in the orders table) were imported, as shown in Figure 5-8.

			Table Import Wizard		?	×
Impor	ting					
The im click th	he Sto	peration might take seven p Import button.	eral minutes to complete. To stop t	he import operat	tion,	
		Cancelled: 0	-			
D	etails			5000055.1	End. 0	
	١	Vork Item	Status		Message	
	a	uery	Success. 48 rows transferred.			
				<u>S</u> top Import	Close]

Figure 5-8. Success screen showing 48 rows imported

4. Click Close to complete the procedure. The results are given in Figure 5-9, which shows the employee last name from the employees table and the employee ID from both the employees and orders tables.

CHAPTER 5 USING SQL IN EXCEL

Image: Image													
File Home Design Advanced					^ 0								
Paste Clipboard	ata Type : * prmat : * 5 * % \$ 200 \$00 \$	Clear All Sort by Filters Column	Find Calculations	Diagram View Show Hidden View View									
					*								
📶 id 💽 last name 💽 employee id 💽	Add Column				^								
9 Hellung-Lar 9													
3 Kotas 3													
4 Sergienko 4													
6 Neipper 6													
9 Hellung-Lar 9													
3 Kotas 3													
4 Sergienko 4													
8 Giussani 8													
9 Hellung-Lar 9													
3 Kotas 3													
4 Sergienko 4													
1 Freehafer 1													
1 Freehafer 1													
1 Freehafer 1													
1 Freehafer 1													
1 Freehafer 1													
7 Zare 7													
6 Neipper 6					~								
					^								
					~								
Query													
Record: 14 4 9 of 48 > >1													
					Local - Lawy								

Figure 5-9. Result table from SQL query

Using SQL to Extract Summary Statistics

SQL can be used to extract summary statistics from a table or tables using aggregation functions. This example will extract a single number—the sum of unit price times quantity from the orders and orders detail tables.

For this example, follow steps 1-6 from the previous example in the "Importing an External Database" section.

1. Enter the following SQL statement, as shown in Figure 5-10, to extract the sum of price time's quantity for all records in the orders table.

```
Select sum([order details].[unit price] *[order details].quantity) from
orders,[order details]
```

Table Import Wizard	?	×
Specify a SQL Query		
Type or paste a SQL query to select data to import from the source database.		
Friendly Query Name: Query		
SQL Statement:		
Select sum([order details].[unit price] *[order details].quantity) from orders.[order details]		~
Validate	<u>D</u> esigr	1
The SQL statement is valid.		
< <u>B</u> ack <u>N</u> ext > <u>F</u> inish	Cance	el

Figure 5-10. SQL query

2. Press Enter and wait until the Success screen is shown. Then click Close. The result is shown in Figure 5-11. Notice that the result is a single number, which is the sum of price times quantity for all order detail records.

CHAPTER 5 USING SQL IN EXCEL

1 2	T T	ŧ I		PowerPivot for Excel - Book15												
F F Paste	Get External Data +	Refresh	Advanced A I I PivotTable	Data Type : Decimal Number ♥ Format : General ♥ \$ ♥ % ⊅ 號 🐝 Formatting	Clear All Sort by Filters Column - Sort and Filter	Find	∑ AutoSum ◄ ☐ Create KPI Calculations	Data View View View								
[Expr100	0] 🗸															
Expr100	00 💌 Add	Column														
32	70576															

Figure 5-11. Results is a single number

Generating a Report of Total Order Value by Employee

In this example, a report will be generated of total order value by employee. To create this report, three tables will be linked: employees, orders, and order details.

For this example, follow steps 1-6 from the previous example in the "Importing an External Database" section.

1. Enter the following SQL statements, as shown in Figure 5-12, to extract the sum of price times quantity for all orders broken down by employee. Click Validate to make sure that the SQL statement is entered correctly.

```
Select employees.[last name], sum([order details].[unit price] * [order details].
quantity) from employees, orders, [order details] where employees.id=orders.
[employee id] and orders.[order id]=[order details].[order id] group by
employees.[last name]
```

Table Import Wizard ?	×									
Specify a SQL Query										
Type or paste a SQL query to select data to import from the source database.										
Friendly Query Name: Query										
SQL Statement:										
select employees.[last name], sum([order details].[unit price]*[order details].quantity) from employees, orders,[order details] where employees.id=orders.[employee id] and orders.[order id]=[order details]. [order id] group by employees.[last name]	~									
<u>V</u> alidate <u>D</u> esign										
The SQL statement is valid.										
< <u>Back</u> <u>N</u> ext > <u>Finish</u> Cancel										

Figure 5-12. SQL query

2. Note that two relationships are used to link three tables, based on the relationships shown in Figure 5-3. No fields are displayed from the orders table, but that table is used as a link table. The following are the relationships from Figure 5-3 that are being used in this example. Click Finish to execute the SQL statement.

```
employees.id=orders.[employee id]
orders.[order id]=[order details].[order id]
```

3. This query generates eight rows, as shown in Figure 5-13.

		Table Import Wizard		?	×
mportin	g				
he impo ick the	ort operation might take s Stop Import button.	several minutes to complete. To stop the imp	ort opera	tion,	
			Total: 1	Cancelled: 0	-
	Success	Su	rolai. 1	Error: 0	
Det	ails:			End. 0	
	Work Item	Status		Message	
	Query	Success. 8 rows transferred.			1
		<u>S</u> top I	mport	Close]

Figure 5-13. Success screen showing eight rows generated by this query

4. The actual data generated is shown in Figure 5-14—a listing of the value of orders by employees.

🌒 💷 🕤 - d	>		PowerPivot for Excel - Book1												
File Home	Design Advan	ced													
Paste Get Exte Data	Paste Get External Refresh PivotTable		e : Text • Text • 5 0.00 00 00 000	2 2 2 2 2 2 2 2 2 2 2 2 2 2	Find Create KP		Diagram View Data View View View								
[last name] •			in the second se	Jorrand		concontrollio									
a last name	Expr1001	dd Column													
Cencini	2997.5														
Freehafer	22255.25														
Giussani	680														
Hellung-Lar	19974.25														
Kotas	5787.5														
Neipper	6378														
Sergienko	6278														
Zare	3786.5														

Figure 5-14. Total value of orders by employee

Using MSQuery

Another alternative is to use SQL with Microsoft Query. This example will show another way to access the Northwind database used earlier in this chapter. Follow these steps:

- 1. Click the Data tab to view the ribbon and then click "From Other Sources."
- 2. Then click the last item on the menu, "From Microsoft Query," as shown in Figure 5-15.

	5	· ¢	- 💷 -								B	look1 - E	ixcel				ų. ų				? 1
FILE	- F	IOME	INSERT	PAGE LAYOU	FORMULAS	DAT	A	REVIEW	VIE	W POW	ER QUE	RY	POWE	RPIVOT	Team						
From	From	From	From Oth Sources	Existing Connections	Refresh All + Redit	erties Links	⊉↓ ×↓	Z A Z Sort	Filter	Clear Reapply Advance	ed Co	Text to	Flash Fill	Remove Duplicates	Data Validation	Consolidat	e What-If Analysis	Relationships	Group	Ungroup Sul	* Show Hide
		Get E		From SQL Server											Data	fools				Out	ine
				Create a connection	on to a SQL Server	table. Imp	ort d	ata into E	xcel as a												
C1				From Analysis Ser Create a connection into Excel as a Tab	vices on to a SQL Server le or PivotTable re	Analysis S port.	ervio	es cube. le	mport da	ita											
1	Α	B	3000.	From Windows A	zure Marketplace						J.		ĸ	1	M	N	0	P	Q	R	S
1			Ē	Create a connection Import data into E	on to a Microsoft V xcel as a Table or P	Vindows A VivotTable	zure	DataMark rt.	ket Feed.												
2				From OData Data	Feed																
3				Create a connection Table or PivotTable	on to an OData Dat e report.	a Feed. In	port	data into	Excel as	a											
5				From XML Data In	mport																
6				Open or map a XN	AL file into Excel.																
7				From Data Conne	ction Miland																
8			12	Import data for an	unlisted format b	using the	e Dat	a Connec	tion Wiza	ard											
9				and OLEDB.																	
10				From Microsoft O	uery																
11		_		Import data for an and ODBC Function	unlisted format b	y using the	e Mic ihilits	rosoft Qu	ery Wiza	ird											
12				and object once	onancy is miniced in	n compar	ionity	in previo	NUT YEI JIG	/1124											
13																					
14																					
15																					

Figure 5-15. Screen showing how to access Microsoft Query
3. Select MS Access Database in the Choose Data Source window, as shown in Figure 5-16. Uncheck "Use the Query Wizard to create/edit queries" at the bottom of the window. Then click OK.

Choose Data Source	e ×
Databases Queries OLAP Cubes	OK
<new data="" source=""> Excel Files*</new>	Cancel
MS Access Database*	<u>B</u> rowse
	Options
	<u>D</u> elete
Image: Use the Query Wizard to create/edit querie	s

Figure 5-16. Choose Data Source window

4. Select Database2, which was downloaded previously, as shown in Figure 5-17, and click OK.

Database N <u>a</u> me	Directories:	OK
Database2.accdb	c:\users\neil\documents	<u> </u>
Database1.accdb	C:\ ^	Cancel
Database2.accdb	🗁 Users	Help
	Neil	
	Documents	Read Only
	Custom Office 1	Exclusive
List Files of Type:	Drives:	
Access Databases (*.m.	• 🖃 c: 🔹	Network

Figure 5-17. Select Database window

5. Use the Add Tables window to add the employees, orders, and order detail tables by double-clicking each table, as shown in Figure 5-18.

CHAPTER 5 ■ USING SQL IN EXCEL

FLE H	· c ² · 🔄	RT PAGELAYOUT FO	RMULAS DATA REVIEW	W VIEW PO	Book1 - Excel WER QUERY POWE	8PIVOT Team	-111] + □	Ungroup	7 1 5 1 5 1 1 1 1 1 1 1 1 1 1 1 1 1	Nei how Detail fide Detail	//×
		P 191	ΞΣ 24 24 1 (1) [2	D				0	utline	r <u>a</u>	^
G2 A 1 2 3 4 5 6 7 8 9 10 11 12	Enployee * detess Attachments Buniness Ph Company	es h Dider Details Deter Afscated Dider Afscated Dider Dider Dider Dider Afscated Dider Dider Dider Dider Afscated Dider Dider Dider Afscated Dider Dider Dider Dider Dider Afscated Dider Dider	Query from	MS Table: Dide: Detail Order Detail Order Picto Order State Order State Order State Order State Order State Order State Order State Order State	Add Tables	Add Close Upsions.		R	5	T	V •
14	H A Record	F H									
16											
17	Create a new g	wery									_
19											
20											
< >	Sheet1	(+)		_		1 4		 			•
READY											 100%

Figure 5-18. Microsoft Query window to add tables

6. Click Close and expand the window where the tables are displayed, to see the relationships that have been defined, as shown in Figure 5-19.

\$	Microsoft Query	- 🗆 🗙								
File Edit View Format Table Criteria Records Window Help	File <u>E</u> dit <u>View</u> Forma <u>t</u> Ta <u>b</u> le <u>Criteria Records Window</u> <u>H</u> elp Follow: 대하 도쿄 중 오 도 시간 이 전 10 (11) (11)									
Po Query from	MS Access Database									
Employees Address Order Details Orders Address * Customer ID Attachments * Customer ID Business Phi Cay ID Order Details Company Order ID Product ID Order ID Product ID Product ID Product ID Product ID Paid Date Paid Date Paid Date Paid Date Paid Date Paid Date Paid Date Porter ID Unit Price Ship Cdry Job Tile Unit Price Ship Cdry State/Provin Ship Cay Ship Cay										
Add fields from the field lists to the Data pane										

Figure 5-19. Query screen showing tables and relationships

7. To enter an SQL query, click the SQL button under the Menu bar and enter the SQL query, as shown in Figure 5-20.

Select sum([order details].[unit price] *[order details].quantity) from
orders,[order details]

	5	· () ·						Book2 - E	cel						? [3]
FILE	н	MC	INSERT PAGE LA	YOUT FORMULAS	DATA	REVIEW VI	EW POWER	QUERY R	POWERPIVOT	Team					
From	From	From	From Other Sources •	g Refresh All - Conr	ections 24 (erties 24	Sort Filter	To Reapply Advanced	Text to Columns	Flash Remo Fill Duplica	we Data	Consolidate	What-If Relationships	Group Ungr	oup Subtota	*∃ Show De =∃ Hide Det
		Get Ext	8				Micros	oft Query					×	Outline	
A1		* 1	File Edit View	ormat Table Criteri	Records W	indow Help									
				el 🛲 🖓 🗣 🔽	=Σ AJA			SOL			×				
1 2 3 4 5	A	В	Employees Address Attachments Business Phon City	Dister Details * Date Allocated Discount ID Inventoy ID	Orders Customer ID Employee ID Notes Order Date	SQL statement select sum[or orders, [order	: der details][unit pri details]	ce]"[order deta	als) quantity) from	n ^ Cr	DK ncel				S T
6 7 8 9 10 11 12 13 14			Company Country/Region E-mail Address Fax Number Fax Name Home Phone ID Job Täle Last Name Mobile Phone Notes	Order ID Product ID Purchase Order ID Quantly Status ID Unit Price	Deder ID Paid Date Agymert Type Ship Address Ship City Ship Country/I Ship State/Prix Ship State/Prix Ship State/Prix Shipped Date Shipper ID Shippin Dete					v					
15															
16			WI dia di	1. Ind									_		
18			In I Hecord	(F)									_		
19															
20			View/edit the SQL sta	tement directly							F				
21															

Figure 5-20. Query screen showing SQL query

8. Click OK for the prompt "SQL Query can't be represented graphically. Continue anyway?", as shown in Figure 5-21.



Figure 5-21. Microsoft Query pop-up window

CHAPTER 5 ■ USING SQL IN EXCEL

The single-number response will be displayed as shown in Figure 5-22.

8	Microsoft Query	
Eile Edit ⊻iew Format Table <u>C</u> riteria <u>R</u> ecords <u>Wi</u>	indow Help	
Po	Query from MS Access Database	5
Expr1000 3270576.0		

Figure 5-22. Results of query

Summary

This chapter has shown several ways in which to use SQL with Excel and PowerPivot to summarize data from a large table or set of tables. It has demonstrated how to join tables based on specified conditions and how to use aggregation to extract summary statistics.

CHAPTER 6

Designing Reports with Power View

Power View is a reporting tool that creates its own type of worksheet. For users who are not otherwise proficient in Excel, it is a way of creating a "sandbox" in which to explore a subset of the data without changing it. Power View allows bringing data to life by creating dashboards containing quick and easy visualizations based on PowerPivot models or PivotTables. It is possible to create multiple Power View worksheets based on the same data.

Elements of the Power View Design Screen

The Power View screen consists of the following elements, as shown in Figure 6-1:

- 1. A design surface on the left
- 2. A filters area in the middle, where data can be filtered based on field values
- **3.** The Power View Fields pane, where the fields to be included in the report are selected. The top part includes all the available fields, and the bottom part shows the selected fields.

FILE HOME FILE HOME Copy Paste Clipboard	C - I IX INSERT	PAGE LAYOUT FORMU A Font * A Text Size * Themes Themes	LAS DATA REVIEW VIE Set Image Transparency Image Position	gdpdata fro W POWER VIEW C C C C C C C C C C C C C C C C C C C	POWER QUERY	POWERPIVOT	Team Picture Arrange	? 10
		Click To built in the f	here to add a title	2		Filters VEW To filter the view, dr field fut,	ag fields from the	Power View Fields Active ALL DIT RealGDPByStatePerCapital Drag fields between areas below: PIELDS

Figure 6-1. Power View design surface

Considerations When Using Power View

Always start with a table view, and then base the chart on the table. It is possible to add multiple tables and/ or charts to the design surface. Simply click in a blank area to add a new chart.

Include on the design surface only fields that are necessary for the report. Unless they are relevant to the report, it is generally a good idea to not to include the key values that are used to link tables.

Types of Fields

There are generally two types of fields that are treated very differently.

- 1. Category or descriptive fields that are nonnumeric
- 2. Numeric fields that can include calculated fields or measures. The summation symbol is used to indicate a numeric field whose aggregation type can be changed.

Understanding How Data Is Summarized

Perhaps the trickiest part of working with Power View is understanding how data is summarized. By default, numeric fields are summed, even if they are fields in which the sum makes no sense, such as ID number or year. The type of aggregation can be easily changed by clicking the down arrow after the field name in the lower part of the Power View Field pane. The aggregation choices are as follows:

- Do Not Summarize
- Sum
- Average
- Minimum

- Maximum
- Count(Not Blank)
- Count(Distinct)

Note that for data to be plotted on a chart, the field must be summarized using one of the preceding mathematical functions.

A Single Table Example

This first example will show how to use Power View based on a single spreadsheet of the real GDP (gross domestic product) per capita data by state for each year from 1997 to 2011. The instructions to download this data from the Azure Marketplace were covered in Chapter 3. To create the Power View Query, perform the following steps:

- 🛯 🖯 ち・ぐ・ 🗏 泳: odpdata from azure - Excel PIVOTTABLE TOOLS ? 🗐 A X FILE HOME INSERT REVIEW VIEW POWER QUERY PAGE LAYOUT FORMULAS DATA POWERPIVOT DESIGN Team ANALYZE X Cut 闄 ∑ AutoSum • 11 • A A = = = - - - - - - - - Wrap Text Calibri Fill -Copy . Conditional Format as Cell Paste Insert Delete Format Sort & Find & S Format Painter ℓ Clear * Formatting * Table * Styles Filter - Select -Clipboard Font Number Styles Alignment Cells Editing $- : X \checkmark f_x$ Sum of RealGDPPerCapital A1 A R D F G н K М Ν C E R S 1 Sum of RealGDPPerCapital Column Labels * Row Labels 2011 Grand Total 3 Alabama 4 Alaska 55892 54362 56322 55949 57955 5 Arizona 29917 31745 34736 35267 36315 36725 6 Arkansas 7 California 8 Colorado 39202 40576 45101 45590 45298 45778 47599 47659 9 Connecticut 49162 50316 54302 54370 59613 58874 56254 56797 59318 61751 63568 10 Delaware 11 District of Columbia 116169 117449 122096 128456 132630 136964 142535 146063 147172 149168 12 Far West 37063 38707 42844 42185 42535 43293 44857 46400 47747 48240 47773 45604 45479 13 Florida 34198 34414 34912 35848 36805 36520 37634 40076 39875 39705 14 Georgia 39793 40229 15 Great Lakes 16 Hawaii 39701 38778 39385 39966 41139 42542 17 Idaho 25200 25806 30332 29768 30028 30267 32138 34224 33553 31984 32559 18 Illinois 41961 43198 43145 43171 43941 44906 45166 46105 19 Indiana 20 Jowa 33444 33595 35186 36151 37420 40037 42439 41451 40831 41365 21 Kansas 33595 34593 35344 36347 36717 36908 37734 37645 38307 39390 40958 41059 39440 39438 39484 00386 00366 -----33440. 33303 33446 33346 40700/ (+) Real GDP Per Capita By State (1 1 4
- 1. Open the GDP data from the Azure Spreadsheet, which is shown in Figure 6-2.

Figure 6-2. Spreadsheet for real data by GDP per capita by state

2. Click the Insert tab and select Power View Reports, then wait while the Power View screen loads, as shown in Figure 6-2. If prompted to install SilverLight, click Yes. Note the design surface on the left, the Power View Fields pane on the right, and the Filters pane in the middle.

3. In the Power View Fields pane on the right side of the screen, click the right-facing triangle to the left of RealGDPByStatePerCapita to display the fields. Then drag the year field to the Filters pane, as shown in Figure 6-3.



Figure 6-3. Setting a filter

4. To display only the most recent year, 2011, click the leftmost icon to the right of Year in the Filters pane to activate the List Filter Mode to see the list of years and select 2011, as shown in Figure 6-4.



Figure 6-4. Setting a filter, expanded view

5. Then, in the Power View Fields pane, click the box in front of Area and RealGDPPerCapita, and a table will be created, as shown in Figure 6-5.

FALE HOME INSERT PAGE LAYOUT FORMULAS DATA RE Table Bar Column Other Map Tites Tite Silicer Card Silicer Card Silicer Silicer Card Silicer Silicer Card Silicer Silicer Card Silicer Silicer <th>gdpdata from azure - Excel</th> <th>POWERPINOT Team</th> <th>? 10 - 9 /</th>	gdpdata from azure - Excel	POWERPINOT Team	? 10 - 9 /
Click here to add	d a title	x > x 4 8 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Power View Fields × ACITYE ALL ■ ReaKOPByStatePerCapital ② C ARLA ② T ReaKOPPRCapital ③ T VIAR Prag fields between areas below: TELE BY FILDS ③ ARLA ∑ ReaKOPPPerCapital ●
Real GDP Per Capita By State (1 Power View1 +	1 4		E E

Figure 6-5. Table showing sum of real GDP per capita

6. Note that by default, the sum of the RealGDPPerCapita is shown, which is meaningless. If you want to change that setting, at the bottom of the Power View Fields pane, click the down arrow after RealGDPPerCapita and select "Do Not Summarize." However, to plot the charts shown following, it is necessary to leave it set to Sum. The results are shown in Figure 6-6. Note the scrollbar to the right of the table, which allows scrolling through all the states. To expand the table, click it and drag the handles.

Click here to add a title	Table Bar Column Other Map - Chart - Chart - Chart - Tipe - Swedt Visualization - Swedt Visualization - Swedt Visualization - Chart - Chart - Tipe - - Title Silee - Chart - Chart - Tipe - - Title Silee - Chart - Chart - Title Silee - Options - Sile - Chart - Title Silee - Options - Options	gdpdata from azure - Excel	POWERPINOT Team	7 00 - 6 S
	Click here to add	a title		Power View Fields × Activit ALL ▲IB ReadSDPByStatePerCapital > ○ # Piris > ○ X Pirs > ○ Area > ○ X RealGOPPerCapital >

Figure 6-6. Table showing real GDP per Capita by state

7. A title can be added by clicking where it says to "Click here to add a title" and entering "Real GDP Per Capita by State for 2011," as shown in Figure 6-7.

	1 B B B	r ÷		gdpdata from	azurepowerview - E	cel			? E - 6 ×
FLE HOM	E INSERT PA	AGE LAYOUT FORMU	LAS DATA REVI	W VIEW POWER VIEW	POWER QUER	POWERPING	OT Team		Neil + 🖸
A Cut Paste	Undo Redo Tr	A Font * A Text Size * hemes Background *	Set Image Tran Image - Position -	sparency Refresh Relationship	s Fit to Field Window List	Filters Area	Text Picture Box	Arrange -	
Clipboard	Undo/Redo	Themes	Background Ima	pe Data	View		insert	Arrange	
	AREA Alabama Alabama Aranan Aranana Califonda Convectour Delenaro Datote of Columbia Far West Florida Georgia	Real GDP Pee ReadCOPPerCeptal \$1301.00 91.05200 31.04200 91.04200 95.04200 95.04200 95.04200 95.04200 95.04200 95.04200 95.04200 95.04200 95.04200 95.045000 95.045000 95.045000 95.045000 95.045000 95.045000 95.0450000 95.045000 95.04500000000000000000000000000000	r Capita by	State for 2011		Filters VIW * 321 * 321 * 321 * 321 * 320		€ X ■ ● X 0 0 0 0 0 0 0 0 0 0 0 0 0	Power View Fields × ACTIVE ALL Immediate Perceptial BARA Immediate Perceptial The Second Perceptial Immediate Percepticon The Second Perceptial
4 9	Real GDP Per Cap	pita By State (1 Pow	er View1 Power Vi	ew2 (+)					•

Figure 6-7. Table with title added

Viewing the Data in Different Ways

Power View supports multiple formats for viewing data. With the Design tab selected, click the Table icon on the left to see the options: table, matrix, and card. The card view, which creates a card for each state, is shown in Figure 6-8. The matrix view is like a table but allows drilling down. It will be demonstrated in later examples. To go back to the table format, simply click Table under the Table icon.

Table Bar Column Other Map - Chart Chart Chart Sileer Switch Visualization - Switch Visualization - Switch Visualization	gdpdata from azu A REVIEW VEW POWER VEW General A A Show Totals Show Totals Number Tet	srepoverview - Excel DESIGN POWER QUERY POWERPHYOT Bring Send ntward - Backward - Arrange	Team	? 10 - 9 × Neil - P
Real GDP Per Capi Addama Adda Add	ta by State for 2011, State of the second s	Filters vetw i TALL * VEW i TALL * 2011 940 940 940 940 940 940 940 940 940 940	x > x 4 E x 4 E 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Power View Fields × ▲ Active ALL ▲ BeadGOPPSStatePerCapital
Real GDP Per Capita By State (1 Power View1	Power View2 (+)	: [4]		Þ

Figure 6-8. Card view

Creating a Bar Chart for a Single Year

To create a bar chart showing the real GDP per capita for each state, follow these steps:

- 1. Click anywhere in the table and select the Design tab. Click Bar Chart. Select Clustered Bar. Because there is a single data source, either clustered or stacked will show the same result.
- 2. The results will appear as shown in Figure 6-9. A single bar for GDP per capita is shown for each state for the year 2011.

FILE HO Paste Copy Paste	ME INSERT	PAGE LAYOUT FORMU A Font * A Treet Size * Background * Themes	LAS DATA REVIEW VI Set Image Transparency Image * Position *	gapasas from az EW POWER VIEW Refresh Relationships Data	POWER QUERY Fit to Field Window List View	Filters Area	IVOT Team	Arrange Arrange	Nei - P
	RealCOPPerCap	Click ital by AREA	here to add a titl	e		Filters VIEW			Power View Fields × ACTIVE ALL Improvement Improvement Improvement Improvement <tr< td=""></tr<>
4 3	Real GDP Per	Capita By State (1 Pow	ver View1			1	d.		Þ



3. To resize the chart, click it and drag the handles.

Note that you can hover the mouse over any bar in the chart to see the GDP per capita value for that state.

Column Chart

A column chart is a bar chart flipped on its side, as shown in Figure 6-10.





106

Displaying Multiple Years

If the filter is changed to include multiple years, by clicking the All box at the top of the Years list in the Filters pane, an instance of the state for each year will be generated in the table, as shown in Figure 6-11. It is also necessary to click Year in the Power View Fields list pane, so that the year is displayed.

Note	🕼 🖯 ち・ぐ・ 泳 🗐・ 🕫		gdpdata from azurejul20.xlsx - Ex	cel		7 8 - 8 3
AKA. YKAA. NadCobe-Capital POWEL VIEW TELLS ACA. YKAA Akabama 1999 2844.00 Akabama 1999 2844.00 Akabama 1999 2844.00 Akabama 1999 2845.00 Akabama 2000 20920.00 Akabama 2000 20920.00 Akabama 2000 20955.00 Akabama 2001 23.60 Akabama 2001 23.64.00 Akabama 2001 23.64.00 Akabama 2001 24.64.00 Akabama 2001 25.66.00 Akabama 2001 54.94.00 Akabab	FILE HOME INSERT PAGE LAW	OUT FORMULAS DATA RVIEW VIEW I I I I IIII IIIII IIIIIII IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	POWER VIEW DESIGN P 	d -	()	Neil + D
Alaka 1999 50 890.00 Alaka 1999 55 892.00 Alaka 2000 54.22.00 Alaka 2002 57.92.00 Alaka 2004 57.95.50 Alaka 2005 59.43.00 Alaka 2005 59.73.00 Alaka 2005 59.90.00 Alaka 2005 59.90.00 Alaka 2006 59.95.70.0 Alaka 2019 61.52.00 Alaka 2019 61.52.00	AEX. TVEIX. TVEIX. <td></td> <td></td> <td>PICES VEW 1 TALES ● VEM 1 TALES</td> <td></td> <td>POWEP VIEW FieldS ▲ ACTIVE ALL ▲ ■ ■ ARA □ ∑ > □ ∑ 10 > □ ∑ ™ > □ ∑ ™ > □ ∑ ™ > □ ∑ ™ > □ ∑ ∑ Next Prog fields between areas below: TLE BY</td>			PICES VEW 1 TALES ● VEM 1 TALES		POWEP VIEW FieldS ▲ ACTIVE ALL ▲ ■ ■ ARA □ ∑ > □ ∑ 10 > □ ∑ ™ > □ ∑ ™ > □ ∑ ™ > □ ∑ ™ > □ ∑ ∑ Next Prog fields between areas below: TLE BY
	Aukia 1999 519.0100 Aukia 1999 519.02.00 Aukia 2000 54.12.00 Aukia 2000 54.02.00 Aukia 2000 55.42.00 Aukia 2001 57.92.00 Aukia 2001 57.94.00 Aukia 2005 55.44.00 Aukia 2006 57.95.00 Aukia 2006 59.900.00 Aukia 2006 59.900.00 Aukia 2006 59.920.00 Aukia 2007 61.323.00					FIELDS ⊕ AREA • YEAR • ∑ RealGDPPerCapital •

Figure 6-11. Table for real GDP per capita for multiple years

If the view is changed to matrix, then it appears as shown in Figure 6-12. Note that it is necessary to tell it not to sum the years, by clicking the arrow after Year in the Fields section of the Power View Fields pane and selecting "Do not summarize."

Image: Solution of the solut	gdpdata from azurepowerview - Exe ATA REVIEW VIEW POWER VIEW DESIGN PC	el WER QUERY POWERPIVOT Team	? 13 - 6 Neil -
Scut Scut All All Text Size* Paste Undo Redo Themes Set Clipboard Undo/Redo Themes Set	Image Transparency Position - Transparency Refresh Relationships Data - Field View	Fiters Area View Box Insert	
Real GDP Per	Capita 2008-2011	Filters C ×	Power View Fields × Active ALL • TRealGDPByStatePerCapital
2009 81.746.00 2010 31.655.00 2011 31.301.00 Total 127.67.00 Alaska 2009 63.252.00 2010 61.340.00 2011 61.453.00		398 3 398 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3 300 3	∑ FIPS ∑ ID Ø∑ RealGDPPerCapital Ø∑ YEAR
Total 245.02.00 Arizona 2009 35.74.00 2010 34.074.00 2011 2011 35.02.00 2011		2006 □ 2009 □ 2009 □ 2009 □ 2009 □ 2001 □	Drag fields between areas below: TILE BY
Total 14.3,55.00 Arianss 2009 31.101.00 2009 30.079.00 31.211.00 2011 31.142.00 31.142.00			∑ RealGDPPerCapital ▼ ROWS @ AREA ▼
Tettal 25,013.00 Califor 2008 48,259.00 2009 48,029.00 2010 45,581.00			YEAR COLUMNS V
Power View1 Power View2 Power View3	\oplus	[4]	

Figure 6-12. Matrix view of GDP per capita for multiple years

Adding a Map

To display the data using a map, follow these steps:

- 1. Click the table to activate the Design tab.
- 2. Click the Map icon. If a map of the world is displayed, use the Zoom and Pan icons at the upper right to display only the desired area of the map, as shown in Figure 6-13. Hover the mouse over a bubble to see the actual number associated with it.

PLE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW Table Bar Column Other Map Table Sticer Card Show Totals Sticer Sticer Sticer Card Store Card Store Table Sticer Sticer<	gdpdata from azurepowerview - Excel VIEW POWER VIEW DESIGN LAYOUT POWER eneral - 196 - 100 - 20 Forward - Backward - Number Text Arrange	QUERY POWERPIVOT Team	? 🗹 - /8 <u>A</u> Nei -
Click here to a	dd a title	▼ L2 VEW MAA VEW MAA 2011 ● (¥201) ● (¥97) ● (¥99) ● (1997) ● (¥99) ● (1997) ● (1	Power View Fields × ACTIVE ALL → IT RealGDPByStatePerCapital → IT RealGDPByCapital → IT RealGDPBrCapital → IT Real
Power View1 Power View2	: [t]		

Figure 6-13. Map view showing GDP per capita by state

Note that bubbles are displayed for each state proportional to the GDP per capita for that state. In the Filters pane, individual or multiple years can be selected to see the data for those years.

Using Tiles

Another way of viewing the data is by using tiles. Simply select a table and click the Tiles icon on the Design toolbar to see a display, as shown in Figure 6-14. The default layout has a tab strip at the top of the display that allows clicking each state to see the values for that state.

RE HOME FLE HOME Table Bar Column - Chart - Chart- Switch Visualiz AREA RealG Arizona	INSERT PAGE LAYOUT FORM	ARUAS DATA REVIEW	yiew Power view General A Bumber Test y State for 2 'olorado Connecti	repowerview - Escel DESIGN POWER QUERY Dring Send Arrange 20008-20111 Delaware Distrik Colu	POWERPHOT	Filters VIEW 9 VEAR 16 2011, 2010, 20	Power View Fields ACTIVE ALL → ■ ReadOPByStatePerCapital → ■ ReadOPByStatePerCapital → ■ FIPS → ∑: P0 → ∑: ReadOPPerCapital → ∑: YEAR Diag fields between areas below: TLE BY ● AREA FIELDS	×
4 + Pow	rer View1 Power View2	-			-			•

Figure 6-14. Tile format

An alternative is the tile flow format with the tiles at the bottom as shown in Figure 6-15. These views can be toggled by clicking on the Tile Type icon on the Design toolbar.

C → · · · · · · · · · · · · · · · ·	gdpdata from azurepowerview - Excel EVIEW VIEW POWER VIEW DESIGN POWER QUERY POWERPIVOT T	eam ? 3 - 5 . Neil - 5
Table Bar Column Other Map Table Silcer Cord Soler Sol	Grand	
Real GDP per Capita	by State for 2008-2011	Filters Power View Fields × View → Real@DPByStatePerCapital → weaked@DPbyStatePerCapital → D view → Real@DPbyStatePerCapital → D view →
AREA RealGDPPerCapital YEAR Alaska \$9,057.00 2008 61,040.00 2009 61,853.00 2010 63,252.00 2011 2011 2011		∑ RealGDPPerCapital ∑ YEAR
Ϋ́.	1	Drag fields between areas below: TILE BY @ AREA
_		FIELDS
Alabama	Alaska Arizona Arkansa Californi Colorad Connet Delawar ^{District} s a o icut e columbi	
	:	

Figure 6-15. Tile Flow format

Relational Example

For this example, Access database2 (the Northwind Database) will be used. Some of the imported tables are shown in Figure 6-16. This database can be downloaded from the Apress web site.

mpo	9 ort operation might take seve Stop Import button.	aral minutes to complete. To stop the	import opera	ition,	
•	Success		Total: 44 Success: 44	Cancelle Erro	d: C or: C
Deta	ails:	Queture.		Maaaaaa	
	Order Details Status	Success 6 mws transferred		Message	- 1
0	Orders	Success 48 rows transferred			
0	Orders Status	Success, 4 rows transferred.			Ì
ø	Orders Tax Status	Success. 2 rows transferred.			d
0	Privileges	Success. 1 rows transferred.			
0	Products	Success. 45 rows transferred.			
Ø	Purchase Order Details	Success. 55 rows transferred.			
Ø	Purchase Order Status	Success. 4 rows transferred.			
Ø	Purchase Orders	Success. 28 rows transferred.			
0	Sales Reports	Success. 5 rows transferred.			
Ø	Shippers	Success. 3 rows transferred.			
0	Strings	Success. 62 rows transferred.			
0	Suppliers	Success. 10 rows transferred.			
Ø	Customers Extended	Success. 29 rows transferred.			
-					

Figure 6-16. Northwind tables

Follow the steps described in Chapter 4 to import Access data.

- 1. Click the PowerPivot tab to open the ribbon.
- 2. Click the Manage button at the top-left end of the ribbon.
- **3.** Once the PowerPivot window is open, you can load data by clicking *From Database* in the Get External Data group. This provides a choice of database sources. For this example, *From Access* will be selected.

- 4. In the Table Import Wizard dialog, click the Browse button to find the Access file you want to import. For this example, use Database2.accdb. This file is available from the Apress web site.
- 5. Click *Test Connection*. If the test is successful, click Next.
- 6. PowerPivot opens the Table Import Wizard to take you through the process of loading data. Choose *Select from a list of tables and views to choose the data you want to import.* Click Next.
- 7. Click the top-left box to select all tables. Note that in the lower right corner of the dialog, there is an option to *Preview & Filter*. This option can be used to select a subset of data from very large datasets. In this example, all rows in all tables will be imported. Then click Finish to complete the import.
- 8. The tables imported are shown in Figure 6-16.

The relationships between the tables are shown in Figure 6-17.



Figure 6-17. Diagram view

This example will use the relation between customer ID in the orders table and the ID field in the customers table.

Another view of the relationships is shown in Figure 6-18, which is accessed by clicking the Relationships icon on the Data tab.

	Manag	ge Relationships		? ×
Status	Table 🛎	Related Lookup Table	^	<u>N</u> ew
Active	Employee Privileges (Employee ID)	Employees (ID)		Edit
Active	Employee Privileges (Privilege ID)	Privileges (Privilege ID)		
Active	Inventory Transactions (Customer Order ID)	Orders (Order ID)		<u>A</u> ctivate
Active	Inventory Transactions (Product ID)	Products (ID)		Desetivate
Active	Inventory Transactions (Purchase Order ID)	Purchase Orders (Purchase Order ID)		Deactivate
Active	Inventory transactions (transaction type)	Orders (Order ID)		Delete
Active	Order Details (Order ID)	Orders (Order ID)		
Active	Order Details (Order 10)	Products (ID)		
Active	Order Details (Status ID)	Order Details Status (Status ID)		
Active	Orders (Customer ID)	Customers (ID)		
Active	Orders (Employee ID)	Employees (ID)		
Active	Orders (Shipper ID)	Shippers (ID)		
Active	Orders (Status ID)	Orders Status (Status ID)		
Active	Orders (Tax Status)	Orders Tax Status (ID)		
Active	Purchase Order Details (Inventory ID)	Inventory Transactions (Transaction ID)		
Inactive	Purchase Order Details (Product ID)	Products (ID)		
Inactive	Purchase Order Details (Purchase Order ID)	Purchase Orders (Purchase Order ID)		
Active	Purchase Orders (Created By)	Employees (ID)	~	
				Close

Figure 6-18. Relationships

To create a Power View report, follow these steps:

- 1. Click Power View Reports on the Excel Insert tab.
- 2. First create a table by selecting fields from the Power View Fields tab on the right side of the screen.
- **3.** From the customers table, select the company, first name, and job title fields, as shown in Figure 6-19.

Image: State of the state	VER QUERY POWERPIVOT Team	7 10 7 19 A Nei - F
Click here to add a title	Filters < × vew + trace	Power View Fields × ACTIVE AL Constant of the second seco
Sheet1 Power View1 ⊕	3 (1)	

Figure 6-19. Table showing company, first name, and job title

4. Next, select Price Total from the Order Price Totals table. Doing this depends on the relationship between the customers to orders table to the order price totals table. The results are shown in Figure 6-20.

Image:	chapterreportspowerview - Excel kS DATA REVIEW VIEW POWER VIEW DESIGN POWER	QUERY POWERPIVOT Team	? E = 8 > ▲ Neil - □
Table Bar Column Other Map Tiles Tile Tile Tile Tile Tile Sicer Switch Visualization Tales Sicer Tiles Tiles Sicer	Card Show Totals Style - Levels - Options Number Tetal Arange		
Congany A nota Congany A Ana Congany A Ana Congany A Ana Congany A Ana Congany A Ana Congany A Ana Congany B Anritash Congany B Anritash Congany B Anritash Congany B Anritash Congany B Anritash Congany B Anritash Congany C Thoma Congany C Con Ang Congany C Son A	k here to add a title	Filters • × Wew I TABLE To filter the view, drag fields from the field list.	Power View Fields × ACIVE ALL Imventoy Sold Imventoy Sold Imventoy Transaction I Imventoy Transac
← → Sheet1 Power View2 Power View9	Power View10 Power View8 Power View11 (+)	: [(]	•

Figure 6-20. Table showing order totals

5. Note that for companies with multiple orders, there are multiple rows shown in the table. A matrix view, shown in Figure 6-21, makes this more obvious.

FILE HOME INSERT FILE HOME INSERT Table Bar Column Other • Chart - Chart - Chart Switch Visualization	A A A A A A A A A A A A A A A A A A A	FORMULAS DATA REVIEW	Chapterreportspowerview - VIEW POWER VIEW DESIGN General A A Bring S Forward - Back Number Text	Excel POWER QUERY POWERPINOT T end end oward *	eam ? 13 - 5 Nei - 5
	Company Tint Name Company A Ana Company A Kana Company E Amrianuh Company C Tooman Company C Tooman Company C Soo Jung Company O Constine	Click here No Trie Peer Test Owner 55600 Auronasing Managar 513000 Auronasing Managar 513000 Purchasing Representative 513000 Purchasing Representative 513000 Purchasing Managar 5135000 Purchasing Managar 5135000 Purchasing Managar 5135000 Purchasing Managar 5135000	to add a title	* Pilers	Power View Fields × ACTIVE ALL Last Name Last Name Mobile Phone Notes StateProvince Verb Page ZIPPostal Code Compare Price Totals Control VE BY The BY The BY Compare Price Total
• · · · Power View	11 (+)			1 4	•

Figure 6-21. Matrix view

Customer and City Example

This example will show customers by city. The objective is to show the value of orders from each city. Follow these steps:

- 1. Click Power View Reports on the Excel Insert tab.
- 2. Select company and city from the Customers table, as shown in Figure 6-22. Adjust the width of the company column in the table, if necessary, by selecting the line between columns in the title bar and dragging.

Image: State of the state	VER QUERY POWERPIVOT Team	2 2 5 Nei -
Click here to add a title	Filters C × www track To filter the view, drag fields from the field lat.	Power View Fields ACTIVE ALL ACTIVE ALL Address Business Friend Company Company Company CountryRegion E-mail Address Fall Address Fall Address Fall Address Fields between areas below: TILE BY FIELDS Company Comp
← → Sheet1 Power View2 (+)	: [4]	

Figure 6-22. Company and city table

3. Select the price total field from the order price totals table. Note that a prompt appears to create a relation. Because no relation is created, the total of all orders is shown, as in Figure 6-23.

Element E	mreports - Excel ?
Table Bar Column Other Map - Chart Chart Chart Bar Switch Visualization Table Siter Siter Siter Column Table Siter Chart Show Totals Siter Column Chart Chart Table Siter Chart Table Siter Column Tabl	Bring - Send Forward - Backward - Arrange
Click here to add a title	Filters < × VEW TABLE POwer View Fields ACTVE ALL Power View Fields ACTVE ALL Power View Fields ACTVE ALL Power View Fields Indicate Details Status Indicate Details Indicate Details Status Indicate Details Indicate D
Sheet1 Power View2	: (

Figure 6-23. Table showing company, city, and order price totals

4. Create the relationship between the orders table and the order price totals table based on the order ID, as shown in Figure 6-24, and click OK.

Col <u>u</u> mn (Foreign):
Order ID
Related Co <u>l</u> umn (Primary):
OrderID
R

Figure 6-24. Creating relationship between the orders and order price totals tables

5. Next, click the arrow after Price Total in the Fields list and select "Do Not Summarize," as shown in Figure 6-25. Note that the value of each order is now shown, instead of the total of all orders.

II II ↔ · ↔ · III → · · · · · · · · · · · · · · · ·	chapterreports - Excel	VER QUERY POWERPIVOT Team	? 10 - 6 Nel -
Table Bar Column Other Map • Chart • Chart • Chart • Switch Visualization Tiles Ti	Silcer Card Show Totals Silcer Card Show Totals Silcer Options Number Tet Arange		
Company Cry Pres Total Company A Seattle 174800 Company A Seattle 174800 Company A Seattle 175800 Company AB Memphi Company B Memphi Company B Memphi Company B Memphi Company B Memphi Company C Seattle 17500 Company C Lei Angela 10000 Company C Lei Angela 10000 Company C Denive Company C Denive C Deniv	lick here to add a title	Filters VINV TABLE PO	Wer View Fields >> VE ALL Order Details Order Details Status Order Details Status Order Price Totals S Order Subtotals Order
Sheet1 Power View2		: [4]	

Figure 6-25. Table showing individual order totals

6. Selecting the matrix view by clicking Table on the Design tab and selecting Matrix and clicking "Fit to Window" will result in the display shown in Figure 6-26. Note that a title has been entered. The view can be scrolled using the scrollbars on the right.

1 5 . 6	÷- i×⊺					chapter	eports - Excel					? 3	- 8
FILE HOME	Undo Redo	AGE LAYOUT A Fon A Tom Themes Bac Theme	FORMULAS t * t Size * kground * s	DATA R et Image T ge - Position - Background I	ransparency mage	POWER VIEW	DESIGN Fit to Window View	Filters Area	RY Pi Power 1 View 1 Ir	OWERPIVOT	Arrange		Neil *
-	_		Order	Totals	s by Co	ompan	У				Filters VIEW MAI	Power View Fields	
Company Company A	City Seattle	Price Total \$736.00 \$1.674.75	Î									Order Details Excended Order Details Status Order Price Totals D Order D	
Company BB	Memphis	\$1,505.00 \$230.00	1									∑ Price Total □ □ Order Subtotals □ □ Order Summary	a.
Company C	Los Angelas	\$13.800.00 \$0.00 \$620.00	1								To filter th field list.	Drag fields between areas below: TILE BY	
Company CC	Denver	\$1,930.00 \$127.50 \$1,218.00										ROWS	
Company D	New York	\$1,560.00	ъ									Company City Price Total	
< > \$	Sheet1 Pow	er View2	۲					1	4		P.	COLUMINS	Ψ

Figure 6-26. Matrix view

7. Clicking the pop-out icon at the upper right of the table display will result in the view shown in Figure 6-27.

G → · · · · · · · · · · · · · · · ·	chapterreports - Excel	PIVOT Team	? 🖬 - 8 Neil -
Table Bar Column Other Map • Chart Chart Chart Tiper Switch Visualization Tiles Sile	Card Show Totals Ship'e Levels* * P Options Number		
Company City Price Total		A MOOTO A	Power View Fields ×
Company A Seattle \$736.00		VIEW MAT	ACTIVE ALL
\$1.674.75			
Company AA Las Vegas			D Crder Details
\$1.505.00			D I Order Details Extended
Company BB Memphis \$230.00			D I Order Details Status
\$1.402.50			Order Price Totals
\$13,800.00			∑ OrderID
Company C Los Angelas \$0.00			
\$620.00			Order Subtotals
\$1,930.00			D Corder Summary
Company CC Denver		To filter th	Drag fields between areas below:
\$127.50		field list.	TILE BY
\$1,218.00			
\$1,560.00			Σ VALUES
Company D New York			
\$184.00			
\$360.00			KOWS .
\$665.00			@ City •
S3.520.00			Price Total •
\$127.50		-	COLUMNS
Sheet1 Power View2	1	P.	

Figure 6-27. Pop-out view of the matrix

8. To filter out records that do not have an order, drag the Price Total field to the Filters pane and drag the left bar to the right to some arbitrary small number greater than zero, as shown in Figure 6-28.

1 E 5· e· 🔳	i× □• •	chapterreports - Ex	xel	7 1 - 6
Picc Plowie index Cut Scopy Paste Copy Clipboard Undo/Re	Al Text Size* Al Text Size* Al Text Size* bio Themes Background*	Transparency Refresh Relationships Data	Field Fiters View View Box	
Company City Company A Seattle Company A Seattle Company A Seattle Company B8 Memph Company B8 Memph Company C Los Ang Company CC Denver Company CC Denver Company CC Denver Company CC Denver Company D New Yo Company D New Yo Company D New Yo	Price Total \$736.00 \$1,674.75 as \$1,505.00 is \$1,402.50 is \$1,402.50 is \$1,300.00 \$1,27.50 \$1,27.50 \$1,218.00 \$1,27.50 \$1,218.00 th \$184.00 th \$380.00 th \$380.0	tals by Company	Filt view so	Power View Field.* ACIVE ALL Control ALL Control ALL Control ALL Control Control Control ALL Control Control Control ALL Control Control Control ALL Control Control Cont
Sheet1	Power View2 Power View4 (+)		1 (I)	

Figure 6-28. Filtering out companies with no orders

Or click on the Price Total field name and set the filter as greater than or equal to 1 and click Apply Filter, as shown in Figure 6-29.

Chaptere FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW POWER VIEW	ports - Excel DESIGN POWER QUERY POWERPIVOT Team
Image: Section of the section of th	ports - Excel POWER QUERY POWERPIVOT Team
Sheet1 Power View2 (+)	

Figure 6-29. Another way to filter out companies with no orders

Showing Orders by Employee

The next example will show number of orders by employee, by taking the following steps:

- 1. Create a new Power View sheet by clicking Power View on the Power View tab.
- 2. Enter the title "Orders by Employee."
- **3.** In the Power View Fields pane, select the Last Name field from the Employees table.

Then select the Order ID field from the Orders table. The results are shown in Figure 6-30. Note that by default, it sums the Order IDs, which is a meaningless number.



Figure 6-30. Table showing employees and orders

4. In the Fields area of the Power View Fields pane, click the arrow after Order ID and select *count (non blank)*. The results are shown in Figure 6-31. Note that a count of the orders is now shown in the table.

Image: Image	chapterreports - Excel VIEW POWER VIEW DESIGN POWER QUERY PO	WERPIVOT Team	7 @ - 6 Neil -
A Cut Second A Tort Size Paste Undo Redo Themes Background * Cipboard Undo/Redo Themes Background *	y Refresh Relationships Data View Ins	ert Arrange	
Orders by Employer	e Filters væv i talls	< x Pr	Dwer View Fields × ThVE ALL □ Σ Taxes □ Orders □
Hellung-Justein 10 Kotasa 6 Nelpar 4 Sergiento 8 Zive 2 Total 48	To filter the view, field list.	drag fields from the Dr	Deployee ID Notes Order Date Order Date Z Order ID Paid Date Payment Type v ag fields between areas below: EBY
			LDS Last Name • If Count of Order ID •
Sheet1 Power View2 Power View9 Power View8	: 4		

Figure 6-31. Table showing employees and number of orders per employee

5. To create a chart of the data, right-click the table and select Copy. Click in the blank area next to the table, right-click, and select Paste to get another copy of the table, as shown in Figure 6-32.

🚺 🖥 ち・ぐ・ 💷 🗵 🖬・ ፣		chapterreports - Excel		? E - 8 ×
File HOME INSERT PAGE LAW Image: Second	OUT FORMULAS DATA REV Time Sicer Card Show Tet Style Levels * Options	EW VEW POWERVIEW DESIGN General - A als 9 + 35 #3 #3 Number Text Arrange	POWERQUERY POWERPIVOT Team	Power View Fields
Lant Yama Count of Order D Genich 12 Glosseh 2 Holloy-Lassen 10 Rota 6 Segunta 8 Segunta 8 Zen 2 Tetal 4	Concers by EIIIpi	loyee	VRW YABLE	ACTIVE ALL Last Name Mobile Phone Notes StateProvince Drive Dage Drives Customer ID E Customer ID E Customer ID FREDS Last Name Count of Order ID
4 > Sheet1 Power View2	Power View9 Power View8	(+)	1	•

Figure 6-32. Making a copy of the table prior to plotting

6. To plot the chart, select the rightmost table that was just pasted, click the Bar Chart icon on the Design tab, and select Stacked Bar to get a bar chart, as shown in Figure 6-33.

Chapterreports - Excel Chapterreports - Excel FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW POWERVIEW DESIGN LL	AVOUT POWER QUERY POWERPINOT Tea	- 6
Table Bar Column Other Map 3000 Tiles Tiles <th>d -</th> <th></th>	d -	
Orders by Employee	Filters C ×	Power View Fields
Cencio Courto of Order 10 Cencio 4 Preshafer 12 Gissani 2 Prenafer 10 Kota i 6 Nioper 4 Stergienko 8 Totai 44 O 10	To filter the view, drug fields from the field list.	Last Name Mobile Phone Notes StateProvince Veb Page ZiPPostal Code Code S: Cutomer ID S: Employee ID we
		Drag fields between areas below:
		∑ VALUES # Count of Order ID •
		AX05
		Last Name 🔻
		LEGEND
		VERTICAL MULTIPLES

Figure 6-33. Bar chart of orders by employee

Aggregating Orders by Product

This example will show the number of orders by product. Follow these steps:

- 1. Click Power View Reports on the Excel Insert tab.
- 2. Select the Product Name field in the products table in the Power View Fields pane, as shown in Figure 6-34.

PLE HOME INSERT PAGE LAYOUT PLE HOME INSERT PAGE LAYOUT Parte Copy Undo Redo Themes Bac Clipboard Undo/Redo Themes	FORMULAS DATA REVIEW VIE tt * f Size * Set Image Transparency Image ? Polition * Background Image	chapterreportspowerview - Excel W POWER VIEW DESIGN POWER QU DESIGN POWER QU POWER VIEW DESIGN POWER QU POWER VIEW PO	VERY POWERPIVOT Team	2 10 - 6 13 Nei - 10
Click here Product Name Northwist Tasses Almosis Northwist Tasses Beer Northwist Tasses Beer Northwist Tasses Bowe Min Northwist Tasses Bowe Min Northwist Tasses Down Min Northwist Tasses Color Northwist Tasses Co	e to add a title	Filters •	× Power View Fields × ACTIVE ALL Category Description Discontinued ∑ ID ∑ Uit Price Product Name ♥ ₩ Drag fields between areas below: THLE 8V FreUDS Product Name ♥	
Sheet1 Power View2 Po	ower View9 Power View10 Power Vie	ews 🛞	1	Þ

Figure 6-34. Product list

3. Next, select the Quantity field in the Order Details Table, as shown in Figure 6-35. Note that the sum of the quantity is shown.

1 - 5-	¢• ≣ ≥×	■ • =	500144445			chapterreport	spowerview - Exc	el		IIII		7 0 - 8 4
Clipboard	Undo Redo	A Font A Font A Themes Backg	ize * Set	Image Tra Position - Background Im	insparency age	POWER VIEW	Fit to Field Window List View	Filters Area	Power Tes View Bo	t Picture	Arrange Arrange	Nell +
	Product Tane Notimine Task Notimine Task	rs Almonds Ins Barr I Barr In Dysebury Spread Ins Dysebury Spread Ins Dysebury Spread Ins Charles Ins Charl Ins Charles Ins Ch	Click he 2024100 4020 4020 4020 4020 20200 8500 8500 20200 4020 400 40	re to ad	d a title			x ~ Filters		Pow Activity Drag f	er View Fields × ALL ALL ALL ALL ALL ALL ALL ALL ALL ALL	E
4 5	Sheet1 Pov	ver View2 Pow	er View9 P	ower View10	Power View	8 🕘 🕀			: 4			Þ

Figure 6-35. Quantity of products ordered

4. Click anywhere in the table and use the handles to expand it, as shown in Figure 6-36.

Image: Image	chapterreportspowerview - Excel ATA REVIEW VIEW POWER VIEW DESIGN POWER QUERY	POWERPIVOT Team
Table Bar Column Other Map - Chart- Chart- Chart- Switch Yusulization - The State	Show Tetak Levels - Options Number Tet Arange	
Product Name Product Name Northwid Trader Almonds Northwid Trader Almonds Northwid Trader Almonds Northwid Trader Charling Northwid Trader Sharling Northwid Trader Manuale Albon Northwid Trader Manuale Northwid Trader Manuale Northwid Trader Manuale Northwid Trader Manuale Northwid Trader Sharling Northwid Trader S	to add a title	Power View Fields × Active ALL - Inder Details - Date Allocated 2: Discount 2: Discount 2: Discount 2: Directory ID 2: Product ID 2: Product ID 2: Product ID 2: Status ID 2
Sheet1 Power View2 Power View9 Pow	er View10 Power View8	()

Figure 6-36. Expanded view of table

5. To create a chart, right-click the table and select copy. Click to the right of the table on the design surface and select Paste. Then, on the Design tab, select Bar Chart and Stacked Bar. At the top of the design screen, enter the title "Product Quantity Sold." The result is shown in Figure 6-37.

Product Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Country Country Country Product Name Product Name Product Name Country Country Product Name Product Name Product Name Product Name Country Country Product Name Product Name Product Name Product Name Country Country Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Product Name Nomited Addition Country Country Product Name Product Name Product Name Nomited Addition Country Product Name Product Name Product Name Product Name Nomited Addition Country Product Name Product Name Product Name Product Name Nomited Addition Country Product Name Product Name	N B 5-	¢- □ ≥ □• =	FOR	chapterreport:	spowerview - Excel		7 8 - 8
Product Quantity Sold ************************************	Pice Hom S Cut Paste Copy Clipboard	Undo/Redo Themes	ekground * Background Ima	e view POWERView powerview powe	Fit to Field Filters Window List Area View	Power Coert Powerviol team	Nell *
Northwind Tasker Long Gan Raje A000 Northwind Tasker Long and Stark Starker A article Deformed And Starker A article Deforme		Product Name Northwind Takes Bar Northwind Takes Bar Northwind Takes Days Northwind Takes Cal Northwind Takes Cal Northwind Takes Can Date Cal Northwind Takes Can Date Cal Northwind Takes Can Northwind Takes Can Northwind Takes Can Northwind Takes Can Northwind Takes Can Northwind Takes Cal Northwind Takes Cal Northwind Takes Cal Northwind Takes Date Stars Northwind Takes Cal Northwind Takes Source Northwind Takes Source Northwind Takes Source Northwind Takes Source	Product Quantit	y Sold Name	X ~ Fitters	Power View Fields × ACTIVE ALL ACTIVE ALL Date Allocated Discount Discou	

Figure 6-37. Bar chart of product quantity sold

Note that in the bar chart, only the first nine characters of the name are shown. It is possible to hover the mouse pointer over any bar in the bar chart to see the full product name and quantity sold. To display more characters, click the chart window to select it, and drag the handles to expand the size, as shown in Figure 6-38.



Figure 6-38. Bar chart with expanded product name

Summary

This chapter has just scratched the surface of showing how to use Power View to easily generate data tables and charts to bring data to life. It has covered how to analyze sales by city, by employee, and product. Creating simple maps has also been demonstrated. Chapter 9 will show how to create more sophisticated maps, using Power Map.

CHAPTER 7

Calculating with Data Analysis Expressions (DAX)

This chapter covers calculations with Data Analysis Expressions (DAX) and measuring performance with Key Performance Indicators (KPIs). It will focus on some practical examples of the use of DAX and KPIs.

Understanding Data Analysis Expressions

As explained in Chapter 4, Data Analysis Expressions (DAX) is a formula language used to create calculated columns and fields based on the tabular data model of PowerPivot. DAX formulas differ from regular Excel formulas, which are cell-based, in that DAX formulas apply to all rows in the specified columns. DAX formulas are used in the following two contexts:

- Calculated columns. Calculated columns can be added to an existing PowerPivot table. When a column contains a formula, the value is calculated for each row. To create a formula, click any cell in a new column and press Enter to start entering a formula that is based on column references. Column names can be referenced by pointing to them with the mouse or by entering the names enclosed in square brackets. Columns from other worksheets can be referenced by entering the worksheet name before the column name. If the worksheet name contains a space, it can be enclosed in single quotes. Up to 64 levels of functions can be nested in a DAX expression.
- 2. Calculated fields or measures. These are placed in the calculation area at the bottom of the PowerPivot window. Calculated fields often include aggregation functions, such as sum or average, which by default reference the entire column. DAX functions are very similar to Excel functions, but DAX functions are column-oriented. All available DAX functions can be viewed by clicking the fx icon on the Design tab of the PowerPivot window. When a calculated field is created, the formula is preceded by an identifier, which is like a variable name and can be used to reference the calculated field.

Caution When using DAX, it is best to avoid using the same name for a calculated field and a calculated column. To further avoid ambiguity, it is best when, referencing a column, to use the fully qualified column name, which includes the name of the worksheet.

CHAPTER 7 CALCULATING WITH DATA ANALYSIS EXPRESSIONS (DAX)

DAX Operators

DAX uses the following:

the standard arithmetic operators

- + Addition
- Subtraction
- * Multiplication
- / Division
- Exponentiation

the comparison operators

- = Equal
- > Greater than
- < Less than
- >= Greater than or equal to
- <= Less than or equal to
- <> Not equal to

the text concatenation operator

& Concatenation

and the logical operators

- && And
- ∥ OR

Summary of Key DAX Functions Used in This Chapter

SUM: A simple sum, such as in Excel

SUMX: Iterates through one row at a time in order to do calculations

CALCULATE: Allows calculating based on a filter condition

LEFT: Returns the specified number of characters from the start of the named text string

Updating Formula Results

There are two ways to update the results of formulas.

- 1. Refreshing data involves updating the data in the underlying data source, such as SQL Server, MySQL, or Access.
- **2.** Recalculating involves updating the results of formulas based on changes to the formulas.

ANALYZING SALES DATA

This chapter will continue the example used at the end of Chapter 4—the Contoso database available for download from Tinyurl.com/PowerPivotSamples. Note that the URL is case-sensitive. Once the data is downloaded, open the PowerPivotTutorialSamples spreadsheet.

One way of thinking about this example is to use a star model architecture that includes a fact table containing the numbers that have to be aggregated, surrounded by dimension tables, such as products and stores. In this example, the dboFactSales table is the fact table, as shown in Figure 7-1. To create this example, follow these steps:

1. Click Manage on the PowerPivot tab and then click on Diagram View to see the tables in Diagram View as shown in Figure 7-1.



Figure 7-1. Diagram View

2. Click Data View and then click the dbo_FactSales tab at the bottom of the window to view the dbo_FactSales tab. Scroll all the way to the right and notice the final column, TotalProfit, which is a DAX calculated field based on TotalSales minus TotalCost, as shown in Figure 7-2. Click any cell in the column to see the formula displayed just below the ribbon.

CHAPTER 7 CALCULATING WITH DATA ANALYSIS EXPRESSIONS (DAX)

ile Ho	me Design	Advanced							
aste Dipboard	Get External Ref	resh PivotTabl	Data Type : Auto (Format : General • \$ • % ?	Currency) - 22↓ 3 Z↓ 500 200 F g Sc	iear All Sort b filters Column ort and Filter	y Find	Calculations	Data View Diagram View	rea
TotalProfi	t] •	fx =[TotalSa	ales]-[TotalCost]						
Disce	ountAmount	TotalCost	TotalSales	ETLLoadID	🔽 LoadDa	te 🖬 U	IpdateDate 🖬	TotalProfit	Add Column
0		\$0 \$435	5.87 \$85	.91	6 1/1/200	8 12 1	1/1/2008 12:0	\$419.04	
0		\$0 \$1,212	2.66 \$2	637	6 1/1/200	8 12 1	1/1/2008 12:0	\$1,424.34	
0		\$0 \$1,509	9.57 \$2	961	6 1/1/200	8 12 1	1/1/2008 12:0	\$1,451.43	
0		\$0 \$864	4.72 \$2,60	9.91	6 1/1/200	8 12 1	1/1/2008 12:0	\$1,745.19	
0		\$0 \$2,354	4.94 \$5	121	6 1/1/200	8 12 1	1/1/2008 12:0	\$2,766.06	
0		\$0 \$1,273	3.23 \$3	843	6 1/2/200	8 12 1	1/2/2008 12:0	\$2,569.77	
0		\$0 \$68	8.31 \$1	18.5	6 1/2/200	8 12 1	1/2/2008 12:0	\$80.19	
0		\$0 \$45	5.81 \$8	9.91	6 1/2/200	8 12 1	1/2/2008 12:0	\$44.1	
0		\$0 \$658	8.08 \$1	431	6 1/2/200	8 12 1	1/2/2008 12:0	\$772.92	
0		\$0 \$1,067	7.85 \$2	322	6 1/2/200	8 12 1	1/2/2008 12:0	\$1,254.15	
0		\$0 \$683	3.64 \$1	341	6 1/3/200	8 12 1	1/3/2008 12:0	\$657.36	
0		\$0 \$5	7.51 \$12	5.01	6 1/3/200	8 12 1	1/3/2008 12:0	\$67.5	
0		\$0 \$642	2.33 \$1,25	9.91	6 1/3/200	8 12 1	1/3/2008 12:0	\$617.58	
0		\$0 \$197	7.28	387	6 1/4/200	8 12 1	1/4/2008 12:0	\$189.72	
0		\$0 \$1,37	1.96 \$2	691	6 1/4/200	8 12 1	1/4/2008 12:0	\$1,319.04	
0		\$0 \$7,45	1.73 \$22	491	6 1/4/200	8 12 1	1/4/2008 12:0	\$15,039.27	
0		\$0 \$596	5.07 \$1,1	59.1	6 1/5/200	8 12 1	1/5/2008 12:0	\$573.03	
0		\$0 \$54	4.63 \$1	07.1	6 1/5/200	8 12 1	1/5/2008 12:0	\$52.47	

Figure 7-2. Calculated column

Creating Measures or Calculated Fields

This section shows how to create measures or calculated fields based on the data in the tables.

- 1. To add a calculated field or measure, go to the first cell below the data grid at the bottom of the TotalProfit column and select the cell. Double-click the AutoSum icon on the tab to get the sum. If the calculation area or measure grid at the bottom of the data grid is not visible, click the Calculation Area icon at the right end of the tab.
- 2. Next, select the cell below the sum cell just created, click the down arrow after the AutoSum icon, and from the drop-down, select average, as shown in Figure 7-3. It may be necessary to expand the column to see the totals. Do that by clicking the bar separating the column headings, and when a double arrow appears, drag to the right. Note that an identifier "Sum of Total Profits" and "Average of Total Profits" is created. This identifier acts as both a label and a variable name used to reference the calculated field.
| 1 2 2 2 2 2 2 | | | PowerP | vot for Excel - Cop | y of PowerPiv | otTutorialSar | mplechap7.x | isx | | - | D |
|---------------------|-----------------------------------|--------------|--------------------|---|---|------------------------------------|--------------|----------------|---|-----------|-----|
| Home Design | Advanced | | | | | | | | | | |
| Paste Append | From From Data From Other Sources | g Refre | sh PivotTable | Data Type : -
Format : •
\$ • % • *** *** | Image: Sort A to Z Image: Sort Z to A Image: S | Clear All Sort b
Filters Column | y Find | ∑ AutoSum • | ata Diagram Show Calculation
ew View Hidden Area | | |
| alProfit] - | & Average of TotalProfit:=AVER/ | GEITotals | (rofit) | romating | Sort a | na riiter | ring | calculations | view | | |
| | | to citro tan | | and Palac | | | | Lindsto Data | TatalOcofia | | |
| | scountquantity DiscountAmot | 60 KO | ¢1 376 46 | 510150165 | \$2,600,01 | | 1/2/2002 12 | 1/8/2008 12-0 | rotalPront | 222.45 | Add |
| 50 | 0 | 20 | \$1,370.40 | | \$2,099.91 | 0. | 1/8/2008 12 | 1/8/2008 12:0 | > | 407.24 | |
| 50 | 0 | 50 | \$1,274.70 | | \$2,112 | 6 | 1/9/2008 12 | 1/9/2008 12:0 | \$ | \$121.22 | |
| 50 | | 00 | \$107.00 | | \$242.91 | 0 .
6 . | 1/9/2008 12 | 1/9/2008 12:0 | | \$100 73 | |
| 50 | 0 | 50 | \$197.28 | | \$387 | 6 | 1/9/2008 12 | 1/9/2008 12:0 | | \$189.72 | |
| 50 | 0 | 30 | \$551.55
¢50.59 | | \$116.91 | 6 | 1/10/2008 12 | 1/10/2008 12:0 | | \$57.22 | |
| 50 | 0 | 00 | \$779.05 | | \$1 692 | 6 | 1/10/2008 1 | 1/10/2008 12: | | \$012.05 | |
| 00 | 0 | 50 | \$7 040 29 | | \$4,032 | 6 | 1/10/2008 1 | 1/10/2008 12: | | 296 61 | |
| 30
\$0 | 0 | so
so | \$550.53 | | \$1 079 91 | 6 | 1/10/2008 1 | 1/10/2008 12: | * | \$529.38 | |
| ŝ | 0 | ŝ | \$68.22 | | \$206.01 | 6 | 1/10/2008 1 | 1/10/2008 12: | | \$137.79 | |
| \$0 | 0 | 50 | \$1,985.76 | | \$5,993.46 | 6 | 1/10/2008 1 | 1/10/2008 12: | | \$4.007.7 | |
| so | 0 | \$0 | \$780.03 | | \$1.530 | 6 | 1/11/2008 1 | 1/11/2008 12: | | \$749.97 | |
| \$0 | 0 | \$0 | \$231.75 | | \$504 | 6 | 1/11/2008 1 | 1/11/2008 12: | | \$272.25 | |
| \$0 | 0 | SO | \$45.81 | | \$89.91 | 6 | 1/11/2008 1 | 1/11/2008 12: | | \$44.1 | |
| \$0 | 0 | \$0 | \$814.95 | | \$1.772.1 | 6 | 1/11/2008 1 | 1/11/2008 12: | | \$957.15 | |
| \$0 | 0 | \$0 | \$2,946.06 | | \$8,892 | 6 | 1/11/2008 1 | 1/11/2008 12: | s | 5,945,94 | |
| \$0 | 0 | \$0 | \$4,651.74 | | \$14,040 | 6 | 1/12/2008 1 | 1/12/2008 12: | \$ | 388.26 | |
| \$0 | 0 | \$0 | \$142.2 | | \$278.91 | 6 : | 1/12/2008 1 | 1/12/2008 12: | | \$136.71 | |
| | | | | | | | | | Sum of TotalProfit 2: \$6,969,994.8 | 66.984 | _ |
| | | | | | | | | | Average of TotalProfit: \$1,758.99 | 99 | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| DimChannel dbo_DimC | Date dbo_DimEntity dbo_DimProduct | dbo_DimPro | oductSubcatego | y dbo_FactSales Proc | luctCategory as | ContosoStores | Stores | | | | - |
| H 4 1 of 3 962 476 | | | | home and a second se | | | ana ana an | | | | 1 |

Figure 7-3. Calculated fields

3. Follow the same steps with the Total Sales column to add the sum and average, as shown in Figure 7-4.

0 3 Secret		P	owerPivot for Excel - Copy of PowerPivotTutorialSam	plechap7.xlsx	- 0
ile Home Design	Advanced				6
ste lipboard	m From Data From Other ase * Service * Sources Connect Get External Data	ng Refresh Pive	Data Type: - 2↓ Sort A to Z Sort A to Z Format: - S • % 5 * 20 * 20 * 20 * 20 * 20 * 20 * 20 *	Find Calculations Find Calculations Find Calculations View View Hidden Area View View Hidden Area	
otalSales] 👻 🎵	Average of TotalSales:=AVERA	AGE([TotalSales])			
ReturnAmount 💽 Discou	untQuantity 💌 DiscountAmo	unt 🖪 TotalCost	TotalSales ETLLoadiD) 🗷 LoadDate 🗷 UpdateDate 🗷 TotalProfit	
\$0	0	\$0 \$1,37	.46 \$2,699.91	6 1/8/2008 12 1/8/2008 12:0	\$1,323.4
\$0	0	\$0 \$1,274	.76 \$2,772	6 1/9/2008 12 1/9/2008 12:0	\$1,497.2
\$0	0	\$0 \$111	.69 \$242.91	6 1/9/2008 12 1/9/2008 12:0	\$131.2
\$0	0	\$0 \$193	.28 \$387	6 1/9/2008 12 1/9/2008 12:0	\$189.7
\$0	0	\$0 \$593	.93 \$1,161	6 1/9/2008 12 1/9/2008 12:0	\$569.0
\$0	0	\$0 \$55	.58 \$116.91	6 1/10/2008 1 1/10/2008 12:	\$57.3
\$0	0	\$0 \$778	.05 \$1,692	6 1/10/2008 1 1/10/2008 12:	\$913.9
\$0	0	\$0 \$2,040	.39 \$4,437	6 1/10/2008 1 1/10/2008 12:	\$2,396.6
\$0	0	\$0 \$550	.53 \$1,079.91	6 1/10/2008 1 1/10/2008 12:	\$529.3
\$0	0	\$0 \$68	.22 \$206.01	6 1/10/2008 1 1/10/2008 12:	\$137.
\$0	0	\$0 \$1,985	.76 \$5,993.46	6 1/10/2008 1 1/10/2008 12:	\$4,007
\$0	0	\$0 \$780	.03 \$1,530	6 1/11/2008 1 1/11/2008 12:	\$749.
\$0	0	\$0 \$231	.75 \$504	6 1/11/2008 1 1/11/2008 12:	\$272.3
\$0	0	\$0 \$43	.81 \$89.91	6 1/11/2008 1 1/11/2008 12:	\$44
\$0	0	\$0 \$814	.95 \$1,772.1	6 1/11/2008 1 1/11/2008 12:	\$957.1
\$0	0	\$0 \$2,946	.06 \$8,892	6 1/11/2008 1 1/11/2008 12:	\$5,945.5
\$0	0	\$0 \$4,65	.74 \$14,040	6 1/12/2008 1 1/12/2008 12:	\$9,388.2
\$0	0	\$0 \$14	2.2 \$278.91	6 1/12/2008 1 1/12/2008 12:	\$136.7
			Sum of TotalSales 2: \$12,334,891,468.764 Average of TotalSales: \$3,112.9252	Sum of TotalProfit 2: \$6,90 Average of TotalProfit: \$1	59,994,866.984 ,758.9999

Figure 7-4. Additional calculated fields

Analyzing Profitability

This example will show how to insert a new column to analyze the profitability of each transaction.

1. Add a calculated column for profitability or profit divided by sales by clicking anywhere in the rightmost column named "Add Column" and pressing "=". Then click the TotalProfit column, press /, and click the TotalSales column to define the formula and then press Enter. The result is shown in Figure 7-5.

🖲 🕼 🛃 🖘 🗠 🕫	t ()		Powerf	Pivot for Excel - Copy of Pow	erf	PivotTutorialSa	amplechap7.xlsx					- 0
File Home De:	lign											^
Paste Append	From From Data From Other Database * Service * Sources Get External Data	Existing Connections	PivotTable	Data Type : Auto (Decimal Number Format : General • \$ • % \$ 38 428 Formatting	ŋ	오↓ Sort Smalles 김↓ Sort Largest 兖 Clear Sort	t to Largest to Smallest Clear All Filters	Sort by Column *	Find	∑ Au Cre Calcu	toSum • eate KPI Data Diagram View View	Show Calculatio
[CalculatedCo •	fr =[TotalProfit]/[TotalSal	les]			_							
ReturnAmount	DiscountQuantity	ntămount 💌 Tota	dCost 🖬 1	fotalSales 🔽 FTU oadID	¥	LoadDate	UpdateDate	TotalPro	6e		CalculatedColumn1	Add Column
\$0	0	\$0	\$435.87	\$854.91	6	1/1/2008 12	1/1/2008 12:0			419.04	0.490156858616697	, roo coronar
\$0	0	\$0	\$1,212.66	\$2.637	6	1/1/2008 12	1/1/2008 12:0		\$1	424.34	0.540136518771331	1
\$0	0	\$0	\$1,509.57	\$2,961	6	1/1/2008 12	1/1/2008 12:0		\$1	451.43	0.490182370820669	
\$0	0	\$0	\$864.72	\$2,609.91	6	1/1/2008 12	1/1/2008 12:0		\$1	745.19	0.668678230283803	1
\$0	0	\$0	\$2,354.94	\$5,121	6	1/1/2008 12	1/1/2008 12:0		\$2	,766.06	0.540140597539543	
\$0	0	\$0	\$1,273.23	\$3,843	6	1/2/2008 12	1/2/2008 12:0		\$2	,569.77	0.668688524590164	8
\$0	0	\$0	\$68.31	\$148.5	6	1/2/2008 12	1/2/2008 12:0			\$80.19	0.54	
\$0	0	\$0	\$45.81	\$89.91	6	1/2/2008 12	1/2/2008 12:0			\$44.1	0.49049049049049	1
\$0	0	\$0	\$658.08	\$1,431	6	1/2/2008 12	1/2/2008 12:0		4	772.92	0.540125786163522	
\$0	0	\$0	\$1,067.85	\$2,322	6	1/2/2008 12	1/2/2008 12:0		\$1	,254.15	0.540116279069767	1
\$0	0	\$0	\$683.64	\$1,341	6	1/3/2008 12	1/3/2008 12:0		\$	657.36	0.490201342281879	
\$0	0	\$0	\$57.51	\$125.01	6	1/3/2008 12	1/3/2008 12:0			\$67.5	0.539956803455724	
\$0	0	\$0	\$642.33	\$1,259.91	6	1/3/2008 12	1/3/2008 12:0		\$	617.58	0.490177869847846	1
\$0	0	\$0	\$197.28	\$387	6	1/4/2008 12	1/4/2008 12:0		-	\$189.72	0.490232558139535	
\$0	0	\$0	\$1,371.96	\$2,691	6	1/4/2008 12	1/4/2008 12:0		\$1	319.04	0.490167224080268	8
\$0	0	\$0	\$7,451.73	\$22,491	6	1/4/2008 12	1/4/2008 12:0		\$15	,039.27	0.668679471788715	
\$0	0	\$0	\$596.07	\$1,169.1	6	1/5/2008 12	1/5/2008 12:0		-	573.03	0.490146266358737	1
\$0	0	\$0	\$54.63	\$107.1	6	1/5/2008 12	1/5/2008 12:0			\$52.47	0.489915966386555	
<					÷	-		-		-		
dbo_DimChannel dbo_0	DimDate dbo_DimEntity dbo_DimI	Product dbo_DimProc	luctSubcatego	ry dbo_FactSales ProductCatego	ory	se ContosoStore	s Stores					
Record M 4 9 of 3.962	476 F H	-			-		and the second second					
					_							Ecol.

Figure 7-5. Calculated field for profit/sales

 To format the column to display percentages, select the new column and then click the down arrow after Format on the Home tab and select Percentage. The result is shown in Figure 7-6.

File Home Design									
Paste Append Paste Replace Fro Datab Cipboard	From Data From Other See * Service * Sources Connection Get External Data	Refres	h PivotTable	Data Type : Auto (Decimal Number) • Format : Percentage • \$ • % 5 30 400 Formatting	출↓ Sort Smalle え↓ Sort Larges ඌ Clear Sort	t to Smallest Clear All Soft and Filter	t by ann • Find Find Calculati	um • J KPI Data Diagram S View View H ions Vie	Show Calculatio
alculatedCo 👻 🍂	=[TotalProfit]/[TotalSales]								
ReturnAmount 🔜 Disco	untQuantity 🗷 DiscountAmou	nt 🖬 Tot	alCost 💌 T	otalSales 💽 ETLLoadiD 💌	LoadDate	UpdateDate 🔽 Tot	talProfit 💽 Ca	alculatedColumn1 💽	Add Column
\$0	0	\$0	\$435.87	\$854.91 6	1/1/2008 12	1/1/2008 12:0	\$419.04	49.02 %	
\$0	0	\$0	\$1,212.66	\$2,637 6	1/1/2008 12	1/1/2008 12:0	\$1,424.34	54.01%	
\$0	0	\$0	\$1,509.57	\$2,961 6	1/1/2008 12	1/1/2008 12:0	\$1,451.43	49.02 %	
\$0	0	\$0	\$864.72	\$2,609.91 6	1/1/2008 12	1/1/2008 12:0	\$1,745.19	66.87%	
\$0	0	\$0	\$2,354.94	\$5,121 6	1/1/2008 12	1/1/2008 12:0	\$2,766.06	54.01 %	
\$0	0	\$0	\$1,273.23	\$3,843 6	1/2/2008 12	1/2/2008 12:0	\$2,569.77	66.87%	
\$0	0	\$0	\$68.31	\$148.5 6	1/2/2008 12	1/2/2008 12:0	\$80.19	54.00 %	
\$0	0	\$0	\$45.81	\$89.91 6	1/2/2008 12	1/2/2008 12:0	\$44.1	49.05 %	
\$0	0	\$0	\$658.08	\$1,431 6	1/2/2008 12	1/2/2008 12:0	\$772.92	54.01%	
\$0	0	\$0	\$1,067.85	\$2,322 6	1/2/2008 12	1/2/2008 12:0	\$1,254.15	54.01%	
\$0	0	\$0	\$683.64	\$1,341 6	1/3/2008 12	1/3/2008 12:0	\$657.36	49.02 %	
\$0	0	\$0	\$57.51	\$125.01 6	1/3/2008 12	1/3/2008 12:0	\$67.5	54.00 %	
\$0	0	\$0	\$642.33	\$1,259.91 6	1/3/2008 12	1/3/2008 12:0	\$617.58	49.02 %	
\$0	0	\$0	\$197.28	\$387 6	1/4/2008 12	1/4/2008 12:0	\$189.72	49.02 %	
\$0	0	\$0	\$1,371.96	\$2,691 6	1/4/2008 12	1/4/2008 12:0	\$1,319.04	49.02 %	
\$0	0	\$0	\$7,451.73	\$22,491 6	1/4/2008 12	1/4/2008 12:0	\$15,039.27	66.87 %	
\$0	0	\$0	\$596.07	\$1,169.1 6	1/5/2008 12	1/5/2008 12:0	\$573.03	49.01%	
\$0	0	\$0	\$54.63	\$107.1 6	1/5/2008 12	1/5/2008 12:0	\$52.47	48.99 %	

Figure 7-6. Formatting as percentage

3. To rename the column, right-click the column heading, select Rename Column, and enter Profit/Sales, as shown in Figure 7-7.

🐌 💷 🚽 So =			Powerf	Pivot for Excel - Copy of Poweri	PivotTutorialSa	mplechap7.xlsx			- 8 ×
File Home Des	ign								^
Paste Append Paste Paste Replace	From From Data From Other Database * Service * Sources	Existing Connections	PivotTable	Data Type : Auto (Decimal Number) * Format : Percentage * \$ * % > 36 40	출↓ Sort Smalles え↓ Sort Largest ☆ Clear Sort	t to Largest to Smallest Clear All Filters Column •	Find Create K	Data Diagram Sho View View Hidd	w Calculation
Clipboard	Get External Data			Formatting		Sort and Filter	Find Calculation	s View	
[Profit/Sales] •	f _w =[TotalProfit]/[TotalSa	iles]							
🚄 ReturnAmount 🖃	DiscountQuantity 🗷 Discou	intAmount 💽 Totak	iost 🗷 1	fotalSales 🛛 💀 ETLLoadID 💌	LoadDate 💌	UpdateDate 💌 TotalPro	fit 💌 Prof	it/Sales 💽 🖌 🗛	dd Column
\$0	0	\$0	\$435.87	\$854.91 6	1/1/2008 12	1/1/2008 12:0	\$419.04	49.02 %	
\$0	0	\$0 \$1	,212.66	\$2,637 6	1/1/2008 12	1/1/2008 12:0	\$1,424.34	54.01 %	
\$0	0	\$0 \$1	,509.57	\$2,961 6	1/1/2008 12	1/1/2008 12:0	\$1,451.43	49.02 %	
\$0	0	\$0	\$864.72	\$2,609.91 6	1/1/2008 12	1/1/2008 12:0	\$1,745.19	66.87 %	
\$0	0	\$0 \$3	,354.94	\$5,121 6	1/1/2008 12	1/1/2008 12:0	\$2,766.06	54.01 %	
\$0	0	\$0 \$1	,273.23	\$3,843 6	1/2/2008 12	1/2/2008 12:0	\$2,569.77	66.87 %	
\$0	0	\$0	\$68.31	\$148.5 6	1/2/2008 12	1/2/2008 12:0	\$80.19	54.00 %	
\$0	0	\$0	\$45.81	\$89.91 6	1/2/2008 12	1/2/2008 12:0	\$44.1	49.05 %	
\$0	0	\$0	\$658.08	\$1,431 6	1/2/2008 12	1/2/2008 12:0	\$772.92	54.01 %	
\$0	0	\$0 \$1	,067.85	\$2,322 6	1/2/2008 12	1/2/2008 12:0	\$1,254.15	54.01 %	
\$0	0	\$0	\$683.64	\$1,341 6	1/3/2008 12	1/3/2008 12:0	\$657.36	49.02 %	
\$0	0	\$0	\$57.51	\$125.01 6	1/3/2008 12	1/3/2008 12:0	\$67.5	54.00 %	
\$0	0	\$0	\$642.33	\$1,259.91 6	1/3/2008 12	1/3/2008 12:0	\$617.58	49.02 %	
\$0	0	\$0	\$197.28	\$387 6	1/4/2008 12	1/4/2008 12:0	\$189.72	49.02 %	
\$0	0	\$0 \$1	,371.96	\$2,691 6	1/4/2008 12	1/4/2008 12:0	\$1,319.04	49.02 %	
\$0	0	\$0 \$7	,451.73	\$22,491 6	1/4/2008 12	1/4/2008 12:0	\$15,039.27	66.87 %	
\$0	0	\$0	\$596.07	\$1,169.1 6	1/5/2008 12	1/5/2008 12:0	\$573.03	49.01 %	
\$0	0	\$0	\$54.63	\$107.1 6	1/5/2008 12	1/5/2008 12:0	\$52.47	48.99 %	
100									
<									>
dbo DimChannel dbo D	imDate dbo DimEntity dbo Dim	Product dbo DimProduc	tSubcatego	any dbo FactSales ProductCategory	ContosoStores	s Stores			

Figure 7-7. Renaming a calculated column

4. To create a calculated field that is the average of all the profit/sales, go down to the calculation area below the data area. Select the cell at the bottom of the Profit/Sales column and then click the down arrow after the AutoSum icon on the ribbon and select average. A calculated field will be created, as shown in Figure 7-8. Expand the column as necessary to display the full value.

ាត់តិត ត =	PowerPivot for Excel - Copy	of chap7 aug 4PowerPivotTutorialSa	imple.xlsx	- 🗆 🗙
F Home Design				^ C .
Paste From From Data From Other Existing Database - Service - Sources Connection Clipboard	Refresh PivotTable S %	ercentage • 2 to the fill of	∑ AutoSum ~ Create KPI Calculations ∑ AutoSum ~ Data Diagram S View View H View View H	how Calculation Area
[Profit/Sales] • f _x Average of Profit/Sa	les:=AVERAGE([Profit/Sales])			*
📶 TotalSales 🛛 🗹 ETLLoadID 🔽	LoadDate 💌 UpdateDate 💌	TotalProfit 💌	Profit/Sales	Add Column ^
\$854.91 6	1/1/2008 12 1/1/2008 12:0	\$419.04	49.02 %	
\$2,637 6	1/1/2008 12 1/1/2008 12:0	\$1,424.34	54.01 %	
\$2,961 6	1/1/2008 12 1/1/2008 12:0	\$1,451.43	49.02 %	
\$2,609.91 6	1/1/2008 12 1/1/2008 12:0	\$1,745.19	66.87%	
\$5,121 6	1/1/2008 12 1/1/2008 12:0	\$2,766.06	54.01 %	
\$3,843 6	1/2/2008 12 1/2/2008 12:0	\$2,569.77	66.87 %	
\$148.5 6	1/2/2008 12 1/2/2008 12:0	\$80.19	54.00 %	
\$89.91 6	1/2/2008 12 1/2/2008 12:0	\$44.1	49.05 %	
\$1,431 6	1/2/2008 12 1/2/2008 12:0	\$772.92	54.01 %	
\$2,322 6	1/2/2008 12 1/2/2008 12:0	\$1,254.15	54.01 %	
\$1,341 6	1/3/2008 12 1/3/2008 12:0	\$657.36	49.02 %	
\$125.01 6	1/3/2008 12 1/3/2008 12:0	\$67.5	54.00 %	
\$1,259.91 6	1/3/2008 12 1/3/2008 12:0	\$617.58	49.02 %	
\$387 6	1/4/2008 12 1/4/2008 12:0	\$189.72	49.02 %	
\$2,691 6	1/4/2008 12 1/4/2008 12:0	\$1,319.04	49.02 %	
\$22,491 6	1/4/2008 12 1/4/2008 12:0	\$15,039.27	66.87 %	
\$1,169.1 6	1/5/2008 12 1/5/2008 12:0	\$573.03	49.01 %	
\$107.1 6	1/5/2008 12 1/5/2008 12:0	\$52.47	48.99 %	~
Sum of TotalSales 2: \$12,334,		Sum of TotalProfit 2: \$6,969,994,866	Average of Profit/Sales: 53.31 %	^
Average of TotalSales: \$3,112		Average of TotalProfit: \$1,758.9999		
<				>
dbo_DimChannel dbo_DimDate dbo_DimEntity dbo_D	imProduct dbo_DimProductSubcateg	ory dbo_FactSales ProductCategory are C	ontosoStores Stores	
Record: 14 4 1 of 3,962,476 > >				
				0

Figure 7-8. Average of Profit/Sales

5. To add calculated fields for minimum and maximum profit/sales, select the cell in the calculation area just below the one where the average is. Click the arrow after AutoSum on the ribbon, select maximum, and then repeat and select minimum. Calculated fields for average minimum and maximum will be created, as shown in Figure 7-9.

e Home Design						-
te	ther Existing	Refresh Piv	Data Type : Format : Pe	rcentage - 21 Clear All Sort by Find	Create KPI Data Diagram S	show Calculation
Database • Service • Sour	ces Connection	s •	• 3 · 70	2 .00 +.0 ZP Filters Column -	View View H	idden Area
ofit/Sales] • & Minin	num of Profit/S	ales:=MIN/[Pro	ofit/Sales])	solt and Filter Pille	Calculations	vv
TotalSalar	STU opdin	LoadDate 🔽	LindateDate	TotalBrofit	Profit/Salac	Add Column
\$854.91	6	1/1/2008 12	1/1/2008 12:0	5419 04	49.02%	Add Column
\$2 637	6	1/1/2008 12	1/1/2008 12:0	\$1 424 34	54.01%	
\$2.951	6	1/1/2008 12	1/1/2008 12:0	\$1.451.43	49.02 %	
\$2,609.91	6	1/1/2008 12	1/1/2008 12:0	\$1,745,19	66.87%	
\$5.121	6	1/1/2008 12	1/1/2008 12:0	\$2,766.06	54.01 %	
\$3,843	6	1/2/2008 12	1/2/2008 12:0	\$2,569,77	66.87%	
\$148.5	6	1/2/2008 12	1/2/2008 12:0	\$80.19	54.00 %	
\$89.91	6	1/2/2008 12	1/2/2008 12:0	\$44.1	49.05 %	
\$1,431	6	1/2/2008 12	1/2/2008 12:0	\$772.92	54.01 %	
\$2,322	6	1/2/2008 12	1/2/2008 12:0	\$1,254.15	54.01 %	
\$1,341	6	1/3/2008 12	1/3/2008 12:0	\$657.36	49.02 %	
\$125.01	6	1/3/2008 12	1/3/2008 12:0	\$67.5	54.00 %	
\$1,259.91	6	1/3/2008 12	1/3/2008 12:0	\$617.58	49.02 %	
\$387	6	1/4/2008 12	1/4/2008 12:0	\$189.72	49.02 %	
\$2,691	6	1/4/2008 12	1/4/2008 12:0	\$1,319.04	49.02 %	
\$22,491	6	1/4/2008 12	1/4/2008 12:0	\$15,039.27	66.87 %	
\$1,169.1	6	1/5/2008 12	1/5/2008 12:0	\$573.03	49.01 %	
\$107.1	6	1/5/2008 12	1/5/2008 12:0	\$52.47	48.99 %	
Sum of TotalSales 2: \$12,334,				Sum of TotalProfit 2: \$6,969,994,866	Average of Profit/Sales: 53.31 %	
Average of TotalSales: \$3,112				Average of TotalProfit: \$1,758.9999	Maximum of Profit/Sales: 71.0	
					Minimum of Profit/Sales: 42.41	1

Figure 7-9. Adding calculated fields for minimum and maximum profit/sales

Using the SUMX Function

The SUM function can be used to sum a single column. By contrast, SUMX iterates through each column and performs a calculation. So, for instance, SUMX could be used to sum the product of price time's quantity for each row. To do so, follow these steps:

- 1. Select the next cell down in the first column of the calculation area.
- 2. Enter the following formula to test SUMX and format it as currency, as shown in Figure 7-10:

```
testsumx:=SUMX(dbo_FactSales,dbo_FactSales[UnitPrice]*
dbo_FactSales[SalesQuantity])
```

Paste Append Paste Replace Ste Copy	From Database	From Data From Other Service - Sources	Existing Connections	efresh PivotTable	Data Type : - Format : Currency - \$ - % - 26 428	2 Sort A to Z 2 Sort Z to A 2 Clear Sort	Jear All Sort by Filters Column	Find D	AutoSum - Create KPI Di Vi	ata Diagram Show Calculation ew View Hidde	
[SalesKey] - X	√ fr te	stsumx:=SUMX(dbo	FactSales,dbo	FactSales[UnitP	rice]*dbo FactSales[S	alesQuantity])		THIS S	arcolaciona	TTEN .	
SalesKey		Dat 🕫 🖬 char	ne 🐔 🖬 S	or 🗟 🖬 Pro	duc 🐔 🖬 Promoti	ionKey 🖬 Cur	rencyKey 🖬 U	InitCost 🖬	UnitPrice	SalesQuantity 🔽 ReturnQuan	tity 🔽 Retur
Index Section 4	551966	20080101	2	199	2018	11	1	\$48.43	\$94.99	9	0
	551995	20080101	2	199	1452	11	1	\$134.74	\$293	9	0
	552014	20080101	2	199	254	11	1	\$167.73	\$329	9	0
	552038	20080101	2	199	1615	11	1	\$96.08	\$289.99	9	0
	552126	20080101	2	199	206	11	1	\$261.66	\$569	9	0
	552193	20080102	2	199	1104	11	1	\$141.47	\$427	9	0
	552214	20080102	2	199	781	11	1	\$7.59	\$16.5	9	0
	552222	20080102	2	199	2442	11	1	\$5.09	\$9.99	9	0
	552238	20080102	2	199	737	11	1	\$73.12	\$159	9	0
	552389	20080102	2	199	1453	11	1	\$118.65	\$258	9	0
	552424	20080103	2	199	2174	11	1	\$75.96	\$149	9	0
	552447	20080103	2	199	1628	11	1	\$6.39	\$13.89	9	0
	552633	20080103	2	199	1998	11	1	\$71.37	\$139.99	9	0
	552659	20080104	2	199	1782	11	1	\$21.92	\$43	9	0
	552736	20080104	2	199	310	11	1	\$152.44	\$299	9	0
	552850	20080104	2	199	552	11	1	\$827.97	\$2,499	9	0
	552895	20080105	2	199	2175	11	1	\$66.23	\$129.9	9	0
	552946	20080105	2	199	794	11	1	\$6.07	\$11.9	9	0
lestsumx: \$12,617,81	4,712.97]									

Figure 7-10. Using the SUMX function

Using the CALCULATE Function

The CALCULATE() function is used to calculate an expression based on a filter. The following example calculates total sales through the store channel. Follow these steps:

- 1. Select the top leftmost cell in the Calculation area.
- 2. Enter the formula, as shown in Figure 7-11. The DAX formula is

```
StoreSales:=CALCULATE(SUM(dbo_FactSales[TotalSales]),
dbo_DimChannel[ChannelName]="Store")
```

File Home	Design					Excer copy or for		preenaps	011201					~ 6
Paste Da Clipboard	From From Dat tabase • Service • Get E	a From Other Sources	Existing Connections	Refresh	PivotTable	Data Type : ~ Format : General ~ \$ ~ % ? 3% 4% Formatting	2↓ X↓ Clear All Sort by Filters Column - Sort and Filter	Find	∑ A 2 Cal	RutoSum • Treate KPI culations	Data Diag View Vie	gram Show Hidder View	Calculation Area	
[SalesKey]	 	Store Sale	s:=CALCULAT	re(SUM(d	bo_FactSal	es[TotalSales]), dbo_	DimChannel[Channel	Name]="	Store	")				×
SalesKey	🖬 Di	at 👘 🖬	channe 🖣	🕯 🖬 Sto	n 🕬 💌	Produc ঝ 💌 Pro	omotionKey 💽 Curr	encyKey		UnitCost	🖬 UnitPrie	ie 🖬 Sa	lesQuantity	
	551966	20080101		2	199	2018	11		1	\$48.4	3 \$	94.99		
	551995	20080101		2	199	1452	11		1	\$134.7	4	\$293		
	552014	20080101		2	199	254	11		1	\$167.7	3	\$329		
	552038	20080101		2	199	1615	11		1	\$96.0	8 \$2	89.99		
	552126	20080101		2	199	206	11		1	\$261.6	6	\$569		
	552193	20080102		2	199	1104	11		1	\$141.4	7	\$427		
	552214	20080102		2	199	781	11		1	\$7.5	9	\$16.5		
	552222	20080102		2	199	2442	11		1	\$5.0	9	\$9.99		
	552238	20080102		2	199	737	11		1	\$73.1	2	\$159		
	552389	20080102		2	199	1453	11		1	\$118.6	5	\$258		
	552424	20080103		2	199	2174	11		1	\$75.9	6	\$149		
	552447	20080103		2	199	1628	11		1	\$6.3	9 \$	13.89		
	552633	20080103		2	199	1998	11		1	\$71.3	7 \$1	.39.99		
	552659	20080104		2	199	1782	11		1	\$21.9	2	\$43		
	552736	20080104		2	199	310	11		1	\$152.4	4	\$299		
	552850	20080104		2	199	552	11		1	\$827.9	7 \$	2,499		
	552895	20080105		2	199	2175	11		1	\$66.2	3 \$	129.9		
	552946	20080105		2	199	794	11		1	\$6.0	7	\$11.9		
Store Sales:	69603454													^
<														>
dbo_DimChanne	dbo_DimDate	dbo_DimEnt	tity dbo_Dim	Product c	bo_DimProd	uctSubcategory dbo_F	actSales ProductCatego	ry an Con	tosoSt	ores Stores	5			
Decord: 14 4	1 / 3 962 476									50.000 000 000 000				-0

Figure 7-11. Using the CALCULATE function to show total store sales

Note that the formula starts with a label, followed by a colon. StoreSales is both the label and identifier for this calculated field. CALCULATE will then sum the TotalSales field from the dbo_FactSales table, in which the channel name is equal to store. When typing a function name or field name, Intellisense pops up a list of what it thinks you want to enter. Highlight the desired value and press Tab to select.

1. Note that the total sales number is not formatted. This can be corrected by selecting the field and clicking the down arrow after the Format icon on the ribbon and selecting currency. The column also needs to be expanded, which can be achieved by clicking the bar between column headings and dragging. The result is shown in Figure 7-12.

1 2 3 0					PowerP	vivot for Excel - Cop	y of PowerPiv	otTutorialSamp	lechap7.xl	5X		- 0
F Home Design												
Paste Append Paste Replace ste E Copy Clipboard	rom abase -	From Data From Other Service • Sources Get External Data	Existing Connections	Refresh	PivotTable	Data Type : - Format : Currency • \$ • % > 26 4% Formatting	2↓ Sort A to Z 3↓ Sort Z to A 20 Clear Sort Sort	Clear All Sort by Filters Column -	Find Find	Calculations	Data Diagram Show Calculation View View	
[SalesKey] 👻	fr Sto	re Sales:=CALCULA	TE(SUM(dbo	FactSale	[TotalSale	s]), dbo_DimChanne	[ChannelNam	e]="Store")				
SalesKey		Dat 🖈 🖬 char	nne 🥵 🖬	Stor	🕯 🖬 Proc	duc 🦟 🖬 Promot	ionKey 🖬 Ci	urrencyKey 🖃 l	JnitCost	UnitPrice	SalesQuantity 🔳 ReturnQuanti	ty 🔄 ReturnA
5	51966	20080101	2		199	2018	11	1	\$48.4	3 \$94.9	9 9	0
5	51995	20080101	2		199	1452	11	1	\$134.7	4 \$29	3 9	0
5	52014	20080101	2		199	254	11	1	\$167.7	3 \$32	9 9	0
5	52038	20080101	2		199	1615	11	1	\$96.00	\$289.9	9 9	0
5	52126	20080101	2		199	206	11	1	\$261.6	5 \$56	9 9	0
5	52193	20080102	2		199	1104	11	1	\$141.4	7 \$42	7 9	0
5	52214	20080102	2		199	781	11	1	\$7.5	9 \$16.	5 9	0
5	52222	20080102	2		199	2442	11	1	\$5.0	\$9.9	9 9	0
5	52238	20080102	2		199	737	11	1	\$73.1	2 \$15	9 9	0
5	52389	20080102	2		199	1453	11	1	\$118.6	5 \$25	9	0
5	52424	20080103	2		199	2174	11	1	\$75.9	5 \$14	9 9	0
5	52447	20080103	2		199	1628	11	1	\$6.3	9 \$13.8	9 9	0
5	52633	20080103	2		199	1998	11	1	\$71.3	7 \$139.9	9 9	0
5	52659	20080104	2		199	1782	11	1	\$21.9	2 \$4	9	0
5	52736	20080104	2		199	310	11	1	\$152.4	\$29	9 9	0
5	52850	20080104	2		199	552	11	1	\$827.9	7 \$2,49	9 9	0
5	52895	20080105	2		199	2175	11	1	\$66.2	3 \$129.	9 9	0
5	52946	20080105	2		199	794	11	1	\$6.0	7 \$11.	9 9	0
Store Sales: \$6,960,345,4	86.20											
_DimChannel dbo_DimDa	te dbo	_DimEntity dbo_Dim	Product dbo_	DimProdu	ctSubcatego	ry dbo_FactSales Pro	ductCategory a	ContosoStores St	ores			
rd: I4 4 1 of 3,962,476	1 H											

Figure 7-12. Formatting as currency

Calculating the Store Sales for 2009

Calculated fields can be based on other calculated fields. The store sales for 2009 can be calculated based on the total StoreSales calculated field. Follow these steps:

- 1. Select the next cell in the leftmost calculation area.
- 2. Enter the following formula:

```
StoreSales2009:=CALCULATE([StoreSales],LEFT(dbo_DimDate[Datekey],4)="2009")
```

The LEFT function is used to extract the year from the Datekey text string. The result is shown in Figure 7-13.

🗐 🖬 🐺 to + ct + = 1			PowerP	Pivot for Excel - Cop	of PowerPivotTutorialSa	nplecha	o7.xlsx			- 6 ×
File Home Design										^ ()
Paste Append Paste Replace Copy	m Data From Other	Existing Refresh	PivotTable	Data Type : - Format : Currency - \$ - % > 28 4%	21 Sort A to Z 21 Sort Z to A Clear All Sort b Filters Column	y Find	∑ AutoSum	a Data View	Diagram Show Hidden Area	
Clipboard	Get External Data			Formatting	Sort and Filter	Find	Calculation	6	View	
[SalesKey] - X 🗸 f _x Stores	Sales2009:=CALCUL	ATE([StoreSales],LE	FT(dbo_Di	mDate[Datekey],4]=":	2009")					¥
🚄 SalesKey 💽	Dat CALCU	ATE(Expression, [Hiter)), [Eiter2], [Eit	Produc., 🖈 🔽 Pron	otionKey 💽 CurrencyKey	🖬 Uni	tCost 💌 Uni	tPrice 💌	SalesQuantity 🗷 ReturnO	luantity 💌 Return
551966	20080101	2	199	2018	11	1	\$48.43	\$94.99	9	0
551995	20080101	2	199	1452	11	1	\$134.74	\$293	9	0
552014	20080101	2	199	254	11	1	\$167.73	\$329	9	0
552038	20080101	2	199	1615	11	1	\$96.08	\$289.99	9	0
552126	20080101	2	199	206	11	1	\$261.66	\$569	9	0
552193	20080102	2	199	1104	11	1	\$141.47	\$427	9	0
552214	20080102	2	199	781	11	1	\$7.59	\$16.5	9	0
552222	20080102	2	199	2442	11	1	\$5.09	\$9.99	9	0
552238	20080102	2	199	737	11	1	\$73.12	\$159	9	0
552389	20080102	2	199	1453	11	1	\$118.65	\$258	9	0
552424	20080103	2	199	2174	11	1	\$75.96	\$149	9	0
552447	20080103	2	199	1628	11	1	\$6.39	\$13.89	9	0
552633	20080103	2	199	1998	11	1	\$71.37	\$139.99	9	0
552659	20080104	2	199	1782	11	1	\$21.92	\$43	9	0
552736	20080104	2	199	310	11	1	\$152.44	\$299	9	0
552850	20080104	2	199	552	11	1	\$827.97	\$2,499	9	0
552895	20080105	2	199	2175	11	1	\$66.23	\$129.9	9	0
552946	20080105	2	199	794	11	1	\$6.07	\$11.9	9	0 v
StoreSales: \$6,960,345,486.20										^
StoreSales2009: \$1,954,384,680.86]									
testsumx: \$12,617,814,712.97	2									
<										>
dbo_DimChannel dbo_DimDate dbo_D	imEntity dbo_DimPro	duct dbo_DimProdu	ct5ubcatego	ry dbo_FactSales Prod	uctCategory IN ContosoStores	Stores				
Record: H 4 1 of 3,962,476 > H										

Figure 7-13. Calculating the 2009 store sales

3. Next, create a calculated field for store sales for the previous year, 2008, by entering the following DAX formula in the cell below, as shown in Figure 7-14.

StoreSales2008:=CALCULATE([StoreSales],LEFT(dbo_DimDate[Datekey],4)="2008")

			PowerP	vivot for Excel - Cop	y of PowerPivotTutorialSam	plechap	o7.xlsx			- 0
Pome Design Paste Append Paste Replace Copy Clipboard Get	Data From Other Existing te * Sources Connection	Refresh	PivotTable	Data Type : - Format : Currency • \$ • % > 28 48 Formatting	2 Sort A to Z 3 Sort Z to A Clear Sort Fitters Column Sort and Fitter	Find	∑ AutoSum Create KPI Calculations	Data View	Diagram Show Calculation View Hidden Area	
salesKey] • fx StoreSa	es2008:=CALCULATE([S	toreSales],L	EFT(dbo_Di	mDate[Datekey],4)="	2008")					
SalesKey 💌 I	Dat 👘 💌 channe	🐨 🖾 Stor		Produc 👫 💌 Pror	notionKey 🗷 CurrencyKey	Uni Uni	tCost 📰 Uniti	rice 🔳	SalesQuantity 🗷 ReturnQua	ntity 🖬 Ref
551966	20080101	2	199	2018	11	1	\$48.43	\$94.99	9	0
551995	20080101	2	199	1452	11	1	\$134.74	\$293	9	0
552014	20080101	2	199	254	11	1	\$167.73	\$329	9	0
552038	20080101	2	199	1615	11	1	\$96.08	\$289.99	9	0
552126	20080101	2	199	206	11	1	\$261.66	\$569	9	0
552193	20080102	2	199	1104	11	1	\$141.47	\$427	9	0
552214	20080102	2	199	781	11	1	\$7.59	\$16.5	9	0
552222	20080102	2	199	2442	11	1	\$5.09	\$9.99	9	0
552238	20080102	2	199	737	11	1	\$73.12	\$159	9	0
552389	20080102	2	199	1453	11	1	\$118.65	\$258	9	0
552424	20080103	2	199	2174	11	1	\$75.96	\$149	9	0
552447	20080103	2	199	1628	11	1	\$6.39	\$13.89	9	0
552633	20080103	2	199	1998	11	1	\$71.37	\$139.99	9	0
552659	20080104	2	199	1782	11	1	\$21.92	\$43	9	0
552736	20080104	2	199	310	11	1	\$152.44	\$299	9	0
552850	20080104	2	199	552	11	1	\$827.97	\$2,499	9	0
552895	20080105	2	199	2175	11	1	\$66.23	\$129.9	9	0
552946	20080105	2	199	794	11	1	\$6.07	\$11.9	9	0
storeSales2009: \$1,954,384,680.86										
estsumx: \$12,617,814,712.97										
storeSales2008: \$2,225,161,671.19										

Figure 7-14. Calculating the 2008 store sales

4. Next, the year-over-year growth can be calculated by entering the following formula and changing the format to percent, as shown in Figure 7-15. The identifiers for the calculated fields have to be enclosed in square brackets. Note that sales decreased by 12% in 2009.

te Daste Replace	rom Data From Other Service * Sources	Existing Refre	sh PivotTable	Data Type : * Format : * \$ • % • 200 +00	Sort A to Z	lear All Sort by	Find	∑ AutoSum ☐ Create KPI	Data View	Diagram Show View Hidden	Calculation Area	
Clipboard	Get External Data			Formatting	Sort an	d Filter	Find	Calculations		View		
SalesKey] ▼ fx Sale	sChange:=([Store	Sales2009]-[StoreS	ales2008])/[St	oreSales2008]								
SalesKey	🖬 Dat 🛸 🖬	channe 🚷 💌 S	tor 👘 🖬	Produc 👘 🖬 Pron	notionKey 💌	CurrencyKey 💌	UnitCo	st 🔽 Uniti	Price 🖬	SalesQuantity	ReturnQu	antity 🖬 Re
55196	6 20080101	2	199	2018	11	1		48.43	\$94.99		9	0
55195	15 20080101	2	199	1452	11	1	\$1	34.74	\$293		9	0
55201	4 20080101	2	199	254	11	1	\$1	67.73	\$329		9	0
55203	8 20080101	2	199	1615	11	1	\$	96.08	\$289.99		9	0
55212	20080101	2	199	206	11	1	\$3	61.66	\$569		9	0
55215	3 20080102	2	199	1104	11	1	\$1	41.47	\$427		9	0
55221	4 20080102	2	199	781	11	1		\$7.59	\$16.5		9	0
55222	20080102	2	199	2442	11	1		\$5.09	\$9.99		9	0
55223	8 20080102	2	199	737	11	1	:	73.12	\$159		9	0
55238	9 20080102	2	199	1453	11	1	\$1	18.65	\$258		9	0
55242	20080103	2	199	2174	11	1	:	75.96	\$149		9	0
55244	7 20080103	2	199	1628	11	1		\$6.39	\$13.89		9	0
55263	3 20080103	2	199	1998	11	1	:	71.37	\$139.99		9	0
55265	9 20080104	2	199	1782	11	1		21.92	\$43		9	0
55273	6 20080104	2	199	310	11	1	\$1	52.44	\$299		9	0
55285	0 20080104	2	199	552	11	1	\$8	27.97	\$2,499		9	0
55285	5 20080105	2	199	2175	11	1	4	66.23	\$129.9		9	0
55294	6 20080105	2	199	794	11	1		\$6.07	\$11.9		9	0
estsumx: \$12,617,814,712.97												
toreSales2008: \$2,225,161,671.1	9											
alesChange: -12.17%												

SalesChange:=([StoreSales2009]-[StoreSales2008])/[StoreSales2008]

Figure 7-15. Calculating a year or a year sales change

Creating a KPI for Profitability

A KPI, or Key Performance Indicator, is based on the relationship between a calculated field and a value. Create a KPI for profitability by initiating the following steps:

1. Create a calculated field for average of profit/sales, as shown in Figure 7-16, by entering the following formula:

Average of Profit/Sales:=AVERAGE([Profit/Sales])

			PC	WEIPIVOLI	of Excel - Copy of P	owerProoffull	maisampiech	aby (version e)	130		
File Hor	ne Design										^
Paste Pas Paste Pas Clipbol	te Append te Replace 17 Database -	From Data From Other Service - Sources Connect Get External Data	ng Refresh	PivotTable	Data Type : - Format : Percentage * \$ • % • 28 4% Formatting	2↓ Sort A to Z 2↓ Sort Z to A 2↓ Clear Sort Sort a	Clear All Sort by Filters Column	Find Calc	vtoSum • Data Diagra View View	Min Show Calculation Hidden Area	
Profit/Sale:] - fr Ave	rage of Profit/Sales:=AVE	RAGE([Profit/:	Sales])							
ReturnA	mount 💌 Discount	Quantity 💌 DiscountAm	iount 🖬 Tot	alCost 💌	TotalSales 🔳 E	TLLoadID 🖬	LoadDate 🖬	UpdateDate	TotalProfit	Profit/Sales	· Add Co
)	\$0	0	\$0	\$435.87	\$854.91	6	1/1/2008 12	1/1/2008 12:0	\$419.04	4	9.02 %
5	\$0	0	\$0	\$1,212.66	\$2,637	6	1/1/2008 12	1/1/2008 12:0	\$1,424.34	5	4.01 %
)	\$0	0	\$0	\$1,509.57	\$2,961	6	1/1/2008 12	1/1/2008 12:0	\$1,451.43	4	9.02 %
)	\$0	0	\$0	\$864.72	\$2,609.91	6	1/1/2008 12	1/1/2008 12:0	\$1,745.19	6	6.87 %
3	\$0	0	\$0	\$2,354.94	\$5,121	6	1/1/2008 12	1/1/2008 12:0	\$2,766.06	5	4.01 %
)	\$0	0	\$0	\$1,273.23	\$3,843	6	1/2/2008 12	1/2/2008 12:0	\$2,569.77	6	6.87 %
3	\$0	0	\$0	\$68.31	\$148.5	6	1/2/2008 12	1/2/2008 12:0	\$80.19	5	4.00 %
)	\$0	0	\$0	\$45.81	\$89.91	6	1/2/2008 12	1/2/2008 12:0	\$44.1	4	9.05 %
3	\$0	0	\$0	\$658.08	\$1,431	6	1/2/2008 12	1/2/2008 12:0	\$772.92	5	4.01 %
3	\$0	0	\$0	\$1,067.85	\$2,322	6	1/2/2008 12	1/2/2008 12:0	\$1,254.15	5	4.01 %
3	\$0	0	\$0	\$683.64	\$1,341	6	1/3/2008 12	1/3/2008 12:0	\$657.36	4	9.02 %
3	\$0	0	\$0	\$57.51	\$125.01	6	1/3/2008 12	1/3/2008 12:0	\$67.5	5	4.00 %
3	\$0	0	\$0	\$642.33	\$1,259.91	6	1/3/2008 12	1/3/2008 12:0	\$617.58	4	9.02 %
)	\$0	0	\$0	\$197.28	\$387	6	1/4/2008 12	1/4/2008 12:0	\$189.72	4	9.02 %
3	\$0	0	\$0	\$1,371.96	\$2,691	6	1/4/2008 12	1/4/2008 12:0	\$1,319.04	4	9.02 %
3	\$0	0	\$0	\$7,451.73	\$22,491	6	1/4/2008 12	1/4/2008 12:0	\$15,039.27	6	6.87%
3	\$0	0	\$0	\$596.07	\$1,169.1	6	1/5/2008 12	1/5/2008 12:0	\$573.03	4	9.01 %
)	\$0	0	\$0	\$54.63	\$107.1	6	1/5/2008 12	1/5/2008 12:0	\$52.47	4	8.99 %
									Sum of TotalProfit	Sum of Profit/Sales: 2112	3735
									Average of TotalPr	Average of Profit/Sales: 5	3.31 🖽
										2	
											>

Figure 7-16. Creating calculate field for Average of Profit/Sales

2. Right-click the calculated field and select Create KPI. Enter an absolute value of .53, or 53%, and adjust the sliders so that green is displayed for values that meet the target, as shown in Figure 7-17, and click OK.

		k	Key Perfo	ormance	Indicato	r (KPI)		?	×
KPI base field (value):	Average	of Profit/Sa	ales					V	
KPI <u>S</u> tatus									
Define target value:									
O Calculated Field:								Y	
<u>A</u> bsolute value:	0.53]
Define status threshol	ds:								
			•			0.4	0.53	•	
↔									
							Target		
Select icon style:						L			
7	8		×	•	슔	at at		00	
9 8	0		Ŷ	$\mathbf{\Delta}$	53	att att		•	
٣	0		1		\$	000		0	
A Descriptions									
							ОК	Cano	el

Figure 7-17. Defining the KPI

Creating a Pivot Table Showing Profitability by Product Line

To create a Pivot Table showing profitability by product line, follow these steps:

- 1. Select the dbo_FactSales tab.
- 2. Click the Pivot Table icon on the ribbon and select Pivot Table.
- 3. Select New Worksheet in the Create Pivot Table dialog and click OK.
- 4. Select the Product Category Name field in the Product Category table. It will appear in the Rows box in the Pivot Table pane on the right of the screen, as shown in Figure 7-18.

	5· c· 🔳 👂	× 💷 - 🔹		Co	py of PowerP	ivotTutorialS	amplechap7 - I	Excel			PIVOT	ABLE TOOLS	1100			? 0 - 8 2
FILE	HOME INSERT	PAGE LAYOUT	FORMULA	S DATA	REVIEW	VIEW	POWER QU	ERY PO	OWERPIVOT	Team	ANALYZ	E DESIGN				Neil - 🖸
Manage Data Mode	Calculated KPIs Fields * * Calculations	C Align Vertically	Add to Data Moo Ta	Update del All bles	Detect Relationships	Settings										
83	• : × ·	✓ f _x Row La	bels													
1		В	с	D	E	F	G	н	i.	J	к	L M	N	-	PivotTable F	ields ->
2	David shale														ACTIVE ALL	
4	AUDIO														Choose fields to add	to report: 🚯 🔻
5	CAMERAS & CA	MCORDERS													▷ Lia dbo FactSales	
6	CELL PHONES														En luci	<u> </u>
7	COMPUTERS														ProductCateg	jory tegoryKey
8	GAMES & TOYS														ProductCa	tegoryLabel
10	MUSIC MOVIES	& AUDIO BOOKS													✓ ProductCa	tegoryName
11	TV & VIDEO	a Aubio Books													ProductCa	tegoryDescription -
12	Grand Total															
13															Drag fields between	areas below:
14															T CHITCOL	LE couner
15															T PILTERS	III COLUMINS
16																
17																
18															E ROWS	Σ VALUES
19															ProductCate •	
20																
21																
4 1	Sheet4 S	heet1 Sheet2	Sheet5	Sheet3	Data for Pr	ofit by Cate	egor (+)	1						•	Defer Layout Upo	date UPDATE
READY														m	圖 町	+ 107

Figure 7-18. Inserting product category name into Pivot Table

5. Next, select the Sales Quantity field in the dbo_FactSales table and the Average of Profit/Sales calculated field, which appears at the bottom of the fields list. Click the calculated field name and select values, goal, and status. This field will appear in the values box, and a Pivot Table will be created, as shown in Figure 7-19.

Manage C	Align Vertically	y Add to Update Data Model All Tables R	Detect Settings				
87	• : 🗙 🗹 f_x сомя	PUTERS					
A A 1 2	8	c	D	E	F	G	PivotTable Fields
3	Row Labels 👻	Sum of SalesQuantity	Average of Profit/Sales	Average of Profit/Sales Goal	Average of Profit/Sales Status		Choose fields to add to report:
4	AUDIO	1353298	53.69 %	0.53			
5	CAMERAS & CAMCORDERS	6551189	56.25%	0.53			God
7	CELL PHONES	13439636	54.25 %	0.53			C Statur
8	GAMES & TOYS	6040572	52.10%	0.53			A R Auerran of Profit/Syler
0	HOMEAPPLIANCES	10379472	51.65%	0.53	ă		Value (Average of Profi
10	MUSIC, MOVIES & AUDIO BOOKS	1523415	56.70 %	0.53			Goal
11	TV & VIDEO	3449196	51.90 %	0.53	õ		✓ Status
12	Grand Total	53320454	53.31 %	0.53	•		
13							
4							Drag fields between areas below:
15							T FILTERS
16							T Values
17							L. vaides
8							
9							E ROWS E VALUES
20							ProductCate
21							Autored T
22							Average of
	Sheet1 Sales by Channel	Profit by Category	ContosoStores Data f	🔶 🕴 📢		Þ	Defer Layout Update UPDAT
	ouraça.						700 EE1 .

Figure 7-19. Adding the KPI into Pivot Table

Summary

This chapter showed how to use the DAX formula language to create calculated fields and calculated columns and how to create Key Performance Indicators to monitor business activity. In contrast to Excel formulas, which reference specific cells, DAX formulas reference columns, but the CALCULATE() function can be used to calculate an expression based on a filter.

CHAPTER 8

Power Query

Microsoft Power Query for Excel is a data-discovery tool for importing data from a wide variety of sources, including relational, semi-structured, and unstructured data available from sources such as OData, Hadoop, and Azure Marketplace. It is a self-service ETL (export, transform, load) tool for extracting, transforming, and loading data. It is more powerful than some of the approaches that have been covered previously. Microsoft is constantly updating Power Query, so some features may have changed by the time this book is published. This chapter demonstrates how to access data—from the Web, in table format, CSV format, and JSON format—and then shape it using the Power Query editor before importing it into Excel.

Installing Power Query

If the Power Query tab is not visible, you will have to install it. Power Query is available as a free download from Microsoft for Excel 2010 and 2013. The download link can be accessed by doing a search on "Power Query download." Be sure to install the 32-bit version if you have installed the 32-bit version of Excel and the 64-bit version if you have the 64-bit version of Excel. To enable Power Query after installation, follow these steps:

- 1. Click the File tab to see the Backstage View.
- 2. Click Options at the bottom of the left pane.
- 3. Select Add-Ins on the left pane.
- 4. At the bottom of the window, pull down the arrow after Manage and select Com Add-Ins and click Go.
- 5. Check Microsoft Office PowerPivot for Excel 2013, Microsoft Power Map for Excel, and Power View, and click OK.

After installation, the Power Query tab will be visible.

Key Options on Power Query Ribbon

This section will cover the most useful options on the Power Query ribbon, which is shown in Figure 8-1.



Figure 8-1. Power Query ribbon

- From Web: Allows extracting tables from a specific web site
- From File: Imports from Excel, CSV, XML, text files, or from a group of files in a folder
- *From Database*: Imports from databases, such as SQL Server, Access, Oracle, and MySQL
- *From Azure*: Imports from Azure services, such as SQL Database, Market Place, HDInsight, Blob Storage, and Table Storage
- *From Other Sources*: Imports from sources including SharePoint Lists, OData feeds, Hadoop (HDFS), Facebook, Salesforce Objects and Reports, and ODBC
- Recent Sources: Shows the source of recent queries
- From Table Excel Data: Imports a range from an Excel spreadsheet
- Merge: Merges two queries to add columns
- Append: Appends one table to another with the same structure
- *Show Pane*: Toggles the display of the Workbook Query pane on the right side of the screen
- *Data Source Settings*: Manages setting for data sources that have been accessed through Power Query
- Options: Sets options for loading data
- Update: Checks online for Power Query updates
- Data Catalog Search: Searches a curated set of web sites and extracts tables
- My Data Catalog Queries: Allows sharing queries with other users

Working with the Query Editor

Power Query allows importing a subset of the data into the Query Editor, so that the data can be shaped by removing unneeded rows or columns or creating calculated columns before importing into Excel or into a data model. There are three tabs in Query Editor: Home, Transform, and Add Column.

Key Options on the Query Editor Home Ribbon

This section covers key options on the Query Editor Home ribbon, which is shown in Figure 8-2.

Volume Volume	Keep Remove ows • Rows • Reduce	Remove Duplicates Remove Errors * Rows	s 21 ZI Split Column •	Data Type: Tex Group By 12 Replace Vi Transform	t ▼ low As Headers ▼ alues	Merge Queries Append Queries Combine Binaries Combine New Query
X Image: Constraint of the second s	Area (km²) 134,940,000 148,940,000 510,072,000	/Web/Tables/890 Area (mi²) 52,100,000 57,510,000 196,940,000	Population ▼ 7,164,553,400 7,164,553,400 7,164,553,400 7,164,553,400	75bd66059e0842/V Density (pop./km²) 53 48 14	1/Data") Density (pop./mi 137 124 36	Query Settings P A PROPERTIES Name Secondary table - List of sovereign states
Clear Soft	143,998 36,191	55,598 13,973	142,325,250 23,379,129	988.38 646	2,559.9 1,673	All Properties APPLIED STEPS
Search	99,461 25,314 41,526	9,774 16.033	50,219,669 10,537,222 16,838,149	416	1,308	Source
⊘ Turkey ⊘ Uganda	27,065 3,287,240 22,072	10,450 1,269,211 8,522	10,413,211 1,244,392,079 8 188 576	385 378.55 370.99	997 980.45 960.87	
 ✓ Ukraine ✓ United Arab Emirates ✓ United Kingdom 	30,528 27,816	11,787 10,740	11,194,824 10,163,000	367 365	951 945	
United States Uzbekistan Venezuela	377,915 300,000 65.610	145,914 115,831 25,332	127,100,000 99,598,451 20,277,597	336 331.99 309	870 859.86 800	
Vietnam World (land only)	331,212 242,910	127,882 93,788	88,780,000 63,705,000	268 262	694 679	
World (land only, excluding Antarctica)	803,940 357,123	310,403 137,886	186,517,533 80,716,000	232 226	600.89 585	
✓ Yemen ✓ Zambia	301,308 47,875	116,336 18,485	60,021,955 9,445,281	199 197	515 510	~

Figure 8-2. Query Editor screen

- *Close & Load:* Closes the editor and loads the data either into an Excel worksheet or into the data model
- *Refresh Preview*: Refreshes the query, reprocessing all steps, and previewing the results
- Advanced Editor: Shows the code used to extract the data in the M query language
- *Choose Columns/Remove Columns*: Allows keeping or removing selected columns. Clicking the Choose Columns icon will show a list of columns that can be unchecked to remove, as shown in Figure 8-3.

				,
Choose Column	S			
choose column				
Choose the columns to	кеер			
Search Columns]
✓ (Select All Columns)				
✓ Ticker symbol				
Security				
SEC filings				
GICS Sector				
GICS Sub Industry				
Address of Headqua	inters			
Date first added				
CIK				
✓ Key				
	0	К	Cancel	
				5

Figure 8-3. Choose Columns dialog

- Keep Rows/Remove Rows: Allows keeping or removing selected rows
- *Remove Duplicates*: Removes duplicates. Be careful about defining the criteria for duplicates. For example, in a list of stock prices by day, removing duplicates based only on the stock symbol would leave only one unique row for each stock.
- *Sort*: Sorts columns in ascending or descending order. Remember that the Query Editor is only working a subset of the data. It might be better to wait until the data is loaded into a worksheet or data model before sorting.
- Split Column: Splits columns based on a delimiter, such as a comma or space
- Group By: Groups rows based on a field or fields

Note that another way to remove or filter out rows is by clicking the arrow to the right of the column heading and de-selecting values, as shown in Figure 8-2.

A Simple Population

This example will show how to use the Data Catalog Search option to retrieve population data in a tabular format for analysis. The Data Catalog Search option searches a curated data catalog of public data web sites, such as Wikipedia and the Census Bureau web site, and allows tables to be easily extracted from those web sites. It is not a full-featured search engine. It seems to understand searches that start with words such as *largest* or *most*. It can also search corporate databases after being logged in to the corporate network. For this example, take the following steps.

- 1. On the Power Query tab, click Data Catalog Search.
- 2. In the Search pane to the right, enter "most populated countries." Hover the mouse pointer over the results to get a preview of the data. Then hover the mouse over the item in the list entitled "Secondary table—List of sovereign states and dependent territories by population density" from Wikipedia. The results are shown in Figure 8-4. Note the prompts at the bottom for Load and Edit.

RIE H	HOME INSE	RT PA	GE LAYOUT	FORMU		Book	SFARCH PC	WERPIVOT To	am	×	? 10 - 6 × Neil - P
A1 A 1 A 1 A 1 A 4 A 4 A 4 A 4 A 4 A 4 A	• : ×	с .	D	E	Pos.	Inst y table – List OF SOVE (By 2P, 2014) reign states and dependent territories by popu County (or dependent territory) World (land only, excluding Antarctica) World (land only) World (land enly) Bangladesh	lation density - Wik Area (km ²) 134,940,000 148,940,000 510,072,000 143,998	Area (mi ²) 52,100,000 57,510,000 196,940,000 55,598	rclopedia Population 7,164,553,400 7,164,553,400 7,164,553,400 142,325,250	De	Online Search ▼× (●) [most populated countried]] PUBLIC (1922)
5 6 7 8 9					2 3 4 5 6	Taiwan South Korea Rwanda Netherlands Haiti	36,191 99,461 25,314 41,526 27,065	13,973 38,402 9,774 16,033 10,450	23,379,129 50,219,669 10,537,222 16,838,149 10,413,211	Ŷ	w Secondary table - List of sov From Wikipedia, last modified on Thurs List of sovereign states and dependent territories by population density
10 11 12 13 14 15 16					Column Pos, Coun Population Data So the http://e	s [10] try (or dependent territory), Area (km ⁹), Area (n source, Key UFCes [1] m.wikipedia.org/wiki/List_of_sovereign_states_t	ni ^a), Population, Der and_dependent_terri	isity (pop./km²), Der tories_by_populatio	nsiły (pop./mi ^a), Date, n_densiły		 W Main Lable – List Of SOVPERIG From Wikipedia, last modified on Thurs List of sovereign states and dependent territories by population density – G FIG net emissions/removal From The World Bank, World Developm GHG net emissions/removals by LUCF
17 18 19 20 21 22					LOAD -	EDIT			LEGAL TE	RMS	refers to changes in atmospheric levels of W 10 most densely populated From Wikipedia, last modified on Thurs World population - Wikipedia, the free encyclopedia
23 24 25 26	Sheet1	۲								•	w With population above 1 mil From Wikipedia, last modified on Thurs ✓ 1 2 3 4 5 Next

Figure 8-4. Preview of population data

3. Click Edit on the lower left of the preview to load a subset of data from the table into the Query Editor, as shown in Figure 8-5. The table lists countries with their population and areas in square miles or square kilometers and shows their population density per square mile or square kilometer.

XI 🙂				Secon	ndary t	able - Lis	t of sovereigr	states and	depe	ndent territories b	y populati - Query Edit	tor	- 0	×
	Home	Transfo	rm Add C	olumn	View									^ Ø
Close & Load •	Refresh Preview Query	Properties	Choose Re Columns Col Mana	emove umns -	Rename Move •	Keep Rows *	Remove Rows • Reduce Rov	ove Duplicates ove Errors •	ĝ↓ ∡↓ Sort	Split Column - By	ata Type: Text • E Use First Row As Headers • 2 Replace Values fransform	Merge Queri Append Que Combine Bir Combine	ies 😲 Send Feedback • eries 🕜 Help naries 🗊 About Help	
E- Po	. v	Country (or d	ependent territ	ory)	T A	rea (km²)	Y Area (mi ²)	Populati	on	▼ Density (pop./km²	Density (pop./mi ²)	▼ Date	-	
1		World (land o	nly, excluding A	intarctica)	1	34,940,000	52,100,000	7,164,55	3,400	53	137	May 22, 201	Query Settings	×
2		World (land o	nly)		1	18,940,000	57,510,000	7,164,55	3,400	48	124	May 22, 201	A PROPERTIES	
3		World (with w	vater)		5	10,072,000	196,940,000	7,164,55	3,400	14	36	May 22, 201	Name	
4 1		Bangladesh			1	13,998	55,598	142,325,	250	988.38	2,559.9	May 22, 201	Secondary table - List of	sove
5 2		Taiwan			3	5,191	13,973	23,379,1	29	645	1,673	February 28,	All Properties	
6 3		South Korea			9	9,461	38,402	50,219,6	69	505	1,308	July 1, 2013		
7 4		Rwanda			2	5,314	9,774	10,537,2	22	416	1,077	August 15, 2	- Arreleo sters	
8 5		Netherlands			4	1,526	16,033	16,838,1	49	405.48	1,050.2	May 22, 201	Source	8
9 6		Haiti			2	7,065	10,450	10,413,2	11	385	997	2012		
10 7		India			3	287,240	1,269,211	1,244,39	2,079	378.55	980.45	May 22, 201		
11 8		Israel			2	2,072	8,522	8,188,57	6	370.99	950.87	May 22, 201		
12 9		Belgium			3	0,528	11,787	11,194,8	24	367	951	April 1, 2014		
13 10		Burundi			2	7,816	10,740	10,163,0	00	365	945	July 1, 2013		
14 11		Japan			3	77,915	145,914	127,100,	000	336	870	May 1, 2014		
15 12		Philippines			3	00,000	115,831	99,598,4	51	331.99	859.86	May 22, 201		
16 13		Sri Lanka			6	5,610	25,332	20,277,5	97	309	800	March 21, 20		
17 14		Vietnam			3	31,212	127,882	88,780,0	00	268	694	2012		
18 15		United Kingdo	om		2	\$2,910	93,788	63,705,0	00	262	679	July 1, 2012		
19 16		Pakistan			8	3,940	310,403	186,517,	533	232	600.89	May 22, 201		
20 17		Germany			3	57,123	137,886	80,716,0	00	226	585	September 3		
21 18		Italy			3	01,308	116,336	60,021,9	55	199	515	November 3		
22 10		Realizing Re				10.10	10 105	A 445 30		107		D		

Figure 8-5. Population data in Query Editor

4. In the Query Editor, data can be manipulated by deleting rows or columns, splitting or merging columns, or removing duplicates. For example, to delete the columns for area or density in kilometers, select both columns by clicking the header of the first column to be deleted and Ctrl-click the header of the second column, as shown in Figure 8-6.

XI	1 😷	I		Secondar	y table - List	of sovereign	states and depend	ent territories by po	pulati - Query Edito	r	- 🗆 🗙
	F	Home	Transform Ac	d Column View	5						^ e
Ci ci	ose &	Refresh Preview •	Advanced Editor	Choose Remove Columns Columns Manage Co	Move •	Keep Remove Rows * Rows *	E-BRemove Duplicates	Sort	Data Type: Text • Use First Row As H Use First Row As H 12 Replace Values Transform	eaders • Appenders • Combin	Queries d Queries ne Binaries bine
-	Pos	. 💌 Co	ountry (or dependent t	erritory)	Area (km²)	✓ Area (mi ²)	Population	Density (pop./km²)	Density (pop./mi²)	Date	Population source
1		W	orld (land only, excludi	ng Antarctica)	134,940,000	52,100,000	7,164,553,400	53	137	May 22, 2014	USCB's World population c
2		W	orid (land only)		148,940,000	57,510,000	7,164,553,400	48	124	May 22, 2014	USCB's World population c
3		W	orld (with water)		510,072,000	196,940,000	7,164,553,400	14	36	May 22, 2014	USCB's World population c
-4	1	Ba	ingladesh		143,998	55,598	142,325,250	988.38	2,559.9	May 22, 2014	Official estimate
5	2	Ta	iwan		36,191	13,973	23,379,129	646	1,673	February 28, 2014	Monthly official estimate
6	3	So	outh Korea		99,461	38,402	50,219,669	505	1,308	July 1, 2013	Official estimate
7	4	Rv	vanda		25,314	9,774	10,537,222	416	1,077	August 15, 2012	Preliminary 2012 census re
8	5	Ne	etherlands		41,526	16,033	16,838,149	405.48	1,050.2	May 22, 2014	Official population clock
9	6	Ha	aiti		27,065	10,450	10,413,211	385	997	2012	Official estimate
10	7	Inc	dia		3,287,240	1,269,211	1,244,392,079	378.55	980.45	May 22, 2014	India Population clock
11	8	Isr	ael		22,072	8,522	8,188,576	370.99	960.87	May 22, 2014	Official population clock
12	9	Be	lgium		30,528	11,787	11,194,824	367	951	April 1, 2014	Monthly official estimate
13	10	Bu	ırundi		27,816	10,740	10,163,000	365	945	July 1, 2013	UN Estimate
14	11	Jaj	pan		377,915	145,914	127,100,000	336	870	May 1, 2014	Monthly official estimate
15	12	Ph	llippines		300,000	115,831	99,598,451	331.99	859.86	May 22, 2014	Official population clock
16	13	Sri	i Lanka		65,610	25,332	20,277,597	309	800	March 21, 2012	Preliminary 2012 census re
17	14	Vi	etnam		331,212	127,882	88,780,000	268	694	2012	Official estimate
18	15	Ur	nited Kingdom		242,910	93,788	63,705,000	262	679	July 1, 2012	Official estimate
19	16	Pa	kistan		803,940	310,403	186,517,533	232	600.89	May 22, 2014	Official population clock
20	17	Ge	ermany		357,123	137,886	80,716,000	226	585	September 30, 2013	Monthly official estimate
21	18	Ita	ily		301,308	116,336	60,021,955	199	515	November 30, 2013	Monthly official estimate
22	19	Do	ominican Republic		47,875	18,485	9,445,281	197	510	December 1, 2010	Final 2010 census result
23	20	No	orth Korea		122,762	47,399	24,052,231	196	508	October 1, 2008	Final 2008 census result
24	21	Sv	vitzerland		41,285	15,940	8,136,700	197	510	December 31, 2013	Quarterly official estimate
	C										>

Figure 8-6. Highlighting columns for deletion

- 5. Then click Remove Columns under Remove Columns on the tab, to remove the selected columns.
- **6.** After the data is shaped the way you want, click Close & Load, to load the data into an Excel spreadsheet.

Performance of S&P 500 Stock Index

This example will show how to retrieve S&P 500 performance data, by taking the following steps:

- 1. On the Power Query ribbon, click Data Catalog Search.
- 2. In the search box, type "S&P 500 performance." Hover the mouse over the first link, labeled "Total annual returns...," as shown in Figure 8-7.

	* :	XV	fx		290 - 200 -	Wikipedia, the free ency	clopedia			
A I	8	с	D	E	Year	Change in Index	Total Annual Return Including Dividends	Value of \$1.00 Invested on 1970	-01-01	the a p con
1					1970	0.01%	4.01%	\$1.04	^	Online Search
2					1971	10.79%	14.31%	\$1.19	10	(a) cha 500 automasca
3					1972	15.63%	18.96%	\$1.41		G sop soo performance
4					1973	-17.57%	- 14.00%	31.21		PUBLIC [4676]
5					1974	-28.7276	= 20.475	20.89		w Total appual raturos Mark
6					19/3	10 168	31.CU%	\$1.62		w rotal annual returns = Mark
7					1970	11.50%	-7.19%	\$1.51		S&D SOO - Wikingdia the free
8					1977	1.05%	= 7.10% 6 56%	\$1.40	~	encyclonedia
9					15/10	1.0076	6.50%	21/42		miledeu maietananza 520.500
10					`				/	Windex maintenance - Sor Sou
1					Column	is [10]				ShD 500 - Wikipedia, the free
2					Year, Char	nge in Index, Total Annua	al Return Including Dividends, Value of \$1.00 In	rested on 1970-01-01,		sop soo - wikipedia, the free
3					5 Year Ann	nualized Return, 10 Year	Annualized Return, 15 Year Annualized Return,	20 Year Annualized Return,		encyclopedia COD COD
4					25 Year Ar	nualized Return, Key				Winvesting - S&P 500
10					Data So	urces [1]				From Wikipedia, last modified on Thurs
0					+ http://	en.wikipedia.org/wiki/^(SSPC			S&P 500 - Wikipedia, the free
0										encyclopedia
0										w History - S&P 500
0					LOAD -	EDIT		LEGAL	TERMS	From Wikipedia, last modified on Thurs
21				1						S&P 500 - Wikipedia, the free
22										encyclopedia
23										w S&P 500 Component Stocks
4										From Wikipedia, last modified on Thurs
25										List of S&P 500 companies - Wikipedia,
									-	the free encyclopedia
26										

Figure 8-7. Data preview

3. Click Edit in the lower left to load the data showing S&P 500 returns for every year from 1970 to 2013 into the Query Editor, as shown in Figure 8-8.

X		=		Total	annual return	s - Market statistics	- S&P	500 - Que	ry Editor			- 🗆 🗙
		Home	Transform A	dd Column View								^ 📀
Clo	ie &	Refresh Preview •	Properties Advanced Editor	Choose Remove Columns Columns • Manage Columns	Keep Remov Rows • Rows	Remove Duplicates	A↓ Z↓ A↓ Sort	Split Column •	Data Type: Tex Data Type: Tex Use First R 12 Replace Va Transform	t • low As Heade alues	Combine	
	Year		Change in Index 💌	Total Annual Return Including Divide	ds 👻 Value	of \$1.00 Invested on 1970-	01-01	▼ 5 Year An	nualized Return	10 Year Ann	0.00	
1	1970)	0.01%	4.01%	\$1.04						Query Settings	×
2	1971	1	10.79%	14.31%	\$1.19				11	. ^	A PROPERTIES	
3	1972	2	15.63%	18.98%	\$1.41						Name	
4	1973	3	-17.37%	-14.66%	\$1.21					•	Total annual returns - Market	statistics - 58
5	1974	1	-28.72%	-26.47%	\$0.89			-2.35%			All Properties	
6	1975	5	31.55%	37.20%	\$1.22			3.21%			All Properties	
7	1976	5	19.15%	23.84%	\$1.51			4.87%			A APPLIED STEPS	
8	1977	,	-11.50%	-7.18%	\$1.40			-0.21%			Source	*
9	1978	8	1.06%	6.56%	\$1.49			4.32%				
10	1979	9	12.31%	18.44%	\$1.77			14.76%		5.86%		
11	1980)	25.77%	32.50%	\$2.34			13.96%		8.45%		
12	1981	1	-9.73%	-4.92%	\$2.23			8.10%		6.47%		
13	1982	2	14.76%	21.55%	\$2.71			14.09%		6.70%		
14	1983	3	17.26%	22.56%	\$3.32			17.32%		10.63%		
15	1984	1	1.40%	6.27%	\$3.52			14.81%		14.78%		
16	1985	5	26.36%	31.73%	\$4.64			14.67%		14.32%		
17	1986	5	14.62%	18.67%	\$5.51			19.87%		13.83%		
18	1987	,	2.03%	5.25%	\$5.80			15.47%		15.27%		
<	-									>		
•												
10 0	OLUN	INS, 48 R	ows								PREVIEW DOWNLOA	DED AT 11:34 AM.

Figure 8-8. Data in Query Editor

4. Select the columns by clicking Choose Columns and unchecking columns that will not be included. For this example, the 5-, 10-, 15-, and 20-year return columns will be removed, as shown in Figure 8-9. Click OK.

Search Columns				
(Select All Colum)	ns)			
✓ Year				
Change in Index				
✓ Total Annual Retu	urn Includ	ling Divid	lends	
✓ Value of \$1.00 In	rested on	1970-01	-01	
5 Year Annualized	Return			
10 Year Annualize	ed Return			
15 Year Annualize	ed Return			
20 Year Annualize	ed Keturn			
✓ 25 Year Annualize	ed Keturn			
e key				

Figure 8-9. Choose Columns dialog

5. To remove the four bottom rows, click Remove Rows and then Remove Bottom Rows and enter 4 as the number of rows to be removed, as shown in Figure 8-10, and click OK.

HO	(<mark>8</mark>) × ₹ (1		Total	annual returns - Market statistics -	S&P 500 - Query Editor		- 0				
Cio	se & Refres	ne Transform A	dd Column View	Image Columns Image Columns<							
m. 25 26 27 28 29 31 31 32 33 34 34 35 36 37 38	Year 1994 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007	Change in Index +1.54% 54.11% 20.26% 51.01% 52.657% 53.01% -10.14% -10.04% 58.58% 5.99% 5.05% 5.95%	Total Annual Return Including Divides 1.32% 27.58% 22.59% 33 24 25 25 25 26 27 27 27 28 28 29 29 29 29 29 29 29 29 29 29 29 20 20 20 20 20 20 20 20 20 20 20 20 20	Indi Value of \$1.00 Invested on 1970-1 \$1.5.1 \$18.59 \$22.86 ROWS to remove from the bottom.	01-01 • 25 Year Annualized Return • 10 98% 10 98% 12 22% 12 25%	Key V A PF N N X	uery Settings NOPERTES me bal annual returns - Market statistics Properties PUED STEPS Source Removed Other Columns Removed Bottom Rows				
39 40 41 42 43 44 45 46 47 48 ,	2008 2009 2010 2011 2012 2013 High Low CAGR Median	-38.47% 23.49% 12.64% 0.00% 13.29% 29.60% 34.11% -38.47% 12.64%	-3 28.49% 25.06% 2.13% 22.39% 37.58% -37.00% 10.40% 15.79%	543.11 549.61 550.65 558.76 577.79	10.54% 9.94% 9.28% 9.71% 10.27% 12.25% 9.28% 12.98%	× .					

Figure 8-10. Remove Rows dialog

6. Then click Close & Load on the left side of the ribbon to import the data into a spreadsheet, to get the results shown in Figure 8-11.

A1 · · · · · · · · · · · · · · · · · · ·	√ ∫x Year C	-						
A 8	C							~
1 Year Change in Index T	Total Annual Return Including Dividende	D	E	F G	н	1		
	our sumar recurrincidance or activation	Value of \$1.00 Invested on 1970-01-01	25 Year Annualized Return	Key			Workbook Queries	* X
2 1970 0.01%	1.01%	\$1.04		1				
3 1971 10.79%	14.31%	\$1.19	-	2			1 quey	D
4 1972 15.63% 1	18.98%	\$1.41		3			3 Total annual returns - Market	1.00
5 1973 -17.37% -	-14.66%	\$1.21	*	4			44 preview rows loaded.	
6 1974 -28.72% -	-26.47%	\$0.89	-	5				
7 1975 31.55% 3	37.20%	\$1.22	•	6				
8 1976 19.15% 2	23.84%	\$1.51		7				
9 1977 -11.50% -	-7.18%	\$1.40	-	8				
10 1978 1.06% 6	5.56%	\$1.49		9				
11 1979 12.31%	18.44%	\$1.77	•	10				
12 1980 25.77% 3	32.50%	\$2.34		11				
13 1981 -9.73% -	-4.92%	\$2.23	•	12				
14 1982 14.76% 2	21.55%	\$2.71		13				
15 1983 17.26% 2	22.56%	\$3.32	•	14				
16 1984 1.40% 6	5.27%	\$3.52	•	15				
17 1985 26.36% 3	31.73%	\$4.64	-	16				
18 1986 14.62% 1	18.67%	\$5.51		17				
19 1987 2.03% 5	5.25%	\$5.80	-	18				
20 1988 12.40% 1	16.61%	\$6.76	•	19				
21 1989 27.25% 3	31.69%	\$8.90	-	20				
22 1990 -6.56% -	-3.10%	\$8.63		21				
23 1991 26.31% 3	30.47%	\$11.26	•	22				
24 1992 4.46% 7	7.62%	\$12.11	-	23				
25 1993 7.06% 1	10.08%	\$13.33	•	24				
26 1994 -1.54%	1.32%	\$13.51	10.98%	25				
27 1995 34.11% 3	37.58%	\$18.59	12.22%	26				
28 1996 20.26% 2	22.96%	\$22.86	12.55%	27		٣		
Sheet1	Sheet2 (+)	1 4				Þ		

Figure 8-11. Data imported into a spreadsheet

Importing CSV Files from a Folder

Sometimes Hadoop data will be stored in files with comma-separated values. Rather than loading it into Hortonworks Sandbox or some other product to be manipulated as a Hadoop, the files with comma-separated values can be loaded directly into Excel and PowerPivot, using Power Query. The following example demonstrates how to do this.

1. First download the stock market zipped file from the following URL:

s3.amazonaws.com/hortonassets/mstut/nyse.zip

2. After the download is complete, double-click the file in your browsers download window to open it, as shown in Figure 8-12.

Opening nyse.zip	×
You have chosen to open:	
🚯 nyse.zip	
which is: Compressed (zipped) Folder (125 MB)	
from: http://s3.amazonaws.com	
What should Firefox do with this file?	
Open with Windows Explorer (default)	
○ Save File	
Do this <u>a</u> utomatically for files like this from now on.	
OK Cancel	

Figure 8-12. Download window

3. Two folders of .csv files will be downloaded, as shown in Figure 8-13.



Figure 8-13. Windows showing file folders

- 4. Drag the folder nyse_prices to the desktop, to make it easier to access.
- 5. Open Excel and click the Power Query tab. Select From File and From Folder, as shown in Figure 8-14.

FILE	■ ラ・ マ・ III IX III・ = FILE HOME INSERT PAGELAYOUT FORMULAS DATA REVI										Book1 - Excel									?	E - Neil
Q [Online Fi Search V	rom H Neb	From File *	From Databas	From e * Azure *	From Other Sources *	Recent Sources *	From Table	Merge App	pend	Show Pane	Workbo	ok Data S ps Sett	る の Source ings	date tions	My Data Catalog Queries	Sign In	Send Fee Help	dback *			
B6		XI	From E	kcel data from a oft Excel wo	rkhook		Excel Data	Combin	e	Workbook Quer	les	Set	ttings		Power bi		Help				
A A 1 2	-	G	From C Import comma	SV data from a -separated	value file.		F	G	н	1	J	К	L	м	N	0	P	Q	R	S	T
3 4 5 6	Г		Import file.	data from a ext	n XML																
7 8 9			From From From From From From From From	data from a older metadata a	text file.																
11	1		Fi	rom Folder																	
12 13 14			6	Microso Tell me r	ft Power Q	uery for Ex	cel														
15 16 17																					

Figure 8-14. Power Query screen to import from File Folder

6. Click Browse and select the nyse_prices folder from the desktop, as shown in Figure 8-15. Click OK.

		>	£
Folder			
Choose a folder.			
Folder Path			
C:\Users\Neil\Desktop\nyse_prices	Browse		
		OK Cancel	

Figure 8-15. Folder screen

The result is a list of .csv files, as shown in Figure 8-16.

	<mark></mark> - =					Query1 - Query Ed	itor			- 🗆 🗙
	le Home	Transform A	dd Column Vie	w						^ 📀
Clo	se & Refresh	Advanced Editor	Choose Remon Columns Column Manage	e S • Columns	Keep Remov Rows * Rows *	Remove Duplicates	Ali Split Group Solit Split By	Data Type: Any • Use First Row As Headers • 1 2 Replace Values Transform	Merge Queries Append Queries Combine Binaries Combine	
	Content +	Name	▼ Extension	▼ Date acces	sed 💌	Date modified	Date created	Attributes Mr Folder Path	¥	
1	Binary	NYSE_daily_prices_A.c	sv .csv	1/16/	2014 7:07:26 PM	3/6/2015 11:27:14 AN	1/16/2014 7:07:26	M Record C:\Users\Neil	Desktop\nyse_prices\	
2	Binary	NYSE daily prices B.c	sv .csv	1/15/	2014 7:07:26 PM	3/6/2015 11:27:18 AN	1/16/2014 7:07:26	M Record C:\Users\Neil	Desktop\nyse_prices\	^
3	Binary	NYSE_daily_prices_C.c	sv .csv	1/16/	2014 7:07:25 PM	3/6/2015 11:27:29 AN	1/16/2014 7:07:26	M Record C:\Users\Neil	Desktop\nyse_prices\	
4	Binary	NYSE_daily_prices_D.c	sv .csv	1/16/	2014 7:07:28 PM	3/6/2015 11:27:38 AN	1/16/2014 7:07:28	M Record C:\Users\Neil	Desktop\nyse_prices\	
5	Binary	NYSE_daily_prices_E.c	sv .csv	1/16/	2014 7:07:28 PM	3/6/2015 11:27:40 AN	1/16/2014 7:07:28	M Record C:\Users\Neil	Desktop\nyse_prices\	
6	Binary	NYSE_daily_prices_F.c	sv .csv	1/16/	2014 7:07:28 PM	3/6/2015 11:27:43 AN	1/16/2014 7:07:28	M Record C:\Users\Neil	Desktop\nyse_prices\	
7	Binary	NYSE_daily_prices_G.c	sv .csv	1/16/	2014 7:07:28 PM	3/6/2015 11:27:44 AM	1/16/2014 7:07:28	M Record C:\Users\Neil	Desktop\nyse_prices\	
8	Binary	NYSE_daily_prices_H.c	sv .csv	1/16/	2014 7:07:30 PM	3/6/2015 11:27:47 AN	1/16/2014 7:07:30	PM Record C:\Users\Neil	\Desktop\nyse_prices\	
9	Binary	NYSE_daily_prices_l.cs	v .csv	1/16/	2014 7:07:30 PM	3/6/2015 11:27:48 AN	1/16/2014 7:07:30	PM Record C:\Users\Neil	Desktop\nyse_prices\	
10	Binary	NYSE_daily_prices_J.cs	iv .csv	1/16/	2014 7:07:30 PM	3/6/2015 11:27:50 AN	1/16/2014 7:07:30	PM Record C:\Users\Neil	\Desktop\nyse_prices\	
11	Binary	NYSE_daily_prices_K.c	sv .csv	1/16/	2014 7:07:30 PM	3/6/2015 11:27:52 AN	1/16/2014 7:07:30	PM Record C:\Users\Neil	Desktop\nyse_prices\	
12	Binary	NYSE_daily_prices_L.cs	5V .C5V	1/16/	2014 7:07:30 PM	3/6/2015 11:27:54 AN	1/16/2014 7:07:30 8	PM Record C:\Users\Neil	Desktop\nyse_prices\	
13	Binary	NYSE_daily_prices_M.	csv .csv	1/16/	2014 7:07:30 PM	3/6/2015 11:27:56 AN	1/16/2014 7:07:30	PM Record C:\Users\Neil	Desktop\nyse_prices\	
14	Binary	NYSE_daily_prices_N.c	SV .CSV	1/16/	2014 7:07:30 PM	3/6/2015 11:27:59 AN	1/16/2014 7:07:30	PM Record C:\Users\Neil	Desktop/nyse_prices/	
15	Binary	NYSE_daily_prices_0.c	SV .CSV	1/16/	2014 7:07:30 PM	3/6/2015 11:28:06 AN	1/16/2014 7:07:30	PM Record C:\Users\Neil	\Desktop\nyse_prices\	
16	Binary	NYSE_daily_prices_P.c	sv .csv	1/16/	2014 7:07:32 PM	3/6/2015 11:28:10 AN	1/16/2014 7:07:32	M Record C:\Users\Neil	Desktop/nyse_prices/	
17	Binary	NYSE_daily_prices_Q.c	sv .csv	1/16/	2014 7:07:32 PM	3/6/2015 11:28:10 AN	1/16/2014 7:07:32	PM Record C:\Users\Neil	Desktop/nyse_prices	
18	Binary	NYSE_daily_prices_R.c	sv .csv	1/16/	2014 7:07:32 PM	3/6/2015 11:28:15 AN	1/16/2014 7:07:32	PM Record C:\Users\Neil	\Desktop\nyse_prices\	
19	Binary	NYSE_daily_prices_S.c.	sv .csv	1/16/	2014 7:07:32 PM	3/6/2015 11:28:17 AM	1/16/2014 7:07:32	PM Record C:\Users\Nell	Desktop/nyse_prices/	
20	Binary	NYSE_daily_prices_T.c	sv .csv	1/16/	2014 7:07:32 PM	3/6/2015 11:28:19 AN	1/16/2014 7:07:32	PM Record C:\Users\Neil	Desktop/nyse_prices/	
21	Binary	NYSE_daily_prices_U.c	SV .CSV	1/16/	2014 7:07:32 PM	3/6/2015 11:28:20 AN	1/16/2014 7:07:32	PM Record C:\Users\Neil	\Desktop\nyse_prices\	
22	Binary	NYSE_daily_prices_V.c	sv .csv	1/16/	2014 7:07:32 PM	3/6/2015 11:28:22 AN	1/16/2014 7:07:32	PM Record C:\Users\Neil	Desktop/nyse_prices/	
23	Binary	NYSE_daily_prices_W.	CSV .CSV	1/16/	2014 7:07:34 PM	3/6/2015 11:28:24 AN	1/16/2014 7:07:34	PM Record C:\Users\Neil	\Desktop\nyse_prices\	
24	Binary	NYSE_daily_prices_X.c	sv .csv	1/16/	2014 7:07:34 PM	3/6/2015 11:28:25 AN	1/16/2014 7:07:34	PM Record C:\Users\Neil	\Desktop\nyse_prices\	~
25	Binary	NYSE daily prices Y.c.	V22. V2	1/16/	2014 7:07:34 PM	3/6/2015 11:28:25 AM	1/16/2014 7:07:34 /	PM Record C:\Users\Neil	Desktop\nyse_prices\	No. of Concession, name
8 CC	DLUMNS, 26 RC	ws							PREVIEW DOWNLOA	DED AT 11:36 AM.

Figure 8-16. List of . csv files

7. To open a file, click Binary in the leftmost column. For this example, select the file in the fifth row. The file is opened in the QueryEditor, and results are shown in Figure 8-17. Remember that the Query Editor can only contain a subset of the total number of records.

XII	<u>₽</u> =			(Query1 - Query Ed	itor				- • ×
	Hom	e Transform	Add Column View							^ (YO
Clos	e & Refresh d • Preview	Properties Advanced Editor Query	Choose Remove Columns Columns • Manage Columns	e Keep Remove Rows * Rows * Redu	Remove Duplicates	2↓ ∡↓ Sort	Split Column - By Tra	Type: Decimal Number • Jse First Row As Headers • Replace Values nsform	Merge Queries Append Queries	
ш.	exchange	* stock_symbol *	date 💌 stock_price_open	▼ stock_price_high	stock_price_low		stock_price_close	stock_volume 💌 stock_pri	ice_adj_close 💌	
2	NYSE	EXI	2/5/2010	42.97	43.12	42.01	42.79	32300	42.79	
3	NYSE	EXI	2/4/2010	43.97	44.13	43.31	43.36	24300	43.36	
4	NYSE	EXI	2/3/2010	44.9	45.17	44.67	44.91	43800	44.91	
5	NYSE	EXI	2/2/2010	44.44	45.07	44.4	44.99	45300	44.99	
6	NYSE	EXI	2/1/2010	44	44.07	43.27	43.89	112500	43.89	
7	NYSE	EXI	1/29/2010	43.93	44.19	43.35	43.4	17600	43.4	
8	NYSE	EXI	1/28/2010	44.45	44.67	43.6	43.89	75700	43.89	
9	NYSE	EXI	1/27/2010	44.15	44.57	44.03	44.52	75300	44.52	
10	NYSE	EXI	1/26/2010	44.13	44.89	44.09	44.47	18300	44.47	
11	NYSE	EXI	1/25/2010	44.79	45.17	44.52	44.54	75200	44.54	
12	NYSE	EXI	1/22/2010	44.87	45.15	44.12	44.12	71400	44.12	
13	NYSE	EXI	1/21/2010	45.94	45.01	44.85	45	80800	45	
14	NYSE	EXI	1/20/2010	45.29	45.35	45.52	45.94	43700	45.94	
15	NYSE	EXI	1/19/2010	46.6	46.93	45.45	46.91	48400	45.91	
16	NYSE	EXI	1/15/2010	47	47	46.35	45.64	50700	45.64	
17	NYSE	EXI	1/14/2010	47.17	47.26	46.82	47.21	19600	47.21	
18	NYSE	EXI	1/13/2010	46.95	47.09	46.57	47.01	15700	47.01	
19	NYSE	EXI	1/12/2010	46.91	45.92	45.51	46.72	65100	45.72	
20	NYSE	EXI	1/11/2010	47.28	47.28	46.73	47.14	18600	47.14	Ť
46.	92									
9 CC	LUMNS, 50 R	ows							PREVIEW DOWNL	OADED AT 1:56 PM.

Figure 8-17. CSV file imported into Query Editor

- 8. Note that this file contains the opening, high, low, and closing stock price and volume for each stock on a given day. For this example, all rows and columns will be retained.
- **9.** Click Close & Load to load the records from the Query Editor into a spreadsheet. The results are shown in Figure 8-18.

	5. 6. [■ i× =• •		Воо	k1 - Excel			TABLE TOOLS		7 0	- 8 ×
FLE Table Nam Query1 Resize Propert	HOME IN E Sum Table Conv ies	SERT PAGE LAYO marize with PivotTable ove Duplicates vert to Range Tools	UT FORMULAS	DATA REVIEW Properties Open in Browser resh Open in Browser resh Open ink ternal Table Data	VIEW POWER Header Row Total Row Banded Row	QUERY POWE First Column Last Column S Banded Colu Table Style Opti	RPIVOT Team	QUERY DESIGN	Tat	de Styles	Nel *
A1	* 1 7	$\langle \sqrt{f_x} \rangle$	exchange								~
A	6	c	D	E	F	G	н	i	-		
1 excha	inge_stock_syn	nbol date st	tock_price_open_st	ock_price_high_stock_	price_low_stock	k_price_close_ste	ock_volume_stock_j	price_adj_close		Workbook Queries	* X
2 NYSE	EXI	2/8/2010	43.04	43.06	42.45	42.51	53000	42.51		1 milen	
3 NYSE	EXI	2/5/2010	42.97	43.12	42.01	42.79	32300	42.79		C Queed	D.
4 NYSE	EXI	2/4/2010	43.97	44.13	43.31	43.36	24300	43.36		2 206 source landed	
5 NYSE	EXI	2/3/2010	44.9	45.17	44.67	44.91	43800	44.91		2,296 rows loaded.	
6 NYSE	EXI	2/2/2010	44.44	45.07	44.4	44.99	45300	44.99			
7 NYSE	EXI	2/1/2010	44	44.07	43.27	43.89	112500	43.89			
8 NYSE	EXI	1/29/2010	43.93	44.19	43.35	43.4	17600	43.4			
9 NYSE	EXI	1/28/2010	44.45	44.67	43.6	43.89	75700	43.89			
10 NYSE	EXI	1/27/2010	44.15	44.57	44.03	44.52	75300	44.52			
11 NYSE	EXI	1/26/2010	44.13	44.89	44.09	44.47	18300	44.47			
12 NYSE	EXI	1/25/2010	44.79	45.17	44.52	44.54	75200	44.54			
13 NYSE	EXI	1/22/2010	44.87	45.15	44.12	44.12	71400	44.12			
14 NYSE	EXI	1/21/2010	45.94	46.01	44.85	45	80800	45			
15 NYSE	EXI	1/20/2010	46.29	46.36	45.52	45.94	43700	45.94			
16 NYSE	EXI	1/19/2010	46.6	46.93	46.45	46.91	48400	46.91			
17 NYSE	EXI	1/15/2010	47	47	46.35	46.64	50700	46.64			
18 NYSE	EXI	1/14/2010	47.17	47.26	46.82	47.21	19600	47.21			
19 NYSE	EXI	1/13/2010	46.95	47.09	46.57	47.01	15700	47.01			
20 NYSE	EXI	1/12/2010	46.91	46.92	46.51	46.72	65100	46.72			
21 NYSE	EXI	1/11/2010	47.28	47.28	46.73	47.14	18600	47.14			
22 NYSE	EXI	1/8/2010	46	46.66	46	46.56	39200	46.56			
23 NYSE	EXI	1/7/2010	45.38	46.02	45.38	46	19800	46			
as anner	ew.	- In Innan	** **	** **	47.70	** **	*****	** **			
	Snéét1	sneet2				1			Ŀ		
READY											+ 100%

Figure 8-18. Data loaded into spreadsheet

Group By

To group by a specific category, go into the Query Editor and follow these steps:

- 1. Reopen the Query Editor, by clicking Launch Editor on the Power Query tab. Then click Transform and Select Group By at the left end of the ribbon.
- 2. In the Group By dialog, select the stock_symbol column as the Group By criteria and leave the default as Count, as shown in Figure 8-19.

日 ち・		U - 1	-				Query1	- Query Editor					- 🗆 🗙
FILE HO			Home	Transform Add	d Column View								^ ()
Edit Propertie Query Edit	Group	D Usi As	First Row Headers • Table	#B Transpose G Reverse Rows Count Rows e	Data Type: Text •	t Column ivot Columns • Column	Expand Aggregate	Split Column Format	Merge Columns 統 Length 참 Parse ㆍ Column	XO Statistics Stand	ard Scientif	Trigonometry - T Da S Rounding - fic Information - olumn Date & T	ste - ne - uration - ime Column
A1		ш.	exchange	* stock_symbol	v date v stock	price_open 💌	stock_price_high	* stock_price	low v stock_price	close 🔻 stoc	k_volume	0	
	50	1	NYSE	EXI	2/8/2010	43.04		43.06	42.45	42.51		Query Settings	×
4	lerie	2	NYSE	EXI	3/5/3010	43.43		(3.13	43.41	63.70	^	OPERTIES	
1 Query1	Ğ	3	NYSE	EXI								× me	
2 The Data		4	NYSE	EXI	6							verv1	
3		5	NYSE	EXI	Group By							Description	
4		6	NYSE	EXI	Specify the columns	to group by.						Properties	
5		7	NYSE	EXI								PLIED STEPS	
6		8	NYSE	EXI	Group by		*					Source	*
7		9	NYSE	EXI	stock_symbol	*	-					Navigation	*
		10	NYSE	EXI								Imported CSV	
5		11	NYSE	EXI	New column name		Operation		Column		+	First Row as Header	
,		12	NYSE	EXI	Count		Count Rourd					Changed Type	
0		13	NYSE	EXI	Count		Count Nows					Grouped Rows	*
1		14	NYSE	EXI	-								
2		15	NYSE	EXI						OK	Cancel		
3		16	NYSE	EXI									
4		17	NYSE	EXI	1/14/2010	47.17		47.26	45.82	47.21			
5		18	NYSE	EXI	1/13/2010	45.95		\$7.09	46.57	47.01			
6		19	NYSE	EXI	1/12/2010	45.91		45.92	46.51	46.72			
7		20	NYSE	EXI	1/11/2010	47.28		47.28	46.73	47.14			
0		21	NYSE	EXI	1/8/2010	45		45.65	45	46.56			
0		22	NYSE	EXI	1/7/2010	45.38		45.02	45.38	46			
a		23	NYSE	EXI	1/6/2010	45.53		45.84	45.53	45.66			
20		24	NYSE	EXI	1/5/2010	45.48		45.67	45.37	45.56	~		
()		<	1								>		
CADY	9 COL	JMNS	50 ROWS									PREVIEW DOWNLOA	DED AT 11:03 AM.

Figure 8-19. Group By dialog

3. Click OK, and the results shown in Figure 8-20 will appear.

BB 5.		9- - 1				Query	1 - Query Editor			- 🗆 🗙
FILE HO		Home	Transform Add	Column View						^ ()
Edit Propertie Query Edit	Group	Use First Rom As Headers Tab	Count Rows	Data Type: Text • 1.2 Replace Values 2.2 Replace Errors 2.2 Replace Errors 3.2 Replace Errors 3.3 Replac	Fill • Pivot Column Unpivot Columns • Any Column	Expand	Split Column - Format Split Column - Text Column	Σ Statistics Standard Scient Number	0 ² Trigonometry - ⁰ Rounding - ntific Information - r Column	Date + Date + Time + Duration + Date & Time Column
A1	×.	m- stock_sy	mbol 💌 Count 💌						Quany Catting	
	5	1 EX	843	7					Query Setting	ys x
1. 000000	Queri	2 EWT	2420	0					A PROPERTIES	
The Det	U U	3 EOD	72:	8					Name	
2 The Data		4 ELX	5010	2					Query1	
3		5 EMN	405	7					All Properties	
4		6 EWM	348	8					A APPLIED STEPS	
5		7 ENH	174	8					Course	*
6		8 ERF	231	7					Navigation	*
7		9 ESV	4805	2					Imported CSV	
8		10 EPD	2890	5					First Row as He	ader
9		11 EAC	223						Changed Type	0.29
10		12 EWI	3480	5					× Grouped Rows	4
11		13 EL	3574							
12		14 605	335							
13		15 EUT	1/0							
14		17 FAT	516							
15		18 FWI	348							
16		19 EWN	348							
17		20 ETG	1518	8						
18		21 ETP	3190	2						
10		22 ETJ	635	2						
20		23 ELN	4510	2						
20		24 ELS	4266	5					~	
4	_	25 EMD	30	s						
READY	2 COLL	JMNS, 50 ROW:	5.						PREVIEW C	DOWNLOADED AT 11:12 AM.

Figure 8-20. Results after doing a Group By on stock_symbol

Importing JSON

This example shows how to use Power Query to extract JSON data from a web site. Figure 8-21 shows the raw JSON data on Nobel Prize winners.

http://api.nobel1/laureate.json × http://query.y22&formatajson × +								
🗲 🕲 api.nobelprize.org/v1/laureate.json	🖾 🥥 🗵 😋 🖓		¢	۵	+	ŵ	ø	Ξ
<pre>"laurestes":[("id":1", "firstname": "Wilhelm Conrad", "surname": "R\u00466 Germany), "bornCountyCode": "DE", "bornCity": "Lennep (now Remscheid),", di [("year':1")oliv", Costegory: "physica", "barker":1", "hocivation": "("in reco subsequently named after nin\", "affiliations":[("name": "Wunich Universi Wucherlander", "diedountryCode": "NL", "moder", "mainter", "prizes":[("year':1") Retherlander", "diedountryCode": "NL", "moder", "mainter", "prizes":[("year':1") Retherlander", "councy": "the Netherland")]])], ("id":"3", "firstname": "Eleter", "surname": "Cheman", "born": "1865-05-25", " Wetherlander", "bornCourtyCode": "NL", "Monitary "1865-05-25", " ("id":"3", "firstname": "Photogen: "display ("surser", "mainter", "loss and the surser", "bornCourty", "bornCity", "Sonormanie", "discourts", "Ison ampetiem upon radiation phonemenie", "difiliations"][["mame": "loss-05-25", " ("id":"4", "firstname": "Arter", "surname": "Size", "modivation": "loss and "surser", "surname": "loss and "surser", "surser, "surser", "surser, "surser", "surser, "su</pre>	<pre>ntgen", "born":"1845-03-27", "died":"1923-0 Search using Google [ry":" LadCountry": "Germany", "died"cuntryCode":"BC," "atButty" 'Munic opilion of the extraordinary services he has rendered by the ligy", "city": "Numlch", "country", "Germany"]]]), ['ligt":", "first 0100", "cy": "bit "bohland", "bornCountry", "bohland", "born "into a service of the service of the service of the service sm upon radiation phenomena\"", "affiliations"; [("name":"Leiden 0100", "cy": "bit "bohland", "bornCountry", "the "''''the Netherlands", "dotnCountry", "the "''''the Netherlands", "bornCountry", "the "i'''the Netherlands", "bornCountry", "the "bornCountry", "Cy", "Amaster opnition of the extraordinary service they rendered by their in University", "city", "Amasterdam, "country", "bornCity", "Paris", "site", "share", "bornCountry", "Faris", "country", "the reference, "bornCountry", "the "state of the extraordinary service they rendered by their in university", "city", "antiversed", "bornCountry", "bornCity", "Paris", "country", "the service of the extraord ","hub069cole Dolytechnique", "city", "Paris", "country", "dire "state", "share", "country", "chernis Bequere School of Industrial Physics and Chemistry", "chernis Bequere "state", [(1)), ('year', "1934-07-04", "bornCountry", "state", "notiv the radiation phenomena in icovered by Professor Henris Bequere "state", [(1)), ('year', "1931', "state", "bornCountry", "state", "notiv ison", "ison", "city, "rest, "country", "state", "state", "notiv ison", "ison", "city, "rest, "state", "bornCountry", "state", "notiv ison", "ison", "state", "state", "state", "bornCountry", "state", "notiv ison", "ison", "city, "rest, "state", "bornCountry", "state", "notiv ison", "ison", "city, "rest, "state", "bornCountry", "state", "notiv ison", "state", "pices", "ison", "state", "bornCountry", "based "ison", "pornCountry, "state, "state", "country", "based "ison", "bornCountry, "state", "country, "state", "state", "state", "state","," "ison", "bornCountry, "state", "country, "states", "states", "st</pre>	Prussia ",""eedi idaower," "Archean gnition dam", "ger searcher "," hornGu arbon", "ger searcher "," bornGu arbon", "," "," bornGu arbon", "," bornd", "," "," bornGu arbon", "," bornd",	<pre>(now pr":") y of ' ', "di of the mdrif' '' rvic' '' '' '' '' '' '' '' '' '' '' '' '' '</pre>	male" the : edCoo e ext :"mal es he "Pari" recoo tion; "recoo tion; s": sear: :"\"; "tion; sear: [{"ne near ions "male"	<pre>',"priz interprise interprise interprise interprise interprise '':("n" '':("n" '':("n" '':("n")))) interprise interp</pre>	"the nary "the nary "izes" ende ende edCo i of ame" i the gnit and d Gi v'rfc oyal	: rays e y s": edCouter the rove, or h	n y

Figure 8-21. Raw JSON data

To import the data using Power Query, follow these steps:

- 1. Click the Power Query tab.
- 2. Click From Web and enter the following URL, as shown in Figure 8-22, and click OK:

http://api.nobelprize.org/v1/laureate.json

FILE	÷ ا	- c> - IOME I	nsert	PAGE LAYOUT	FORM	MULAS (DATA I	REVIEW	VIEW PC	Boo DWER QUERY	ok1 - Excel	ERPIVOT			ųt s				"	10)
Q Online Search	From	m From b File + I	From Database * A	From From O Azure - Source Data	ther Reco	ent res = Table Excel D	Merge ata Co	Append mbine	Show Pane Workbook Qu	Workd Setti	book Data S ngs Sett Set	う (責) U ゆ 臣 C iource ings tings	Ipdate Options	My Data Catalog Queri Power	Sign ies In BI	Send Fee G Help About Help	dback *			
	A	B	c	Jx D	E	F	G	н	L.	J	к	L	м	N	0	P	Q	R	S	T
1						_										_		_		
3 4 5 6 7 8 9 10 11 12 13 14 15						From Enter a URL [http://a	Web Web page	URL. ze.org/v1/li	aureate.json					OK	0	ncel				
16																				
17			_						_											
18																				
19																				
21																				
22																				
23																				
2.4		the second second																		

Figure 8-22. Importing using the From Web option

3. The Query Editor will open. Click the Into Table icon at the upper left of the ribbon, as shown in Figure 8-23.

XI 🙂 - =		Convert	Query1 - Query Editor	- 🗆 🗙
File Home Transform	Add Column View	Record		^ ()
X ✓ fk = Ison laureates List	Document(Web.Contents("	http://api.nobelş	vize.org/v1/laureate.json")) Query Settings APROPERTIES Name Query1 All Properties APPLIED STEPS	×
			Source	¥
1 FIELD			PREVIEW DOWNL	OADED AT 11:13 AM.

Figure 8-23. The Convert screen for importing JSON

4. The result, which displays only the column heading, is shown in Figure 8-24.

🗓 🙂 = Query1 - Query Editor	- - ×
File Home Transform Add Column View Image: Column & Preventies Image: Column & Colum	 ↑ Image Queries ↑ Image Queries ↓ Combine Binaries Combine
× ✓ fit = Record.ToTable(Source) Image: Name Value w 1 taureates Ust	Query Settings × PROPERTIES Name Query1 All Properties APPLIED STEPS
	Source ☆ X Converted to Table
2 COLUMNS, 1 ROW	PREVIEW DOWNLOADED AT 11:13 AM.

Figure 8-24. Top-level JSON import

5. Notice the expand button to the right of Value. Click that expand button. The results are shown in Figure 8-25.

XI 🙂	- -			Query1 -	Query Editor	- 🗆 🗙
File	Home	Transform Add	d Column View			^ ()
Close & Load • Close	Refresh Preview •	Properties	Choose Remov Columns Column Manage Column	Keep Remove Remo	24 X↓ Split Column + By Sort Sort Data Type: Text + Text Solution Data Type: Text + Solution Data Type: Text + Text Solution	Gan Merge Queries Combine Binaries Combine
➤ Nam 1 Isou 2 Isou 3 Isou 4 Isou 5 Isou 6 Isou 9 Isou 10 Isou 11 Isou 12 Isou 13 Isou 14 Isou 15 Isou 16 Isou 17 Isou 18 Isou 19 Isou		Table.Expanse Value w Record	düstColumn(#*Co	verted to Table", "Value")		Query Settings × PROPERTIES Name Query1 All Properties APPLIED STEPS Source * Converted to Table X Expand Value
20 laur 21 laur	reates	Record				
22 laur 23 laur	reates reates	Record				~
2 COLUM	INS, 50+ ROV	rs				PREVIEW DOWINLOADED AT 11:13 AM.

Figure 8-25. JSON column list

XI I 🙂	=				Query1 - Qu	ery Edito	or				- 🗆 🗙
	Home	Transform	Add Column View	N							^ ()
Close & Load • Close	Refresh Preview	Cuery	d Editor Choose Remo Columns Colum Manage Colum	ve Keep Remove ns Rows Rows Reduce Ref	move Duplicates move Errors •	ŽI ŽI Sort	lit Group mm - By	Data Type: Any - Use First Row As Headers + 12 Replace Values Transform	E, 8-8	Merge Queries Append Queries Combine Binaries Combine	
X	√ fx	= Table	ExpandListColumn(#"Co	onverted to Table", "Value")				*	Query Cattings	
- Nor		Mahun 4.4								Query Settings	×
1 1200	nc ·	Percent								PROPERTIES	
2 Jaun	eates	Record	Search Columns to Exp	and					^	Name	
3 laun	eates	Record	Cleart All Column	6						Query1	
4 laun	eates	Record	✓ id							All Properties	
5 laun	eates	Record	✓ firstname							A APPLIED STEPS	
6 laun	eates	Record	🗹 surname							Source	
7 laun	eates	Record	Jorn S							Converted to Table	~
8 laun	eates	Record	died							X Expand Value	
9 laun	eates	Record	✓ bornCountry								
10 laun	eates	Record	bornCountryCode								
11 laun	eates	Record	diadCountry								
12 lauri	eates	Record	diedCountry								
13 laun	eates	Record	diedCity								
14 laun	eates	Record	gender								
15 laun	eates	Record	I prizes								
10 laun	eates	Record	Use original column	name as prefix							
17 laun	cates	Record	and and angline containing								
10 100	coucs	Percent		OK Cancel							
20 18-10	eates	Record									
21 Jaun	eater	Record									
22 Jaun	eates	Record									
23 laun	eates	Record							~		
2 COLUM	NS, 50+ ROW	<i>i</i> s								PREVIEW DOWNLOA	DED AT 11:13 AM.

6. Click the expand button again. The results are shown in Figure 8-26.

Figure 8-26. Select columns dialog box

7. Accept the defaults and click OK to get the results shown in Figure 8-27, which shows the actual records, with a single row for each Nobel laureate.

x II I 🙂	- -					Quer	y1 -	Query E	dito	r				- 🗆 🗙
File	Home	Transform	Add Column	n View	_									^ 6
Close & Load • Close	Refresh Preview *	Properties	ditor Choose Colum	se Remove ins Columns • age Columns	Keep R Rows * R	Remove Dupli	icates s •	ĝ↓ ∡↓ Sort	Spl	t Group mn • By	Data Type: Any • Use First Row As Headers • \$2 Replace Values Transform		Merge Queries Append Queries Combine Binaries Combine	
×	√ fx	= Table.E	xpandRecord	Column(#"Ex	pand Valu	e", "Value",						*	Query Settings	
m- Nar	me 💌	Value.id 💌	Value.firstnam	ie .		Value.surname	- V	alue.born	¥	Value.died	Value.bornCountry	[Query settings	×
1 laur	reates	1	Wilhelm Conra	d		Röntgen	18	845-03-27		1923-02-10	Prussia (now Germany)		PROPERTIES	
2 laur	reates	2	Hendrik Antoor	n		Lorentz	18	853-07-18		1928-02-04	the Netherlands	^	Name	
3 lau	reates	3	Pieter			Zeeman	11	865-05-25		1943-10-09	the Netherlands		Query1	
4 laur	reates	4	Antoine Henri			Becquerel	18	852-12-15		1908-08-25	France		All Properties	
5 laur	reates	5	Pierre			Curie	18	859-05-15		1905-04-19	France		A APPLIED STEPS	
6 laur	reates	6	Marie			Curie, née Sklodowska	18	867-11-07		1934-07-04	Russian Empire (now Poland)		Source	-84
7 laur	reates	8	Lord Rayleigh			(John William Strutt)	18	842-11-12		1919-06-30	United Kingdom		Converted to Table	
8 laur	reates	9	Philipp Eduard	Anton		von Lenard	18	862-06-07		1947-05-20	Hungary (now Slovakia)		Expand Value	
9 lau	reates	10	Joseph John			Thomson	18	856-12-18		1940-08-30	United Kingdom		X Expand Value1	8
10 laur	reates	11	Albert Abrahan	n		Michelson	18	852-12-19		1931-05-09	Prussia (now Poland)			
11 laur	reates	12	Gabriel			Lippmann	18	845-08-16		1921-07-13	Luxembourg			
12 laur	reates	13	Guglielmo			Marconi	18	874-04-25		1937-07-20	Italy			
13 laur	reates	14	Karl Ferdinand			Braun	18	850-06-06		1918-04-20	Hesse-Kassel (now Germany)			
14 laur	reates	15	Johannes Dider	rik		van der Waals	18	837-11-23		1923-03-08	the Netherlands			
15 laur	reates	16	Wilhelm			Wien	18	864-01-13		1928-08-30	Prussia (now Russia)			
16 laur	reates	17	Nils Gustaf			Dalén	18	869-11-30		1937-12-09	Sweden	~		
17 laur <	reates	18	Heike			Kamerlingh Onnes	11	853-09-21		1926-02-21	the Netherlands			
Gugliel	mo													
14 COLUI	MN5, 50+ RO	ws											PREVIEW DOWNLO	ADED AT 11:13 AM

Figure 8-27. JSON data in rows and columns.

8. Scroll all the way to the right to find another expand button, as shown in Figure 8-28.
CHAPTER 8 POWER QUERY

XIII U	• =		Query1	- Query Editor				
File	Home Transform Add	Column View						^ 6
Close & Load • Close	Refresh Preview • Query	Choose Remove Columns Columns •	Keep Remove Rows * Rows * Rows *	tes 21 Z1 Split Gru Column + E	Data Type: Any • Use First Row / 1,2 Replace Values Transform	As Headers 👻	Merge Queries Append Queries Combine Binaries Combine	
×	√ fx = Table.Expand	RecordColumn(#"Expa	nd Value", "Value",				Ouery Settings	~
ode 🔻 Vi	alue.bornCity	Value.diedCountry	Value.diedCountryCode	Value.diedCity	Value.gender 🔽 Val	ue.prizes	- Query settings	~
Le	ennep (now Remscheid)	Germany	DE	Munich	male List		PROPERTIES	
A	rnhem	the Netherlands	NL	nuli	male List		∧ Name	
Zo	onnemaire	the Netherlands	NL	Amsterdam	male List		Query1	
Pa	aris	France	FR	nul	male List		All Properties	
Pa	aris	France	FR	Paris	male List		▲ APPLIED STEPS	
W	Varsaw	France	FR	Sallanches	female List		Source	4
La	angford Grove, Maldon, Essex	United Kingdom	GB	nui	male List		Converted to Table	
Pr	ressburg (now Bratislava)	Germany	DE	Messelhausen	male List		Expand Value	
C	heetham Hill, near Manchester	United Kingdom	GB	Cambridge	male List		× Expand Value1	*
St	treino (now Strzeino)	USA	US	Pasadena, CA	male List			
н	lollerich		null null	nuli	male List			
Be	ologna	Italy	IT	Rome	male List			
Fu	ulda	USA	US	Brooklyn, NY	male List			
Le	elden	the Netherlands	NL	Amsterdam	male List			
G	iaffken (now Parusnoye)	Germany	DE	Munich	male List			
St	tenstorp	Sweden	SE	Stockholm	male List		~	
< G	ironingen	the Netherlands	NL	Leiden	male Ust	>		
Gugliein	mo							

Figure 8-28. Additonal columns

9. Clicking this expand button will generate multiple rows for persons who won the Nobel Prize more than once, as shown in Figure 8-29.

CHAPTER 8 POWER QUERY

01 🙂					Quer	y1 - Quei	y Editor			- 🗆 🗙
	Home	Transform	Add Column V	/iew						^ (
Close & Load • Close	Refresh Preview *	Cuery	tor Choose Re Columns Colu Manage Col	move umns • Keep R Rows • F	emove kows • Reduce Rows	cates 2	Split Column - By	Data Type: Text • Use First Row As Headers • \$2 Replace Values Transform	Combine	
×	√ fx	= Table.Exp	andListColumn(#	"Expand Value1",	"Value.prizes")				* Ouese Cattings	
III. Nac		Value id V	alue firstname		Value rurname	Value h	orn Value diad	Value hornCountry	Query Settings	×
1 1200	reates	1	Sibelm Conrad	L.2.1	Röntgen	1845-03	-27 1923-02-10	Prussia (now Germany)	A PROPERTIES	
2 1207	reates	2 14	andrik Antoon		Lorentz	1853.00	1929-02-10	the Netherlands	∧ Name	
2 1001	reater	3 01	atar		Zeeman	1865-05	-15 1943-10-09	the Netherlands	Query1	
A laur	reates	4 4	ntoine Henri		Becquerel	1852-13	15 1908-08-25	France	All Properties	
5 Jaur	reates	5 Pi	erre		Curie	1859-05	-15 1905-04-19	France		
6 laur	rester	6 14	laria		Curie née Skladowska	1867-11	.07 1934.07.04	Putrian Empire (now Poland	APPLIED STEPS	
7 1201	reater	6 M	arie		Curle, née Sklodowska	1867-11	.07 1934-07-04	Russian Empire (now Poland	Source	*
R taur	reates	8 10	ord Ravleigh		(John William Strutt)	1842-11	-12 1919-06-30	United Kingdom	Converted to Table	
0 laur	reater	9 24	siling Eduard Acton		yon Lenard	1862-04	.07 1947-05-20	Hungary (new Slovakia)	Expand Value	
	reates	10 10	sanh John		Thomson	1856-12	18 1940-08-30	United Kingdom	Expand Value1	8
1 Jaur	reater	11 A	hert Abraham		Michelson	1852-12	-19 1931-05-09	Prussia (now Poland)	A Expand Value.prizes	
2 Jaur	reates	12 6	abriel		Linnmann	1845-08	-16 1921-07-13	Luxembourg		
3 Jaur	reates	13 G	uglielmo		Marconi	1874-04	-25 1937-07-20	Italy		
4 laur	reates	14 K	arl Ferdinand		Braun	1850-06	-06 1918-04-20	Hesse-Kassel (now Germany)	1	
5 laur	reates	15 10	hannes Diderik		van der Waals	1837-11	-23 1923-03-08	the Netherlands	·	
6 laur	reates	16 W	filhelm		Wien	1864-01	-13 1928-08-30	Prussia (now Russia)		
7 laur	reates	17 N	ils Gustaf		Dalén	1869-11	-30 1937-12-09	Sweden		
8 laur	reates	18 He	eike		Kamerlingh Onnes	1853-09	-21 1926-02-21	the Netherlands		
9 laur	reates	19 M	lax		von Laue	1879-10	-09 1960-04-23	Germany		
0 laur	reates	20 Si	r William Henry		Bragg	1862-07	-02 1942-03-12	United Kingdom		
1 laur	reates	21 W	illiam Lawrence		Bragg	1890-03	-31 1971-07-01	Australia		
2 laur	reates	22 0	harles Glover		Barkla	1877-06	-07 1944-10-23	United Kingdom	~	
			Inc. Mart Present Looks Co		Dia a str	1070.00		California (a constraint)		

Figure 8-29. Expanded view with multiple rows for recipients of multiple Nobel Prizes

10. Scrolling to the right, clicking the expand button again, and accepting the defaults, as shown in Figure 8-30, will expand some of the fields that were not shown before, such as the reason for the prize.

XI 🙂	- -		Query1	- Query Editor				- • ×
File	Home Transform Add	Column View						^ (
Close & Load • Close	Refresh Preview - Query	Choose Remove Columns Columns - Manage Columns	Remove Errors • Remove Frors • Reduce Rows	tes 21 AL Split Column Sort	By Lata Type: A Group By L2 Replace	uny - t Row As Headers • Values	Combine	
×	√ fx = Table.Expand	ListColumn(#"Expand Value1	", "Value.prizes")				Ouen/Settings	~
ode V	alue.bornCity	Value.diedCountry	Value.diedCountryCode	Value.diedCity	Value.gender	Value.prizes	P Query Securitys	0
Le	ennep (now Remscheid)	Germany	DE		Land Land		PROPERTIES	
A	mhem	the Netherlands	NL	Search Column	ns to Expand		∧ Name	
Ze	onnemaire	the Netherlands	NL	Select All	Columns)		Query1	
Pi	aris	France	FR	✓ year			All Properties	
Pi	aris	France	FR	✓ category			APPLIED STEPS	
W	/arsaw	France	FR	✓ share			- Canada	ж
W	/arsaw	France	FR	I motivation	n		Converted to Table	54
u	angford Grove, Maldon, Essex	United Kingdom	G8	✓ affiliations			Expand Value	
Pr	ressburg (now Bratislava)	Germany	DE	Use original	column name as pref	iox	Expand Value1	4
C	heetham Hill, near Manchester	United Kingdom	GB				× Expand Value.prizes	
St	treino (now Strzeino)	USA	US		OK	Cancel		
н	ollerich	nul	nu.					
8	ologna	Italy	IT	Rome	male	Record		
Fu	ulda	USA	US	Brooklyn, NY	male	Record		
Le	eiden	the Netherlands	NL	Amsterdam	male	Record		
G	affken (now Parusnoye)	Germany	DE	Munich	male	Record		
St	tenstorp	Sweden	SE	Stockholm	male	Record		
G	roningen	the Netherlands	NL	Leiden	male	Record		
P	faffendorf	West Germany (now Germany)	DE	Berlin	male	Record		
W	/igton	United Kingdom	GB	London	male	Record		
A	delaide	United Kingdom	G8	Ipswich	male	Record		
W	/idnes	Scotland	G8	Edinburgh	male	Record	~	
			nr	C1	ar a * a			

Figure 8-30. Further expanding the rows

11. The results are shown in Figure 8-31.

X II (<mark>3</mark> -	Query1 - Query Editor – 🗖 🗙
File Home Transform Add Column	View ^ 🕑
Close & Close & Cuery Manag	Remove Remove Remove Split Group Data Type: Text * Split
X Image: second conditional second conditindex conditindex conditiona second conditional second conditiona	Alumn(#"Expand Value prizes", "Value prizes", "Yalue prizes", "Year", "category", "share", "motivation", "affiliations", Value prizes", "Year", "category", "share", "motivation", "affiliations", Value prizes", "Year", "category", "share", "motivation", "affiliations", Value prizes, "filiations", Value prizes, "filiations,", Value p
physics 2 physics 2 physics 4	"in recognition of the extraordinary service they rendered by their researc List List All Properties "in recognition of the extraordinary services he has rendered by his discov List All Properties "in recognition of the extraordinary services they have rendered by their): List APPLIED STEPS
physics 4 chemistry 1 physics 1	"In recognition of the extraordinary services they have rendered by their) just Source ♦ "In recognition of her services to the advancement of chemistry by the dis luit Converted to Table Converted to Table "for his investigations of the densities of the most important gases and foul luit Excand Value
physics 1 physics 1	*for his work on cathode rays* Ust Expand Value1 *in recognition of the prest merits of his theoretical and experimental inv. Ust Expand Value prices for bit posting a registion harmware to with a prestmential distribution of the prestmentation of the prestment of the prestmen
physics 1 physics 2	"for his method of reproducing colours photographically based on the phy List "in recognition of their contributions to the development of wireless teleg. List
physics 2 physics 1 physics 1	"In recognition of their controllocities to the development of writesis tetag, Lust "for his discoveries regarding the laws governing the radiation of heat" List
physics 1 physics 1 physics 1	"for his investigations of automatic regulators for use in conjunction with gas L List "for his investigations on the properties of matter at low temperatures wit List "for his discovery of the diffraction of X-krys by crystals" List
physics 2 physics 2 physics 1	"for their services in the analysis of crystal structure by means of X-rays" List "for their services in the analysis of crystal structure by means of X-rays" List "for their services in the analysis of crystal structure by means of X-rays" List
18 COLUMNS, 50+ ROWS	PREVIEW DOWNLOADED AT 11:13 AM.

Figure 8-31. Display showing reason (motivation) for Nobel Prize

12. Decisions can be made about how to group the data. For example, to group the data by category, click the heading of the category column to select the desired category, then right-click and select Group By to see the Group By... dialog, as shown in Figure 8-32.

Group By			
Specify the columns to group by			
Group by	+		
Value.prizes.category *] -		
New column name	Operation	Column	+
Count	Count Rows	•	~ =
Count	Count Rows	•	

Figure 8-32. Group By... dialog with single field

13. Clicking OK will show a count of the prizes by category, as in Figure 8-33.

] <mark>8</mark> × ≠	Query1 - C	Query Editor	-
File Home Transform Add	Column View		
Ilose & Refresh Node - Close Ouerv	Choose Remove Columns Columns Columns Remove Rows Rows Rows Rows Rows Rows Rows Rows	21 The second	Combine
X fr. = Table.Group(0 Value.prizes.category V Count physics 22 chemistry 18 b peace 22 t medicine 22 i literature 11 b economics 8 7 null	#"Expand Value.prizes.affiliations", ("Value.prizes.category") 7 7 7 7 8 8 8 8 8 8 9 6	, {{"Count", each Table.RowCount(_), type number}}}	 Query Settings PROPERTIES Name Query1 All Properties APPLIED STEPS Source Converted to Table Expand Value Expand Value Expand Value prizes Expand Value prizes affiliation X Grouped Rows

Figure 8-33. Count of Nobel Prizes by category

14. It is possible to group by two fields, such as category and year, by clicking the plus sign in the Group By... dialog box to select a second field, as shown in Figure 8-34. The Group By... dialog box can be accessed again by clicking the gear icon after Group Rows at the bottom of the Applied Steps list in the Query Setting pane on the right side of the screen.

Group By			
Specify the columns to group b	ay.		
Group by	+		
Value.prizes.category			
Value prizes year			
New column name	Operation	Column	+
Count	Count Rows	+	

Figure 8-34. Group By dialog with two fields

15. The results are shown in Figure 8-35.

		÷		Query1 -	Query	y Editor			- 🗆 🗙
File		Home Transform	Add Column View						^ (
Close Load Close) &	Refresh Preview • Query	tor Choose Remove Columns Columns	Keep Remove Remove Puplicate:	s ĝi Zi Sort	Data Type: Decimal Number • Split Group By L2 Replace Values Transform		Merge Queries Append Queries Combine Binaries Combine	
Queries *	m. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	 X X Jr Talespritzs.category Physics 	ble.Group(#"Expand Va Value.prizes.year Value.prizes.year Value.pri	alue.prizes.affiliations*, ("Value.prizes.co Count	ategory	y", "Value,prizes,year"), {{"Count", each	~	Query Setting: • PROPERTIES Name Query1 All Properties • APPLIED STEPS Source Converted to Table Expand Value Expand Value Expand Value Expand Value X Grouped Rows	5 ×
	21	physics	1921	1					
	22	physics	1922	1					
	23	physics	1923	1			~	-	

Figure 8-35. Count of Nobel Prizes by category and year

16. To add a third field to show the name of the prize winner, select + again in the Group By... dialog and add the surname of the winner, as shown in Figure 8-36.

CHAPTER 8 POWER QUERY

Group by			
Specify the columns to grou	ip by.		
Group by	+		
Value.prizes.category			
Value.prizes.year			
Value.surname	-		
New column name	Operation	Column	+
Count	Count Rows	*	

Figure 8-36. Group By... dialog with three fields

17. The results are shown in Figure 8-37.

	•	=			Query1 - C	Query	Editor		-	×
		Home Transform A	dd Column View							^ (
Close Load) &	Refresh Preview • Query	r Choose Remove Columns Columns Manage Columns	Keep Remove Rows * Rows * Reduce Row	nove Duplicates nove Errors •	Ź↓ Ž↓ Sort	Split Column * By 2 Replace Values Transform		Merge Queries Append Queries Combine Binaries Combine	
		fr = Tab	le Group(#"Expand Va	lue prizes affiliations" /"	alue prizes cat	ann	"Value prizes year" "Value sumame"		-	
10	L		e.oroup(= Expand va	ideprizestaniiations, (alueprizes.cau	egory	, value.prizes.year, value.sumane j,	-	Query Settings	×
Jerie		Value.prizes.category	Value.prizes.year	Value.surname	Count 👻				A PROPERTIES	
ð	1	physics	1901	Röntgen	1			^	Name	
	2	physics	1902	Lorentz	1			11	Query1	
	3	physics	1902	Zeeman	1				All Properties	
	4	physics	1903	Becquerel	1				An Properties	
	5	physics	1903	Curie	1				APPLIED STEPS	
	0	physics	1903	Curie, nee Sklodowska					Source	*
	1	cnemistry	1911	Curie, nee Skiodowska					Converted to Table	
	8	physics	1904	(John William Strutt)					Expand Value	
	9	physics	1905	von Lenard	1				Expand Value1	*
	10	physics	1905	Thomson	1				Expand Value.prizes	
	11	physics	1907	Michelson	1			-	Expand Value.prizes1	*
	12	physics	1908	Uppmann	1				Expand Value.prizes.affiliations	
	13	physics	1909	Marconi	1				➤ Grouped Rows	*
	14	physics	1909	Braun	1					
	15	physics	1910	van der Waals	1					
	10	physics	1911	Wien	1					
	17	physics	1912	Dalén	1					
	18	physics	1913	Kamerlingh Onnes	1					
	19	physics	1914	von Laue	1					
	20	physics	1915	Bragg	2					
	21	physics	1917	Barkla	1					
	22	physics	1918	Planck	1					
	23	physics	1919	Stark	1	_		~		

Figure 8-37. Display of Nobel Prizes by category and year, with surname of recipient

Summary

This chapter has demonstrated how to use Power Query to import tables from web sites, multiple CSV files from a folder, and JSON data. The examples have shown how to use Power Query to extract, transform, and load data. More examples of using Power Query will be provided in later chapters.

CHAPTER 9

Power Map

Power Map is a powerful visualization tool that allows the creation of 2D and 3D visualizations by plotting Excel-based geospatial data on a map, using Bing Map technology. This chapter demonstrates how to import data using Power Query and then how to plot it using Power Map.

Installing Power Map

Power Map is accessed through an icon on the Excel Insert tab. If that icon is not displayed, it may mean that you do not have the current version of Power Map installed. If that is the case, do a search on "download Power Map" and download the latest version from Microsoft. Then, to enable Power Map, follow these steps:

- 1. Click the File tab to see the Backstage View.
- 2. Click Options at the bottom of the left pane.
- 3. Select Add-Ins on the left pane.
- 4. At the bottom of the window, pull down the arrow after Manage, select Com Add-Ins, and click Go.
- **5.** Check Microsoft Office PowerPivot for Excel 2013, Microsoft Power Map for Excel and Power View, and click OK.

Plotting a Map

The steps to plotting a map include the following:

- 1. Open a data source, which may be a spreadsheet or a data model that includes geographic descriptors, such as city and state or geographic coordinates.
- 2. Select the Insert tab and click Map and Load Power Map.
- **3.** In the Fields list pane on the right side of the screen, select geographic fields to map to, such as city and state or geographic coordinates, and then select a data metric, such as population. Power Map calls Bing Map services to locate the geographic locations on a map, and then the metric value is plotted.
- **4.** By default, a column chart is displayed. This can be changed to a bubble chart or heat map by clicking the appropriate icon.

A Power Map video tour is an animation that tells a story about data by linking individual maps. A video can be created to tell a story based on the data. A video is a collection of tours, and a tour is a collection of scenes.

Key Power Map Ribbon Options

This section highlights some of the key options on the Power Map ribbon, which is shown in Figure 9-1.

0 5	¢									cacrir	mestatsma	y28.xlsx	- Microso	t Power	Map for Ex	cel				?
FILE	Н	OME																		
Play Tour	Add Scene	Create Video	Capture Screen	Themes	Map Labels	Riat Map	Find Location	Add Layer	Refresh Data	Shapes	2D Chart	A Text Bax	Legend	(L) Time Line	Date & Time	Tour Editor	Layer Pane			
	Т	our			М	ар			Layer			Insert		T	ime	Vi	ew			

Figure 9-1. Power Map ribbon

- Tour Section
 - Play Tour. Plays a tour already created
 - Add Scene. Adds a scene to a tour
 - Create Video. Creates a video from tours
- Map Section
 - Themes. Changes the map style
 - Map Labels. Adds country or city names
 - Flat Map. Switches between 2D and 3D views
- Layer Section
 - Add Layer. Adds a new layer
 - Refresh Data. Refreshes data from the source
 - Shapes. Selects the shape of the bar chart
- Insert Section
 - 2D Chart. Adds a 2D chart
 - Text Box. Adds a text box on the map
 - Legend. Adds a legend to the map
- Time Section
 - Time Line. Shows or hides the time line in a time animation
 - Date & Time. Shows or adds the date and time in a time animation
- View Section
 - Tour Editor. Displays or hides the Tour Editor pane (left side of screen)
 - Layer Pane. Displays or hides the Layer Pane (right side of screen)

Troubleshooting

Power Map is very particular about data types. If numeric data that you want to plot is not treated as numeric data, as evidenced by only showing a count aggregation, or if time data is not recognized, follow these steps:

- 1. Check for any extraneous string data in the column and delete the row containing string data.
- 2. Make sure that a numeric data column is formatted as a number and that a date column is formatted as date.
- 3. After changing the format, click Refresh Data on the Power Map ribbon.

EXERCISE 9-1. MAPPING CALIFORNIA CRIME STATISTICS EXAMPLE

This example shows how to use Power Query to import California crime statistics by city and how to use Power Map to plot the data.

1. Go to Power Query and do a Data Catalog search on "California Cities by Population and Crime Rate." Hover your mouse over "Cities—California locations by crime rate" and select Edit to load into the Query Editor. The screen shown in Figure 9-2 will appear.

File Home	Transform Add	Column View							~ 6
X 🗸 fx	= OData.Feed("h	ttps://publicdata.clo	uddatahub.net/We	b/Tables/ddeb5b217f7f4e86acb004	2b6fab7	a96/V1/Data")	*	Ouerv Settings	×
m. City	✓ County	Population 💌	Violent crimes	Violent crime rate per 1,000 persons	· Prope	rty crimes 💌 I	Property crime rate	query sectorige	
1 Adelanto	San Bernardino	32520	199	6.	12	924		A PROPERTIES	
2 Agoura Hills	Los Angeles	20657	12	0.	58	236	^	Name	
3 Alameda	Alameda	75457	160	2	12	1892		Cities - California locations a	by crime rate
4 Albany	Alameda	18960	35	1	85	537		All Properties	
5 Alhambra	Los Angeles	84459	149	1	76	1919		A APPLIED STEPS	
6 Aliso Viejo	Orange	48999	43	0.	88	415		Source	8
7 Alturas	Modoc	2802	36	12	85	62		Jource	
8 American Canyor	n Napa	19873	64	3.	22	616			
9 Anaheim	Orange	344526	1279	3.	71	10070			
10 Anderson	Shasta	10056	87	8.	65	606			
11 Antioch	Contra Costa	105009	1068	10	17	4757			
12 Apple Valley	San Bernardino	70823	221	3.	12	1874			
13 Arcadia	Los Angeles	57295	57	0.	99	1388			
14 Arcata	Humboldt	17408	54		1.1	754			
15 Arroyo Grande	San Luis Obispo	17568	60	3.	42	358			
16 Artesia	Los Angeles	16793	60	3.	57	262			
17 Arvin	Kern	19761	165	8	35	671			
18 Atascadero	San Luis Obispo	28825	192	6.	66	658			
19 Atherton	San Mateo	7050	6	0.	85	124			
20 Atwater	Merced	28891	180	6	23	1400			
21 Auburn	Placer	13787	44	3.	19	249			
22 Avalon	Los Angeles	3795	13	3.	43	64			
23 Avenal	Kings	15704	53	3.	37	132			
24 Azusa	Los Angeles	47111	220	4.	67	1204			
25 Bakersfield	Kern	355696	1929	5.	42	17754	~		
De Dalduia Dad	tas Annalas	76611	161	2	**	+ = 0 =			

Figure 9-2. Crime Statistics in Query Editor

 In Query Editor, click the Add Column tab. On the left end of the tab, click Add Custom Column. Create a new column for state. Set the contents ="CA", as shown in Figure 9-3.

New column name	
State	
Custom column formula:	Available columns:
="CA"	City
	County
	Population
	Violent crimes
	Violent crime rate per 1,000 p
	Property crimes
	Property crime rate per 1,000 V
	<< Insert
earn about Power Query formulas	

Figure 9-3. Adding a State column

3. Check that the data type of all the numeric columns is a decimal number by right-clicking the column heading, selecting Change Type, and selecting Decimal Number. Also, check that there are no non-numeric values in numeric columns. When the data is the way you want it, click Close & Load on the ribbon, to load to a spreadsheet. The spreadsheet should appear as shown in Figure 9-4.

Nome Indext Find Diverse Find Diverse	1 1 5 · 2 ·		0.07 0.000	Book1 - E	cel	TABLE	TOOLS	? 8
Al \bullet i \bullet f_x City A 8 C 0 E F G 2 delanto San Bernardino 2520 199 6.12 924 A lameda Alameda 75467 160 2.112 1892 Alameda Alameda 1560 35 1.85 537 Alameda Alameda 1560 35 1.85 537 Alameda 1360 35 1.85 1.85 537 Alambra Us Angeles 84469 149 1.76 1919 Alamoda Correct Cargon Napa 19873 64 3.22 616 American Caryon Napa 19873 64 3.22 616 American Caryon Napa 19873 64 3.22 616 American Caryon Shata 10056 87 88 0.017 4757 3 Apple Valley San Bernardino 70623 1221 3.12 1874 Arcadia Us Angeles 57295 57 0.99 1388 Arcata Us Modoł 17408 54 3.1 754 Arcadia Us Angeles 1778 60 3.57 262 Anterica Carlyon Napa 19873 64 3.1 754 Anardia Los Angeles 1778 60 3.57 262 Anterica Carlyon Napa 19873 64 3.1 754 Arcadia Us Angeles 1795 57 0.99 1388 Arcata Us Modoł 17408 54 3.1 754 Arcadia Us Angeles 1778 60 3.57 262 Anterica Carlyon Napa 19873 64 3.3 1754 Arcadia Us Angeles 1778 60 3.57 262 Anterica Carlyon Napa 13873 44 3.3 12 1874 Arcadia Us Angeles 1773 60 3.57 262 Anterica Carlyon Napa 13757 44 3.3 12 4.5 6.5 6.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7.5 7	ble Name: ble Name: ities_Californi Resize Table Properties	INSERT PAGE LAY ummarize with PivotTab emove Duplicates onvert to Range Tools	ble Slicer	DULAS DATA REVIEW Properties Properties Properties Open in Browser Dulink External Table Data	Header Row First Column Total Row Last Column Banded Rows Banded Colu Table Style Optic	Filter Button		Table Styles
A B C D E F G G Description Workbook Queries 1 City County Population Violent crime rate per 1,000 persons Property crimes ra	A1 * :	$\times \checkmark f_x$	'City					
I City County Population Wident crimerate per L000 persons Property crims load Pro	A	В	с	D	E	F	G	Warthard Overian
2 Adelanto San Bernardino 32520 199 6.12 924 A Agoura Hills los Angeles 20667 12 0.58 226 A Alameda Alameda 75467 160 2.12 1892 Alameda Los Angeles 4469 149 1.76 1919 7 Also Viejo Orange 46999 43 0.88 415 A Aturas Modo 2800 36 12.85 62 A Anarica Canyon Napa 19873 64 3.22 616 0 Anaheim Orange 344526 1279 3.71 10070 2 Also Contra Costa 105009 1068 10.17 4757 3 Apple Valley San Bernardino 70823 221 3.12 1874 Arcaja Los Angeles 10509 1068 10.17 4757 3 Apple Valley San Bernardino 70823 221 3.12 1874 Arcaja Los Angeles 10509 1068 10.17 757 3 Apple Valley San Bernardino 70823 221 3.12 1874 Arcaja Los Angeles 10509 1068 10.17 757 3 Apple Valley San Bernardino 70823 221 3.12 757 5 Arcata Humboldt 17408 54 3.1 754 6 Arroyo Grande San Luis Obipo 17568 60 3.42 358 0 Atherton San Mateo 7060 6 0.55 124 1 Atwater Merced 2881 180 6.23 1040 0 Atherton San Mateo 7060 6 0.85 124 1 Atwater Merced 2881 180 6.23 1040 2 Auburn Placer 13787 44 3.19 249 2 Auburn Placer 13787 74 4 3.19 249 2 Auburn Placer 13787 19 5.29 5.19 5.20 5.10 5.20 5.20 5.20 5.20 5.20 5.20 5.20 5.2	City 💌	County Po	opulation 💌 V	/iolent crimes 💌 Violent crime	e rate per 1,000 persons 🖬 Prop	erty crimes 💌 Prop	perty crime rate per 1,000 per	WORKDOOK Queries
3 Agour Hills Uos Angeles 20667 12 0.58 236 4 Alameda Alameda 73467 160 2.12 1892 5 Albany Alameda 13690 35 1.85 537 6 Alhambra Los Angeles 84469 149 1.76 1919 Alto Vielo Orange 4599 43 0.88 4415 6 Altrura's Modoc 2802 36 12.85 62 6 Anerican Canyon Napa 19873 64 3.22 616 0 Anaheim Orange 344526 1279 3.71 10070 1 Anderson Shata 10056 87 8.65 606 0 Anaheim Orange 344526 1279 3.71 10070 1 Anderson Shata 10050 1668 10.17 4757 3 Apple Valley San Bernardino 70623 221 3.12 1874 4 Arcadia Los Angeles 5755 57 0.99 1388 5 Arcata Humboldt 17408 54 3.1 754 4 Arcadia Los Angeles 10573 60 3.57 262 8 Arvin Kern 13761 165 8.35 671 9 Atacsadero San Luis Obispo 17568 60 3.42 336 9 Atacsadero San Luis Obispo 17568 192 6.66 658 9 Atacsadero San Luis Obispo 1758 100 6.23 124 1 Atwater Merced 28931 180 6.23 124 1 Atwater Merced 2891 180 760 760 760 760 760 760 760 760 760 76	Adelanto	San Bernardino	32520	199	6.12	924		1 query
4 Alameda 73467 160 2.12 1892 3 Albary Alameda 73467 160 2.12 1892 6 Albary Alameda 1990 35 1.85 537 6 Albarbara Los Angeles 84469 149 1.76 1919 7 Aliso Viejo Orange 44999 1.76 1919 8 Alturas Modo 2802 36 12.85 62 9 American Caryon Napa 19873 64 3.22 6.16 0 Anaheim Orange 344526 1279 3.71 10070 1 Anderson Shata 10056 87 8.65 606 2 Antioch Contra Corita 100509 1068 10.17 4757 2 Antioch Contageles 5725 57 0.99 1388 4 Arcala Los Angeles 5725 57 0.99 1388 4 Arcala Los Angeles 15753 60 3.42 358 6 Arroy o Grande San Luis Obispo 17568 60 3.57 262 6 A	Agoura Hills	Los Angeles	20667	12	0.58	236		* Cities - California locations by
Albany Alismeda 19500 35 1.88 537 Alhambra Los Angeles 84469 149 1.76 1919 Also Viejo Orange 48999 43 0.88 415 Alturas Modoc 2802 36 12.85 62 Anaheim Orange 34526 1279 3.71 10070 Anderson Shatata 10056 87 8.65 606 2 Antoich Contra Costa 105009 1068 10.17 4757 3 Appel valley San Bernardino 70823 221 3.12 1874 4 Arcadia Los Angeles 5725 57 0.99 1388 5 4 Arcadia Los Angeles 11708 54 3.1 754 5 6 Arroyo Grande San Luis Obipo 12761 65 6.35 671 8 Arvin Kern 13776 60 3.57 7262	Alameda	Alameda	75467	160	2.12	1892		100 preview rows loaded.
Almanora Los Angeles 84469 149 1.78 1919 Also Viejo Orange 44599 43 0.88 415 8 Atturas Modoc 2802 36 12.85 62 9 Anrerican Canyon Napa 3873 64 3.22 616 0 Anaheim Orange 344526 1279 3.71 10070 1 Anderson Shasta 10056 87 8.65 606 2 Antioch Contra Costa 105090 1068 10.17 4757 3 Apple Valley San Bernardino 70623 221 3.12 1874 4 Arcadia Los Angeles 5725 57 0.99 1888 3.1 754 6 Arroy of Grande San Luis Obispo 17568 60 3.42 3.58 3.5 671 6 Arroy of Grande San Luis Obispo 17585 8.35 671 4 4.42 3.59 4.4450 4.4450 4.4450 4.4450 4.4450 4.4450 4.4450	Albany	Alameda	18960	35	1.85	537		
Altro viejo Orange 48999 43 0.88 415 Altras Modoc 2802 36 12.85 62 American Caryon Napa 19873 64 3.22 616 Anaheim Orange 34426 1279 3.71 10070 Indreson Shasta 10506 87 8.65 606 Antroich Contra Costa 105009 1068 10.17 4757 Apple Valley San Bernardino 70823 221 3.12 1874 Arcadia Los Angeles 5725 57 0.99 1388 6 Arroyo Grande San Lus Obipo 1758 60 3.42 338 7 Artesia Los Angeles 16793 60 3.57 262 6 Arroyo Grande San Luis Obipo 1758 6.65 6.65 6.65 8 Atscadero San Mateo 7060 6 0.85 124 1 Atvater Merced 2891 180 6.23 1400 2 Auburn Piect 13787 4 3.19	Alhambra	Los Angeles	84469	149	1.76	1919		
Alturas Modoc 2802 30 12.83 62 Ø Amerian Carryon Napa 19873 64 3.22 616 Ø Anaheim Orange 344526 1279 3.71 10070 Ø Anaheim Shasta 10056 87 8.65 606 2 Anticoh Contra Costa 105009 1068 10.17 4757 3 Apple Valley San Bernardino 7623 221 3.12 1874 4 Arcadia Los Angeles 5725 57 0.99 1388 6 Arroyo Grande San Luis Obispo 17568 60 3.42 358 7 Artesia Luis Obispo 17568 60 3.57 262 8 Arvin Kern 15751 165 8.35 671 9 Atascadero San Luis Obispo 28825 192 6.66 658 9 Atascadero San Mateo 7060 6 0.85 124 1 Atwater Merced 28951 180 6	Aliso Viejo	Orange	48999	43	0.88	415		
a American Caryon Napa 19873 64 3.22 6.6 0 Anaheim Orange 344526 1279 3.71 10070 1 Anderson Shata 10056 87 8.65 606 2 Antioch Contra Costa 10090 1068 10.17 4757 3 Apple Valley San Bernardino 70823 221 3.12 1874 4 Arcatal Los Angeles 5725 57 0.99 1388 5 Arcata Humbolut 17408 54 3.1 754 6 Arroy Grande San Luis Obispo 17568 60 3.42 358 7 Artesia Los Angeles 10793 60 3.57 262 8 Arvin Kern 13761 165 8.35 671 9 Atascatero San Luis Obispo 28825 192 6.66 658 9 Atascatero San Luis Obispo 28825 192 6.66 658 9 Atascatero San Luis Obispo 1757 44 3.19 249 1 Atvater Mercel 28891 180 6.23 1400 9 Atascatero 1757 44 3.19 249 2 Auburn Placer 13757 44 3.19 249	Alturas	Modoc	2802	36	12.85	62		
Ornanem Ornange 344526 12/9 3.7.1 100/0 I Anderson Shasta 10506 87 8.65 666 2 Antioch Contra Costa 105009 1068 10.17 4757 3 Apple Valley San Bernardino 70823 221 3.12 1874 4 Arcadia Los Angeles 5725 57 0.99 1388 5 Arcata Humboldt 17408 54 3.1 754 6 Arroyo Grande San Lis Obipo 1758 60 3.42 358 7 Artesia Los Angeles 16793 60 3.57 262 8 Arvin Kern 13761 165 8.35 671 9 Atascadero San Lis Obipo 12825 192 6.66 658 0 Atherton San Mateo 7060 6 0.85 124 1 Atwater Merced 2891 180 6.23 1400 2 Auburn Placot Anonizer 3787 3.93<	American Canyon	Napa	19873	64	3.22	610		
Anderson Instata 1000 av 3.00 2 Anticch Contra Costa 10509 1068 10.17 4757 3 Apple Valley San Bernardino 70623 221 3.12 1874 4 Arcadia Los Angeles 57255 57 0.99 1888 5 Arcata Humboldt 17408 54 3.1 754 6 Arcoyo Grande San Luis Obispo 17568 60 3.42 358 7 Artesia Los Angeles 15793 60 3.52 262 8 Arvin Kern 19761 165 8.35 671 9 Atascadero San Luis Obispo 28825 192 6.66 658 0 Atherton San Mateo 7060 6 0.85 124 1 Atwater Merced 28891 180 6.23 1400 2 Auburn Places 13787 44 3.19 249 2 Auburn Isoet Angeles 3786 19 24.94	J Ananeim	Orange	344526	12/9	3./1	10070		
Andrech Contrat Costa 105009 1068 10.17 4757 2 Androch San Bernafino 70623 221 3.12 13.74 4 Arcata Los Angeles 57255 57 0.99 1388 5 Arcata Humboldt 17408 54 3.1 754 6 Arroyo Grande San Luis Obispo 1756 60 3.42 338 7 Artesia Los Angeles 16793 60 3.57 262 8 Anvin Kern 19761 165 8.35 671 9 Atascadero San Mateo 7060 6 0.85 124 14 Atwater Merced 2891 180 6.23 14000 2 Auburn Placer 13767 44 3.19 249 2 Auburn Isoet Sancia 3787 3.9 5.4	Anderson	Shasta	10056	8/	8.65	000		
a Apple Valley San Bernardino (70823 221 3.12 1874 4 Arcadia Los Angeles 5725 57 0.99 1388 5 Arcata Humboldt 17408 54 3.1 754 6 Arroyo Grande San Luis Obipo 17568 60 3.42 358 7 Artesia Los Angeles 16793 60 3.57 262 8 Arvin Kern 19761 165 8.35 671 9 Atscadero San Luis Obipo 2825 192 6.66 658 9 Atscadero San Luis Obipo 2825 192 6.66 658 9 Atscadero San Luis Obipo 2825 192 6.66 658 9 Atscadero San Luis Obipo 2825 192 6.66 658 1 Atwater Merced 2891 180 6.23 1400 1 Atwater Merced 2891 180 6.23 1400 2 Auburn Placer 13787 44 3.19 249 2 Auburn Placer 3785 19 2.6.4 ↓	2 Antioch	Contra Costa	105009	2001	10.17	4/5/		
Artatala L03 Angeles 57293 57 0.99 1388 Artatala Humbolut 17408 54 3.1 754 6 Arroy o Grande San Luis Obispo 17568 60 3.42 358 7 Artesia Los Angeles 16793 60 3.57 262 8 Arvin Kern 19761 165 8.35 671 9 Atascadero San Luis Obispo 28825 192 6.66 658 0 Athercon San Mateo 7060 6 0.85 124 1 Atwater Merced 28891 180 6.23 1400 2 Aubin Placer 13787 44 3.19 249 2 Aubin Ison Ananter 3785 19 3.9 6.4	s Apple Valley	San Bernardino	70823	221	3.12	18/4		
Antonia Antonia <t< td=""><td>Arcaula</td><td>Lus Angeles</td><td>17409</td><td>57</td><td>0.99</td><td>1388</td><td></td><td></td></t<>	Arcaula	Lus Angeles	17409	57	0.99	1388		
Andreis Los Angles Los Angles <thlos angles<="" th=""> Los Angles Los Angles<</thlos>	6 Arrowo Grande	San Luis Obisen	17569	54	3.1	259		
Interve Los nigetes Avria Diffe 3.0 3.0 202 8 Arvin 19751 165 8.35 671 9 Atascadero San Luis Obispo 28255 192 6.66 658 0 Atherton San Mateo 7060 6 0.85 124 1 Atwater Merced 28951 180 6.23 1400 2 Auburn Placer 13787 44 3.19 249 2 Auburn Los Analor 3705 32 2.92 6.64	7 Artoria	Los Angeles	16702	60	3.42	330		
Attacted or San Luis Obispo 22825 192 6.66 658 0 Atherton San Mateo 7060 6 0.85 124 1 Atwater Merced 2891 180 6.23 1400 2 Auburn Placer 13767 44 3.19 249 2 Auburn Floret 3787 33 2.93 6.64	Antin	Kern	19761	165	9.37	671		
Interform San Mate 7060 6 0.05 0 Atherton San Mateo 7060 6 0.85 124 1 Atwater Merced 28931 180 6.23 1400 2 Auburn Placer 13787 44 3.19 249 2 Auburn Merced 37861 19 2.49 14 2 Auburn Merced 37861 19 2.49 14	9 Atascadero	San Luis Obisno	28825	192	6.66	658		
I Atwater Merced 22891 180 6.23 1400 2 Auburn Placer 13787 44 3.19 249 2 Auburn Placer 2765 13 5.26 1400	Atherton	San Mateo	20023	6	0.85	124		
2 Auburn Placer 13787 44 3.19 249	1 Atwater	Merced	28891	180	6.23	1400		
2 Austan Lee Anneler 2705 12 2.43 6.4 W	2 Auburn	Placer	13787	44	3.19	249		
Sheet1 Sheet2 (+)	2 Auston	Los Angolos	2705	12	2.42	64		
	+ > Shee	et1 Sheet2	•		1		•	

Figure 9-4. Crime data loaded into spreadsheet

Again, check for non-numeric values in number fields that could keep Power Map from interpreting the contents correctly. Because not all the rows are displayed in the Query Editor, for larger datasets, this has to be checked after loading the data into Excel. As shown in Figure 9-5, the last row of data may contain ellipses. If this is the case, delete that last row by right-clicking the row and selecting Delete Table Row.

X	5-	¢• ■ ≥ ■•	-	Boo	k1 - Excel	L	ABLE TOOLS		?	
Tab Citi	le Name: esCaliforni Resize Table Properties	E INSERT PAGE Summarize with Pivo Remove Duplicates Convert to Range Tools	tTable Insert Slicer	MULAS DATA REVIEW	VIEW POWER QUERY P Header Row First Col vser Total Row Last Col Banded Rows Banded Table Style	umn I Filter But umn Columns Options	ton	byles		
A	ı *	$ X \lor f_x$	'City							
1	'City	· County	· Population ·	'Violent crimes 💌 'Violent	crime rate per 1,000 persons 💌	Property crimes 💌	Property crime rate per 1,000 persons 💌	'Key 🕶 'State 👻	J	ĸ
90	Corcoran	Kings	25119	89	3.54	328	13.06	89 CA		
91	Corning	Tehama	7751	51	6.58	281	36.25	90 CA		
92	Corona	Riverside	157342	210	1.33	4143	26.33	91 CA		
93	Coronado	San Diego	19345	20	1.03	504	26.05	92 CA		
94	Costa Mesa	Orange	112635	254	2.26	4079	36.21	93 CA		
95	Cotati	Sonoma	7398	54	7.3	85	11.49	94 CA		
96	Covina	Los Angeles	48588	151	3.11	1651	33.98	95 CA		
97	Crescent City	Del Norte	7723	51	6.6	215	27.84	96 CA		
98	Cudahy	Los Angeles	24201	151	6.24	347	14.34	97 CA		
99	Culver City	Los Angeles	39528	179	4.53	1760	44.53	98 CA		
100	Cupertino	Santa Clara	59769	58	0.97	797	13.33	99 CA		
101	Cypress	Orange	48976	56	1.14	1018	20.79	100 CA		
102								,		
103										
104										
105										
106										

Figure 9-5. Ellipsis in the last row that has to be deleted

4. Check the formatting of each column by highlighting the column, by clicking the letter at the top of the column. Right-click the selected column and select Format Cells from the Context menu. For the population, violent crimes, and property crimes columns, set the format to Number, with 0 places to the right of the point, as shown in Figure 9-6. For the violent crimes and property crimes per 1,000 persons, set the format to Number, with 2 places to the right of the point.

🛛 🖯 5· c	- •		crim	estatisticsbycalifor	viacities.xls	sx - Excel			TABLE	TOOLS					?	에 귀
FILE HOME	INSERT PAGE LA	AYOUT FOR	MULAS DAT	A REVIEW	VIEW	POWER QUERY	POWE	RPIVOT	QUERY	DESIGN						Ne
Table Name: Cities_Californi III Californi IIII Californi IIIII Californi IIII Californi IIIII Californi IIII Californi IIIII Californi IIII Californi IIII Californi IIII Californi IIIII	iummarize with PivotT Remove Duplicates Convert to Range Tools	Table Insert I Slicer	Export Refresh	 Properties Open in Browser Unlink Gable Data 	 ✓ Hea ☐ Tota ✓ Ban 	der Row 🗌 Fi al Row 🗍 L ded Rows 🗌 B Tabl	irst Column ast Column anded Colu le Style Opti	n √ Fi imns ons	ilter Button			Table Sty	les			
C1 * :	$\times \checkmark f_x$	'Population				For	rmat Cell	c		?	×					
A	B	с	D				l l		-				н	1	J	K
1 City	County 💌	Population 💌	Violent crime	Number Align	ment	Font Border	Fill	Protection			er 1,00	0 persons 💌 K	ey 🔜			
2 Adelanto	San Bernardino	32520		<u>⊆ategory</u> :								28.41	1			
3 Agoura Hills	Los Angeles	20667		General	~	Sample						11.42	2			
4 Alameda	Alameda	75467		Currency		Population						25.07	3			
5 Albany	Alameda	18960		Accounting	0	ecimal places:	0 0					28.32	4			
6 Alhambra	Los Angeles	84469		Time	E	Use 1000 Sepa	rator ()					22.72	5			
7 Aliso Viejo	Orange	48999		Percentage	1	enative numbers						8.47	6			
8 Alturas	Modoc	2802		Scientific	1	-1234					3	22.13	7			
9 American Canyor	Napa	19873		Text		1234						31.00	8			
10 Anaheim	Orange	344526		Custom		(1234)						29.23	9			
11 Anderson	Shasta	10056										60.26	10			
12 Antioch	Contra Costa	105009										45.30	11			
13 Apple Valley	San Bernardino	70823										26.46	12			
14 Arcadia	Los Angeles	57295										24.23	13			
15 Arcata	Humboldt	17408		L								43.31	14			
16 Arroyo Grande	San Luis Obispo	17568		Number is used t	or general	I display of numb	pers. Curren	ncy and Acco	ounting offer	specialized		20.38	15			
17 Artesia	Los Angeles	16793		formatting for m	onetary va	lue.						15.60	16			
18 Arvin	Kern	19761										33.96	17			
19 Atascadero	San Luis Obispo	28825										22.83	18			
20 Atherton	San Mateo	7060							or	Cancel		17.56	19			
21 Atwater	Merced	28891							U.	Cancer		48.46	20			
22 Auburn	Placer	13787	1.1	44			3.19		249			18.06	21			
23 Avalon	Los Angeles	3795		13			3.43		64			16.86	22			
34 A	Mage	1570.0		**			3.07		100				-			

Figure 9-6. Formatting column as number

5. Click the Insert tab and select Launch Power Map under Map. If you have previously created a map, you may be prompted to create a New Tour. Accept the default of City, County, and State as the geographic fields, as shown in Figure 9-7. Click Next.



Figure 9-7. Setting the geogrpahic fields

6. Click Map Labels on the ribbon to display city names. In the Layer pane, right-click Property Crimes per 1,000 persons and then click Add to Height, as shown in Figure 9-8. Note that the Bing Mapping service is showing a certainty of 94% for the geographic mapping in the upper right corner. If you want to go back and edit the geographic fields, click the icon above the 94%.



Figure 9-8. Setting field for height of plot

7. Click the legend box and resize it to be smaller, then drag it to the upper left part of the map. Click the large + on the lower right side of the map several times to zoom in. The results are shown in Figure 9-9. Note that it is possible to point the mouse at any graphic, to see the details about that city crime rate. It is also possible to double-click any location, to zoom in.



Figure 9-9. Map with legend and detailed crime statistics

Plotting Multiple Statistics

It is possible to plot multiple metrics—both property crime and violent crime statistics. To do this, let's continue with the example from the preceding section.

EXERCISE 9-2. MAPPING CALIFORNIA CRIME STATISTICS, PART 2

You can plot the property crime and violent crime statistics by following these steps:

1. In the Layer pane, right-click Property crime rate per 1,000 persons and select Add to Height, as shown in Figure 9-10.



Figure 9-10. Adding violent crimes metric

Note that the default view is stacked column. This can be changed to clustered column by clicking the icon to the right of stacked column, as shown in Figure 9-11. The property crime rate is shown in blue, and the violent crime rate in orange.



Figure 9-11. Display of both property crime and violent crime metric

- 3. Click Add Scene on the left end of the ribbon.
- 4. Click the Text Box icon on the ribbon and add the title "California Cities Crime Rates," as shown in Figure 9-12.

5 0		cacrimiestatsma	y28 xisx - Microsof	Power Map for Excel		5 1
Pile HOME	Themes Map Rat Find A Map Map Location Map	dd Refresh Skapers yer Data Layer	Text Lagend Box	Time Date 8 Line Time View		
Tour 1	× Layer 1	- 67	OREGON	Трано	Witness	5 12 Q
	Add Text Box Segoe UI • 26 • 1 TITLE California Cities Crime Rates	8 7 🔳 *		CONTENT PREVIEW: California Citie	es Crime Rates	fornia_locations_by
	DESCRIPTION				Create	on r crime rate per 1.00 r crimes
			San D	Diego Mexic	alli a ton y	

Figure 9-12. Adding a title to the map

5. Change the view to the bubble view, which is like a pie chart. The text box can be resized and dragged to the appropriate location on the map, as shown in Figure 9-13.



Figure 9-13. Bubble view of data

6. The size of the bubbles can be adjusted by clicking the gear icon at the top right of the Layer pane, selecting Layer Options, and adjusting the size slider, as shown in Figure 9-14.



Figure 9-14. Adusting bubble size

- 7. Click the Add Scene on the left end of the ribbon.
- 8. Another view is the heat map view shown in Figure 9-15. Note that San Francisco and Los Angeles, the two largest cities in California, have the most crime.



Figure 9-15. Heat map view of data

Adding a 2D Chart

A two dimensional chart for Property crime rate per 1,000 persons can be added by clicking 2D Chart on the ribbon. The results are shown in Figure 9-16.



Figure 9-16. 2D chart

EXERCISE 9-3. PLOTTING EUROPEAN UNEMPLOYMENT RATES EXAMPLE

This example will show how to load data about European unemployment rates into Power Query and how to plot it using Power Map. It will also show how to unpivot data in order to add a time animation. Follow these steps:

- 1. Click the Power Query tab and select Data Catalog Search.
- 2. Type "European Unemployment" into the search box. Hover the mouse over each of the results to get a preview of the data, as shown in Figure 9-17.

1 5	· c> · [• • • •	-			000050 00	Book1 - Excel					? E - 8 X
Name Descrip	This Ref	Source Coli Week - Nar	umn Certified mes	Recent Close Searches Search Options Close	EVIEW VIEW	POWERQU	SEARCI					nei · M
НЗ	• : >	< 🗸 f _x		Unemployn From Wikipedia, la Economy of the E	n <mark>ent</mark> - Econ st modified on Th propean Union - 1	omy of the ursday, May 29, . Vikipedia, the fre	e <mark>Europear</mark> 2014. re encyclopedia	Union			×	^
A In	B	c	D E	Unemployment	2005-03	2006-03	2007-03	2008-03	2009-03	2010-03	2011-	
1				Austria	5.1	5.1	4.5	4.1	4.5	4.9	4.	Online Search
2				Belgium	8.4	8.2	7.7	6.9	7.3	8.1	7.2	European Unemployment
4				Denmark	5.4	4.3	4.1	3	5.7	7.6	7.5	
5				Finland	8.5	7.9	7	6.3	7.4	9	8	PUBLIC [435]
6				Germany	9.8	8.7	8.6	7.4	7.6	7.3	6.	w Unemployment - Economy
7				Greece	9.9	9.6	8.6	7.8	7.8	10.2	13	From Wikipedia, last modified on Thurs
8				Ireland	4.5	4.2	4.6	5.6	10.6	13.2	14	Economy of the European Union -
9				Italy	7.8	7.7	6.1	6.6	6.9	8.8	8.1	wikipedia, the free encyclopedia
10				<							>	W LISE OF SOVEREIGHT States In Eu
12				Columns [10]								List of sovereign states in Europe by
13				Unemployment, 2	005-03, 2006-03,	2007-03, 2008-0	03, 2009-03, 2010	0-03, 2011-03, 2	012-03, Key			unemployment rate - Wikipedia, the
14				Data Sources								free
15				Thttp://en.wikip	edia.org/wiki/Ecc	nomy of the Eu	ropean Union					Net bilateral aid flows from
16							, .					From The World Bank, World Developm
17				10803 033								Net bilateral aid flows from DAC donors
18				LOAD - ED	T					LEGA	L TERMS	are the net disbursements of official
19												w Economies of member state
20												From Wikipedia, last modified on Thurs
21												1 2 3 4 5 Next
F	Sheet1	+					4) b	

Figure 9-17. Preview of European unemployment data

3. Select Edit, at the bottom of the preview screen, to load the data into the Query Editor, as shown in Figure 9-18.

01 🙂	× ≑				Unemplo	yment - E	conomy of t	he Europ	ean Union - O	Query Edit	or		- 0
	Home	Transform	Add Column	View									
Close & Load • Close	Refresh Preview *	Properties	or Choose Columns Manage	Remove Columns - Columns	Keep Ren Rows - Ro	Reduce Rov	iove Duplicates iove Errors * vs	Ž↓ Ž↓ Sort	Split Group	Data Type: D Use First 2 Replace \ Transform	ecimal Number * Row As Headers * /alues		Merge Queries Append Queries Combine Binaries Combine
×	✓ fx	= OData.Fee	ed("https://pu	blicdata.cl	ouddatahub	net/Web/Ta	ables/d0a3c79	c9e58416	58bb2ad2c5edc	0c07/V1/D	ata")	*	Quany Sattings
I Au	employment stria	▼ 2005-03 ▼ 5.1	2006-03 v	2007-03	2008-03 4.5	• 2009-03 4.1	v 2010-03 4.5	▼ 2011- 4.9	03 💌 2012-03 4.4	• Key 4.2	• 1	^	PROPERTIES Name
2 Bei	igium omark	8.4	8.2		1.1	2	7.3	8.1	7.7	7.2	2		Unemployment - Economy of the Euro
A Fin	land	85	7.9		7	53	7.4	0	8	7.6	4		All Properties
5 Fra	ince	9.7	9.1		8.6	7.6	8.8	10.1	9.5	10.1	5		APPLIED STEPS
6 Ge	rmany	9.8	87		8.6	7.4	7.6	7.3	6.2	5.6	6		A APPLIED STEPS
7 Gre	eece	9.9	9.6		8.6	7.8	7.8	10.2	13.8	22.1	7		Source
8 Irel	land	4.5	4.2		4.6	5.6	10.6	13.2	14.7	14.8	8		
9 Ital	W	7.8	7.7		6.1	6.6	6.9	8.8	8.3	10.3	9		
10 Lux	embourg	4.3	4.8		4.9	4.4	6.1	5.6	4.3	5	10		
11 Net	therlands	4.9	4.0		3.4	2.8	2.8	4.1	4.2	5	11		
12 Por	rtugal	7.4	7.6		8.2	7.6	8.5	10.5	12.6	15.1	12		
13 Spa	ain	9.9	8.7		8.1	9.5	17.4	19.1	20.7	24.1	13		
14 Sw	eden	6.3	7.2		6.6	5.8	8	8.7	7.7	7.3	14		
15 Uni	ited Kingdom	4.6	5.0		5.5	5.2	6.6	7.9	7.7	8.1	15		
16 Bul	Igaria				7.5	6.1	5.9	8.7	11.4	12.1	16		
17 Cyp	prus	5.1	5.2		4.1	3.7	4.9	6.7	7.3	10.7	17	~	
10 (10	arh Danublir	8.0	77		6.6	11		70	6.0	6.9	1.0		

Figure 9-18. Unemployment data in Query Editor

- 4. Check the shape of the data to see if there any extraneous columns. In this case, the columns include the country names and the unemployment rate by year and a key column that uses sequential numbering. Also, check the rows. Notice that there are rows for US and Japanese unemployment, as well as European.
- 5. Also, check the data type for each column to see if it is the correct type. To check the data type, select the column by clicking the heading and note the data type shown in the upper right section of the ribbon. To change the data type, select a column, click the arrow after Data Type on the ribbon, and select the correct data type. Notice that the data type for 2005 and 2006 is text and that for the other years is Decimal Number. Use this approach to change the type of all numeric columns to Decimal Number.
- 6. Check the column headings. Notice that the heading for the column with the country names is Unemployment, which is not descriptive. To change the column name, right-click the column heading and select Rename. Rename it to country, as shown in Figure 9-19. Notice that the column rename is listed under Applied Steps in the Query Settings pane on the right side of the screen.

XIII	e = 1				Unemplo	yment -	Economy of	the Europe	an Union - (Query Edit	tor	- 🗆 💌
	Home (V01) Se & Refresh	Transform <u>V02</u> Properties Advanced	Add Colur Y03 Editor	nn View V04	Keep Ren		emove Duplicates emove Errors •		lit Group	Data Type: Te Use First	ext • Row As Headers • Values	Merge Queries Append Queries Combine Binaries
Cl	ose	Query	Ma	nage Columns	KOWS · KO	Reduce R	ows	Sort	ann · by	Transform		Combine
	V 6	= Table	RenameColu	mos/Source II*	Unemployme	ent" "cour	ato/"III					
	country	- 10016.	2006.03	2007.03	2008.03	2000	nay m	2011.03	2012.01	Var		Query Settings ×
IIII v	country	• 2005-05	2006-05	2007-05	· 2008-05	· 2009-0	2010-05	+ 2011-03	2012-03	Ney		
8	Ireland	4.5	4.2		4.6	5.6	10.6	13.2	14.7	14.8	8	∧ Name
9	Italy	7.8	7.7		6.1	0.0	6.9	8.8	8.3	10.3	9	Unemployment - Economy of the Europea
10	Luxembourg	4.3	4.8		4.9	4.4	6.1	5.6	4.3	5	10	All Properties
11	Netherlands	4.9	4.0		3.4	2.8	2.8	4.1	4.2	5	11	A APPLIED STEPS
12	Portugal	7.4	7.0		8.2	7.0	8.5	10.5	12.0	15.1	12	- AFFLED STEPS
13	Spain	9.9	8.7		8.1	9.5	17.4	19.1	20.7	24.1	13	Source 🕆
14	Sweden United Vinadam	0.5	1.2		0.0	5.8	0	0.7	7.7	7.3	14	A Kenamed Columns
10	Bulgasia	4.0	5.0		7.5	2.2	0.0	0.7	11.4	12.1	15	
10	Ougaria		6.2		1.5	0.4	3.9	6.7	7.2	10.7	17	
10	Crach Papublic	8.0	7.7		5.6	3.1	4.5	7.0	60	6.0	10	
10	Ertopia	0.0	6.3		3.0	4.4	3.5	1.9	14.2	10.6	10	
20	Estonia	6.0	7.4		7.2	76	0.2	13.5	11.0	11.1	20	
20	I atvia	0.0	7.6		5.0	61	16.1	22.8	16.2	15.4	20	
21	Latvia	0.2	6.4		0.4	4.2	10.1	15.0	16.2	12.6	22	
22	Malta	7.2	0.4		4.0	5.0	67	6.0	6.3	6.1	22	
23	Doland	18.0	16.9		0.0	2.0	7.7	0.5	0.0	0.0	2.3	
24	Portanu	10.0	10.0		6.6	6.2	5.0	7.6	7	7.2	24	
26	Slovakia	16.7	15.7		11 3	0.0	10.5	14.1	14	13.7	25	
20	Clouenia	10.7 6.4	6.2		5.2	45	£	6.2	81	R 1	20	
20	European Linion	8.9	8.4		73	67	83	0.6	0.5	10.3	28	
20	Linited States	5.1	4.7		1.0	5.1	9.5	0.7	0.0	8.2	20	
30	lanan	45	41		4	20	4.4	48	4.6	45	30	~
30		4.5	4.4		-	4.4	4.4	4.0	10	4.5	30	

Figure 9-19. Query Editor after renaming country column

- 7. When the data is shaped the way you want it, select Close & Load at the leftmost side of the Query Editor ribbon to load the data into a spreadsheet. Note that data can also be loaded into a Data Model, but for this example, a spreadsheet will be used.
- 8. This would be a good time to save the spreadsheet and name it "European Unemployment."
- 9. To start the mapping, select the Insert tab and Map and then Launch Power Map.
- 10. The screen shown in Figure 9-20 will appear. Note that in the Layer pane, on the right side of the screen, country is selected as the geographic field. Click Next.



Figure 9-20. Country selected as geographic field

11. Note that dots appear in each country that has data. Click Map Labels on the ribbon to display country names and select the numeric field that will be plotted. In this case, the 2012 unemployment rate will be checked, so that it can be plotted, as shown in Figure 9-21. Note that when the field is checked, it is added to the Height box. The default format is stacked column, and the default aggregation of the numeric values is sum. Pointing the mouse at a column on the map will display the unemployment rate for that country in 2012.



Figure 9-21. 2012 unemployment shown as a bar chart

12. Change the aggregation to none by clicking the arrow at the right side of the Height box and selecting No Aggregation. Resize the legend box by clicking it and using the handles, and drag it to reposition it on the left side of the map, as shown in Figure 9-22.



Figure 9-22. Map after moving legend box and changing aggregation to none

13. Experiment with the different visualization formats by clicking the icons above the word *Height* in the Layer pane, to view the clustered column, bubble, heat map, and region. Note that because only one set of data is being shown, the clustered column and stacked column format are the same. A bubble view is shown in Figure 9-23.



Figure 9-23. Bubble view

- 14. To zoom in on a map, click the large + sign at the bottom right of the map or double-click an area of the map to zoom in.
- 15. To change the opacity, size, or thickness of the bubble, click the gear icon on the right, at the top of the Layer pane. Figure 9-24 shows a heat map view with the color scale set to 180%, the radius of the heat map to 191%, and opacity at 99%. The legend box for Layer 1 shows the color key based on the unemployment rate. Note that you can also decide whether to show zeroes, negatives, and nulls, by checking or unchecking the box.



Figure 9-24. Heat map view

Showing Two or More Values

Two or more values can be displayed, such as showing unemployment rates in two different years. Using the example from the preceding section, Figure 9-25 shows the employment rates for 2011 (orange) and 2012 (blue) as a bubble. Pointing the mouse at the color on the bubble will show the unemployment rate for that year.



Figure 9-25. Bubble view

Figure 9-26 shows the unemployment rates for 2010 (orange) and 2011 (blue) as stacked bar columns. Pointing the mouse at the color on the stacked bar will show the unemployment rate for that year.



Figure 9-26. Stacked bar view

The same data shown with clustered columns, where the columns are side-by-side, is shown in Figure 9-27.



Figure 9-27. Clustered column view

Creating a 2D Chart

A two-dimensional chart showing the progression over time can be plotted by checking a series of years and then clicking 2D Chart on the ribbon. The result is shown in Figure 9-28. Hovering the mouse over the 2D chart will bring up a box with a down arrow after the title, which allows changing the chart type.



Figure 9-28. 2D chart

EXERCISE 9-4. BUILDING A TIME ANIMATION EXERCISE

It is possible to animate a presentation to show the map values at different times. To do this, it is necessary to unpivot the data to indicate that a value is shown for each country for each year. To start from scratch, go back to steps 1–3 at the beginning of "Exercise 9-3. Plotting European Unemployment Rates Example" to load the data into the Query Editor, or click Recent Sources on the PowerPivot tab and select the data there.

 In the Query Editor, click the first year column heading to select the column. Then Ctrl-click the other year column headings to select all year columns, or after selecting the first year column, Shift-click the last year column to select all year columns. Right-click one of the selected column headings and select Unpivot Columns, as shown in Figure 9-29.

FILE Hor Hor Transform Add Column View Image: Column S Perform Advanced Eatre Image: Column S	1 8 5.	XII (8)- = 1			Unemploym	ent - Economy of t	he European Unio	on - Query Editor	
Image: Description Image:	FILE HOM	File Home	Transform Add	Column View					~
Image: Second	ame Description	Close & Refresh Close	Properties Advanced Editor	Choose Remove Columns Columns - Manage Columns	Keep Remove Rows * Rows *	Remove Duplicates	\$↓ ≵↓ Split Column - G	Data Type: Any • Toup By Transform	Combine
A III. Unemployment 2005-03 2006-03 2006-03 2006-03 2006-03 2011-03	1 *	X V fr	= OData.Feed(https://publicdata.c	louddatahub.net	/Web/Tables/d0a3c79	c9e5841658bb2ad2	c5edc0c07/V1/Data")	* Query Settings
3 nine 3.1 3.2 0.8 7.0 0.8 7.0 0.8 7.0 0.6 2.0 0.6 0.8 0.0 0.	A	Unemployment Austria Belgium Denmark Finland	2005-03 200 5.1 5. 8.4 8. 5.4 4. 8.5 7.	006-03 💌 2007-03 1 2 3 9	▼ 2008-03 ▼ 4.5 4.1 7.7 6.5 4.1 3 7 6.3	2009-03 2010-03 4.5 7.3 5.7 7.4	2011-03 49 44 81 7.7 7.6 7.5 9 8	Remove Columns Remove Other Columns Remove Duplicates Remove Errors	A PROPERTIES Name Unemployment - Economy of the Europea All Properties
7 Greece 9.9 9.6 8.6 7.8 7.8 10.2 13.8 8 Iretand 4.5 4.2 4.6 5.6 10.6 13.2 14.3 10 Lucembourg 4.3 4.8 4.9 4.4 6.1 5.6 4.3 6.7 Group By- 11 Netherlands 4.9 4.0 3.4 2.8 4.1 4.4 4.1 5.6 6.9 8.8 Group By- 11 Netherlands 4.9 4.0 3.4 2.8 4.1 4.4 4.1 4.4 4.1 4.4 4.1 4.4 4.1 4.4 4.1 4.4 4.1 4.4 4.1 4.4 4.1 4.4		6 Germany	9.7 9.	7	8.6 7.4	7.6	7.3 6.2	12 Replace Values	APPLIED STEPS
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$		7 Greece 8 Ireland 9 Italy	9.9 9. 4.5 4. 7.8 7.	6 2 7	8.6 7.8 4.6 5.6 6.1 6.6	7.8 10.6 6.9	10.2 13.8 13.2 14.7 8.8 8.3	Change Type Transform	source se
12 Portugal 7,4 7,6 8,2 7,6 8,2 105 126 Move Move 13 Suin 9,9 8,7 8,1 9,5 17,4 19,1 207 2,3 14 13 Suin 6,3 7,2 6,6 5,8 8 8,7 7,7 2,3 14 15 United Kingdom 4,6 5,0 5,5 5,2 6,6 7,9 7,7 8,3 15 16 Bugaria 7,7 5,6 4,4 5,9 6,7 7,7 14 12,1 15 18 Catch Republic 5,1 5,2 4,1 3,7 4,9 6,7 7,3 10,7 17 18 Catch Republic 5,0 7,7 5,6 4,4 5,5 7,9 6,9 6,8 18 19 Istronia 8,8 5,3 4,9 4 1,11 12,5 14,4 10,8 12,1		10 Luxembourg 11 Netherlands	4.3 4. 4.9 4.	s 0	4.9 4.4 3.4 2.8	6.1 2.8	5.6 4.3 4.1 4.2	금 Group By 팸 Unpivot Columns	
15 pum 13 8.7 8.4 9.5 1.4 9.12 1.07 1.44 9.5 14 Sweden 6.3 7.2 6.6 5.8 8 8.7 7.7 7.3 1.4 15 United Kingdom 6.6 5.0 5.5 5.2 6.6 7.9 7.7 8.1 1.5 16 Bugin'a 7.5 6.1 5.9 8.7 7.7 3.1 1.6 17 Openia 5.2 6.6 7.9 7.7 8.1 1.5 18 Czech Republic 8.0 7.7 5.6 4.4 5.5 7.9 6.9 6.8 1.8 19 Etonia 8.8 5.3 4.9 4 1.11 12.5 14.3 10.6 1.9 20 Hungary 6.8 7.4 7.3 7.6 9.2 1.1 1.15 1.14 1.12 2.0 21 Laria 5.1 7.6 <td></td> <td>12 Portugal</td> <td>7.4 7.</td> <td>6</td> <td>8.2 7.6</td> <td>8.5</td> <td>10.5 12.6</td> <td>Move</td> <td>•</td>		12 Portugal	7.4 7.	6	8.2 7.6	8.5	10.5 12.6	Move	•
15 United Singdom 46 50 55 5.2 6.6 7.7 8.1 1.5 16 Bulgaria 7.5 6.1 5.9 8.7 114 12.1 16 17 Oprius 5.1 5.2 4.1 3.7 4.9 6.7 7.3 10.7 17 18 Czech Republic 8.0 7.7 5.6 4.4 5.5 7.9 6.9 6.8 18 19 Istonia 8.8 5.3 4.9 4 11.1 15.5 14.3 10.6 19 20 Hungary 6.8 7.4 7.3 7.6 9.2 11.1 11.8 11.3 20 21 Latvia 9.1 7.6 6.4 6.1 16.1 22.3 16.2 15.4 21		14 Sweden	6.3 7.	2	6.6 5.8	8	8.7 7.7	7.3 14	
10 original 7.5 0.4 5.9 0.4 1.4 1.24 1.24 17 Vopus 5.1 5.2 4.1 3.7 4.9 6.7 7.3 107 18 Cech Republic 8.0 7.7 5.6 4.4 5.5 7.9 6.9 6.8 1.8 19 Estonia 8.8 5.3 4.9 4 1.1 1.55 1.43 10.8 1.9 21 Iatria 6.8 7.4 7.3 7.6 9.2 1.1 1.14 1.12 20 21 Iatria 8.1 7.6 6.4 6.1 1.61 223 1.62 1.54 21		15 United Kingdom	4.6 5.	0	5.5 5.2	6.6	7.9 7.7	8.1 15	
10 Cecch Republic 8.0 7.7 5.6 4.4 5.5 7.9 6.9 6.8 1.8 19 Estonia 8.8 5.3 4.9 4 1.1.1 15.5 14.3 10.6 1.9 20 Hungary 6.8 7.4 7.3 7.6 9.2 1.1 1.15 1.4.3 11.1 2.0 20 Hurdia 9.1 7.6 6.4 6.1 1.6.1 2.2.3 1.6.2 1.5.4 2.1		17 Cyprus	5.1 5	2	41 37	4.9	67 73	10.7 17	
19 Estonia 8.8 5.3 4.9 4 11.1 15.5 14.3 10.6 19 20 Hungary 6.8 7.4 7.3 7.6 9.2 11 11.8 11.1 20 21 Larvia 9.1 7.6 6.4 6.1 16.1 22.3 16.2 15.4 21		18 Czech Republic	8.0 7.	7	5.6 4.4	5.5	7.9 6.9	6.8 18	
20 Hungary 6.8 7.4 7.3 7.6 9.2 11 11.8 11.1 20 21 Latvia 9.1 7.6 6.4 6.1 16.1 22.3 16.2 15.4 22.1		19 Estonia	8.8 5.	3	4.9 4	11.1	15.5 14.3	10.6 19	
21 Latvia 9.1 7.6 6.4 6.1 16.1 22.3 16.2 15.4 21		20 Hungary	6.8 7.	4	7.3 7.6	9.2	11 11.8	11.1 20	
		21 Latvia	9.1 7.	6	6.4 6.1	16.1	22.3 16.2	15.4 21	
22 Uthuania 9.2 6.4 4.6 4.3 15.1 15.8 16.3 13.6 22		22 Lithuania	9.2 6.	4	4.6 4.3	15.1	15.8 16.3	13.6 22	

Figure 9-29. Context menu showing the Unpivot Columns option

2. The results are shown in Figure 9-30. One row has been created for each country and year.



Figure 9-30. Unpivoted data

3. Note that the column containing the year is labeled Attribute, which is not descriptive. Right-click the column heading and select Rename to change the name to year. Make sure that only the Attribute column is selected. Then change the data type to date, by selecting the column and clicking the down arrow after Data Type on the top-right side of the ribbon and selecting date. Change the name of the unemployment column to country, change the name of the value column to unemployment, and change the type of the column to decimal number. After changing the type of the unemployment column to decimal number, all data will be right-justified. The data should appear as shown in Figure 9-31. Note that the Applied Steps are shown in the Query Settings pane on the right side of the screen.

XII (*	r ≠ I			Unemployment - Economy of the	ne European Union - Query Editor	- - ×
F Close & Load • Close	Home Y01 Refresh Preview •	Transform Ad Y02 Properties Advanced Editor	Choose Remove Columns Columns *	Keep Remove Rows * Rows * Reduce Rows	21 3 T spit Data Type: Decimal Number * mill Use First Now As Headers * Column * by 12 Replace Values Soft Group 12 Replace Values	^
×	√ fx	= Table.Transf	ormColumnTypes(#"F	Renamed Columns", {{"year", type date}.	("unemployment", type number}))	* Quan Cattings
Imp. cod 1 Aux 2 Aux 3 Aux 4 Aux 5 Aux 6 Aux 7 Aux 8 Aux 9 Bel 10 Bel 11 Bel 13 Bel 14 Bel 15 Bel 16 Bel 17 De 18 De 19 De 20 De	untry v Ke stria v Ke stria v Stria v	y vear 1 3/1/2 1 3/1/2 1 3/1/2 1 3/1/2 1 3/1/2 1 3/1/2 1 3/1/2 1 3/1/2 1 3/1/2 2 3/1/2 2 3/1/2 2 3/1/2 2 3/1/2 2 3/1/2 2 3/1/2 2 3/1/2 3 3	unemployment 0005 0005 0007 0007 0009 010 010 010 010 011 012 012 013 014 015 016 017 018 019 010 011 012 013 014 015 016 017 018 019 011 012 013 014 015 016 017 018 019 011 012 013 014 015 016 017 018 0197 0197 0197 0197 </th <th></th> <th></th> <th>POPERTIES Name Unemployment - Economy of the Europea All Properties AIPPLIED STEPS Source Source Kenamed Columns Renamed Columns X Changed Type</th>			POPERTIES Name Unemployment - Economy of the Europea All Properties AIPPLIED STEPS Source Source Kenamed Columns Renamed Columns X Changed Type
21 De 22 De 23 De	nmark nmark nmark	3 3/1/2 3 3/1/2 3 3/1/2	8009 5.7 8010 7.6 8011 7.5			~
4 COLUN	INS, 50+ ROW	s				PREVIEW DOWNLOADED AT 2:18 PM.

Figure 9-31. Spreadsheet after renaming columns

- 4. Select Close & Load to save to a spreadsheet.
- 5. At this point, it is important to check the columns for any extraneous data types. For example, if a date column has non-date values, it will not map properly. The best way to check is to sort the columns from lowest to highest and highest to lowest, by pulling down the arrow at the top of each numeric and date column. This process may reveal a row containing only dots, as shown in Figure 9-32. If this is the case, delete this row by highlighting the row, right-clicking, and selecting delete and table rows.

S 8 5.	c) · 🔝 🔭 🖬 • =	chap9	Peuropeanunemployu	npivotmay28.	xlsx - Excel			TABLE	TOOLS	1.1					? 🖽 –	
FILE HOME able Name: Unemployment_ Resize Table	E INSERT PAGE LAVOUT FC Summarize with PivotTable Remove Duplicates Convert to Range Slicer	Export Refresh	ATA REVIEW Properties Open in Browser OUnlink	VIEW P Heade Total B Bande	OWER QUER r Row F Row L d Rows F	Y POW First Colum Last Colum Banded Col	ERPIVOT n	QUERY ilter Button	DESIGN							sei = K
Properties	Tools	Extern	al Table Data		Tab	ile Style Opt	lions					Table Style	\$			
C2 *	$ X \lor f_x =$															
A I	B C D	E	F G	н	1	J	к	L	м	N	0	P	Q	R	s	т
country 💌	Key 💌 year 📑 unemployment	*														
1																
Austria	1 3/1/2012 4.	.2														
Belgium	2 3/1/2012 7.	.2														
Denmark	3 3/1/2012 7.	.7														
Finland	4 3/1/2012 7.	.6														
France	5 3/1/2012 10.	1														
Germany	6 3/1/2012 5.	.6														
Greece	7 3/1/2012 22	.1														
Ireland	8 3/1/2012 14	.8														
1 Italy	9 3/1/2012 10.	.3														
2 Luxembourg	10 3/1/2012	5														
3 Netherlands	11 3/1/2012	5														
Portugal	12 3/1/2012 15.	.1														
5 Austria	1 3/1/2011 4	4														
Belgium	2 3/1/2011 7.	.7														
7 Denmark	3 3/1/2011 7.	.5														
8 Finland	4 3/1/2011	8														
9 France	5 3/1/2011 9	.5														
Germany	6 3/1/2011 6	.2														
1 Greece	7 3/1/2011 13	.8														
2 Ireland	8 3/1/2011 14	.7														
3 Italy	9 3/1/2011 8.	.3														
a		-														
4 9	Sheet1 Sheet2 +							4								•

Figure 9-32. Extraneous row containing ellipses at top

6. Check the formatting on the year column by selecting the column and right-clicking, making sure that a date format is selected, as shown in Figure 9-33.

Proof Table Recommended Table Proof Table Recommended Proof Table Proof Table Recommended Proof T	RLE	S · c ² · HOME		∓ SE LAYOUT FOR	MULAS	europeans DATA	REVIEW	oivotmay28.xl VIEW	sx - Excel POWER QUERY	POWERPIN	OT QUI	TABLE TOOLS ERY DES	IGN					?
C1 F F	PivotTable	Recommende PivotTables Tables	ed Table Picto	ures Online Pictures &+*	🚔 Store 🎝 My A A	pps - 🖬	Recommer Charts	ded ∰ - E Charts	PivotCl	Power View	Map • Power Map	Line Co Spa	lumn Win/ Loss rklines	Slicer Timel Filters	ine Hype	A Prink Text ks	π Equal Ω Symb	tion * sol
A B C D E F G H J K L M N O P Q R S 1 Justria 1 3/1/2005 5.10 Image: State St	C1	* 1	$\times \checkmark j$	fx 'year														
1 country • Key • vert • innemployment • 2 Austria 1 3/1/2005 5.10 • • • 4 Austria 1 3/1/2005 5.10 • • • • 4 Austria 1 3/1/2005 4.50 • <td>al. a</td> <td>A B</td> <td>с</td> <td>D</td> <td>E</td> <td>F</td> <td>G</td> <td>н</td> <td></td> <td></td> <td></td> <td>M</td> <td>N</td> <td>0</td> <td>р</td> <td>Q</td> <td>R</td> <td>s</td>	al. a	A B	с	D	E	F	G	н				M	N	0	р	Q	R	s
2 Austria 1 3/1/2005 5.10 3 Austria 1 3/1/2016 5.10 4 Austria 3 3/1/2006 5.10 5 Austria 3 3/1/2007 4.50 6 Austria 3 3/1/2007 4.50 7 Austria 3 3/1/2001 4.40 8 Austria 3 3/1/2012 4.20 9 Austria 3 3/1/2012 8.20 9 Austria 3 3/1/2010 8.20 19 Belgium 2 3/1/2007 7.70 19 Belgium 2 3/1/2010 7.20 19 Belgium 3 3/1/2012 7.20	1 count	y 🔽 Key	🖬 year 🔽 u	nemployment 💌						Format Ce	lle		?	×				
3 Avstria 1 2 //2006 5.10 4 Austria 1 3 //2007 4.50 5 Austria 1 3 //2007 4.50 6 Austria 1 3 //2007 4.50 7 Austria 1 3 //2007 4.50 8 Austria 1 3 //2007 4.50 9 Austria 1 3 //2012 4.20 9 Austria 1 3 //2006 8.40 9 Belgium 2 3 //2006 8.40 12 Belgium 2 3 //2007 7.70 13 Belgium 2 3 //2007 7.30 14 Belgium 2 3 //2007 7.30 15 Belgium 2 3 //2007 7.30 16 Belgium 2 3 //2008 6.90	2 Austri	a	1 3/1/2005	5.10						ronnat Ce	115		_					
4 Austria 1 3/1/2001 4.50 5 Austria 1 3/1/2007 4.50 6 Austria 3 3/1/2001 4.50 7 Austria 3 3/1/2001 4.40 8 Austria 3 3/1/2001 4.40 9 Austria 3 3/1/2005 8.40 9 Austria 3 1/1/2005 8.40 9 Austria 3 1/1/2005 8.40 9 Austria 3 1/1/2007 7.70 18 Belgium 2 3/1/2007 7.70 16 Belgium 3 3/1/2002 7.70	3 Austri	a	1 3/1/2006	5.10			Numbe	Alignmen	t Font Bor	der Fill	Protection							
§ Austria 1 \$/1/2007 4.50 6 Austria 1 \$/1/2009 4.50 7 Austria 1 \$/1/2012 4.50 8 Austria 1 \$/1/2012 4.20 9 Austria 1 \$/1/2016 4.10 9 Austria 2 \$/1/2006 8.40 9 Belgium 2 \$/1/2016 8.20 11 Belgium 2 \$/1/2010 8.10 9 Belgium 2 \$/1/2007 7.70 15 Belgium 2 \$/1/2008 6.50 16 Belgium 2 \$/1/2008 6.50 17 Belgium 2 \$/1/2008 6.50 18 Denmark 3 \$/1/2012 7.20 19 Denmark 3 \$/1/2012 7.00 10 Denmark 3 \$/1/2012 7.70 11 Denmark 3 \$/1/2012 7.70 12 Denmark 3 \$/1/2010 7.60 13 Denmark 3 \$/1/2010 7.60 14 Denmark 3 \$/1/2010 7.60 15 Denmark <t< td=""><td>4 Austri</td><td>a</td><td>1 3/1/2010</td><td>4.90</td><td></td><td></td><td>Cateno</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	4 Austri	a	1 3/1/2010	4.90			Cateno											
6 Austria 1 3/1/2009 4.50 7 Austria 1 3/1/2011 4.40 8 Austria 1 3/1/2011 4.40 9 Austria 1 3/1/2012 4.20 9 Austria 1 3/1/2005 8.40 9 Austria 2 3/1/2005 8.40 9 Austria 2 3/1/2005 8.40 12 Belgium 2 3/1/2005 8.40 13 Belgium 2 3/1/2005 8.40 13 Belgium 2 3/1/2005 8.40 14 Belgium 2 3/1/2007 7.70 15 Belgium 2 3/1/2007 7.30 16 Belgium 2 3/1/2012 7.70 17 Belgium 2 3/1/2012 7.70 18 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2012 7.70 10 Denmark 3 3/1/2005 5.40	5 Austri	a	1 3/1/2007	4.50			Genera	al I	Sample									
7 Austria 1 3/1/2011 4.40 Currency 8 Austria 1 3/1/2012 4.20 Dipe 10 Belgium 2 3/1/2005 8.40 Time 11 Belgium 2 3/1/2006 8.20 Strictle 12 Belgium 2 3/1/2007 7.70 Strictle 13 Belgium 2 3/1/2007 7.70 Strictle 14 Belgium 2 3/1/2007 7.70 Strictle 16 Belgium 2 3/1/2007 7.70 Strictle 17 Belgium 2 3/1/2007 7.70 Strictle 18 Belgium 2 3/1/2007 7.70 Strictle Strictle 19 Belgium 2 3/1/2008 6.50 Strictle Strictle Strictle 19 Denmark 3 3/1/2012 7.70 Strictle Strictle Strictle Strictle Strictle 19 Denmark 3 3/1/2001 7.60 Strictle	6 Austri	a	1 3/1/2009	4.50			Numbe	tr	year									
B Austria 1 3/1/2002 4.20 9 Austria 1 3/1/2008 4.10 9 Austria 1 3/1/2008 4.10 9 Austria 1 3/1/2005 6.40 9 Austria 2 3/1/2005 6.40 9 Austria 3/1/2005 6.40 Fraction 9 Austria 3/1/2005 6.40 Fraction 11 Belgium 2 3/1/2007 7.70 12 Belgium 2 3/1/2007 7.70 15 Belgium 2 3/1/2007 7.20 16 Belgium 2 3/1/2007 7.70 15 Belgium 2 3/1/2012 7.20 16 Belgium 2 3/1/2012 7.70 18 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2010 7.60 10 Denmark 3	7 Austri	a	1 3/1/2011	4.40			Accourt	nting	Type:									
9 Austria 1 3 /1/2008 4.10 Image: Strate and the s	8 Austri	a	1 3/1/2012	4.20			Date		*3/14/2012									
10 Belgium 2 2 1/2005 8.40 Fraction 11 Belgium 2 2 1/2006 8.20 0/14/12 11 Belgium 2 3/1/2010 8.10 0/14/12 1/14/14 12 Belgium 2 3/1/2007 7.70 0 0/14/12 14 Mair 1/2007 7.70 0 0/14/12 1/14/14 16 Belgium 2 3/1/2012 7.70 0 0 16 Belgium 2 3/1/2012 7.20 0 0 17 Belgium 2 3/1/2012 7.20 0 0 18 Denmark 3 3/1/2012 7.70 0	9 Austri	a	1 3/1/2008	4.10			Percen	tage	"Wednesday,	March 14, 201	12							
11 Belgium 2 2 1/2006 8.20 12 Belgium 2 3 1/2007 7.70 13 Belgium 2 3 1/2007 7.70 14 Belgium 2 3 1/2007 7.70 15 Belgium 2 3 1/2007 7.70 16 Belgium 2 3 1/2007 7.70 15 Belgium 2 3 1/2007 7.70 16 Belgium 2 3 1/2007 7.70 17 Belgium 2 3 1/2001 7.70 18 Denmark 3 3 1/2012 7.70 19 Denmark 3 3 1/2012 7.70 19 Denmark 3 3 1/2012 7.70 19 Denmark 3 3 1/2012 7.70 10 Denmark 3 3 1/2001 7.60 10 Denmark 3 3 1/2005 5.70	10 Belgiu	m	2 3/1/2005	8.40			Fractio	n	3/14/12				_					
12 Belgium 2 3/1/2010 8.10 13 Belgium 2 3/1/2017 7.70 14 Belgium 2 3/1/2017 7.70 15 Belgium 2 3/1/2007 7.70 16 Belgium 2 3/1/2007 7.70 17 Belgium 2 3/1/2012 7.70 19 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2012 7.70 10 Denmark 3 3/1/2010 7.60 12 Denmark 3 3/1/2010 7.60 12 Denmark 3 3/1/2003 5.70 12 Denmark 3 3/1/2003 5.70 12 Denmark 3 3/1/2005 5.40 13 Straw 3 3/1/2005 4.30	11 Belgiu	m	2 3/1/2006	8.20			Text	IIC .	03/14/12									
13 Belgium 2 3/1/2007 7.70 14 Belgium 2 3/1/2011 7.70 15 Belgium 2 3/1/2007 7.30 16 Belgium 2 3/1/2012 7.20 17 Belgium 2 3/1/2012 7.20 18 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2011 7.50 10 Denmark 3 3/1/2013 7.70 12 Denmark 3 3/1/2013 7.60 18 Deriver 4 3/1/2014 7.70 Date formats display doat to change: in regional date and time settings that are specified for the operating system. Formats without an atterisk 2 regional date and time settings that are specified for the operating system settings. 12 Denmark 3 3/1/2005 5.40 13 Denmark 3 3/1/2005 6.430 14 Denmark 3 3/1/2005 6.430	12 Belgiu	m	2 3/1/2010	8.10			Special		14-Mar-12					·				
14 Belgium 2 3/1/2001 7.70 15 Belgium 2 3/1/2009 7.30 16 Belgium 2 3/1/2012 7.20 17 Belgium 2 3/1/2012 7.70 18 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2012 7.70 10 Denmark 3 3/1/2012 7.70 10 Denmark 3 3/1/2012 7.70 10 Denmark 3 3/1/2010 7.60 10 Denmark 3 3/1/2005 5.70 12 Denmark 3 3/1/2005 5.40 13 Denmark 3 3/1/2005 5.40 13 Denmark 3 3/1/2005 5.40 14 Denmark 3 5/1/2005 5.40	13 Belgiu	m	2 3/1/2007	7.70			Custos		Locale (locatio	n):								
15 Belgium 2 1/1/2009 7.30 16 Belgium 2 3/1/2012 7.20 17 Belgium 2 3/1/2012 7.20 18 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2010 7.60 10 Denmark 3 3/1/2009 5.70 12 Denmark 3 3/1/2005 5.40 13 Denmark 3 3/1/2005 5.40 14 Denmark 3 3/1/2005 5.40	14 Belgiu	m	2 3/1/2011	7.70					English (Unit	d States)				~				
16 Belgium 2 3/1/2012 7.20 17 Belgium 2 3/1/2008 6.90 19 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2011 7.60 20 Denmark 3 3/1/2011 7.50 20 Denmark 3 3/1/2005 5.70 21 Denmark 3 3/1/2005 5.40 23 Denmark 3 3/1/2006 4.30 0 Arrow 0 K (ancel	15 Belgiu	m	2 3/1/2009	7.30														
17 Belgium 2 2/1/2008 6.90 18 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2012 7.70 19 Denmark 3 3/1/2010 7.60 20 Denmark 3 3/1/2011 7.50 21 Denmark 3 3/1/2009 5.70 22 Denmark 3 3/1/2005 5.40 23 Denmark 3 3/1/2005 5.40 24 Denmark 3 3/1/2005 5.40	16 Belgiu	m	2 3/1/2012	7.20														
18 Denmark 3 3 1/2012 7.70 19 Denmark 3 3 1/2010 7.60 20 Denmark 3 3 1/2011 7.50 21 Denmark 3 3 1/2005 5.70 22 Denmark 3 3 1/2005 5.70 23 Denmark 3 3 1/2005 5.40 24 Denmark 3 1/12006 4.30	17 Belgiu	m	2 3/1/2008	6.90					\$									
19 Denmark 3 3/1/2011 7.60 20 Denmark 3 3/1/2011 7.50 21 Denmark 3 3/1/2005 5.70 22 Denmark 3 3/1/2005 5.40 23 Denmark 3 3/1/2005 6.430	18 Denm	ark	3 3/1/2012	7.70			Data (a	make disalars	data and time casi	i numbers as	data walnus	Data farmate	that he air with					
20 Denmark 3 2/1/2011 7.50 21 Denmark 3 3/1/2005 5.70 22 Denmark 3 3/1/2005 5.40 23 Denmark 3 3/1/2005 4.30	19 Denm	ark	3 3/1/2010	7.60			an aste	risk (*) respon	d to changes in re	gional date an	d time settin	gs that are sp	ecified for the					
21 Denmark 3 3/1/2009 5,70 22 Denmark 3 3/1/2005 5,40 23 Denmark 3 3/1/2006 4,30 OK Cancel 0K Cancel	20 Denm	ark	3 3/1/2011	7.50			operati	ng system. Fo	rmats without an a	sterisk are no	t affected by	operating sys	tem settings.					
22 Denmark 3 3/1/2005 5.40 23 Denmark 3 3/1/2006 4.30 OK Cancel	21 Denm	ark	3 3/1/2009	5.70														
23 Denmark 3 2 3/1/2006 4.30 OK Cancel	22 Denm	ark	3 3/1/2005	5.40			· · · · · · · · · · · · · · · · · · ·											
32 December 2 and a feature an	23 Denm	ark	3 3/1/2006	4.30								OK	Cancel					
	110	and a	2 2 1+ 12023				1				_		1000					

Figure 9-33. Applying date format to year column

7. Also, check the format for the unemployment column, making sure that it has a number format with two places to the right of the decimal point, as shown in Figure 9-34.

FILE HON	C ² - □ → □ → = ME INSERT PAGE LAYO	eu DUT FORMULAS D	ropeanunemployunpivotm IATA REVIEW VIEW	wy28.adox - Excel W POWER QUERY POWERPINOT	TABLE TOOLS QUERY DESIGN			, a	Neil •
PivotTable Recon Pivo Tab	mended Table Pictures Or Pictures Or Pictures Illustra	aline tures a+- stions Apps	Recommended Charts	Image: Second	Map Ver Map Sparklines	Win/ Slicer Tim Filters	eline Hyperlink Links	4 π Equation - Ω Symbol Symbols	
D1 ···	$f_x \sim f_x$	D E		Format Cells	? 🔽	0	P Q	R S	т
2 Austria 3 Austria 4 Austria 5 Austria 6 Austria 6 Austria 9 Austria 9 Austria 9 Austria 10 Belgium 11 Belgium 12 Belgium 13 Belgium 15 Belgium 15 Belgium	1 3/1/2005 1 3/1/2006 1 3/1/2010 1 3/1/2007 1 3/1/2007 1 3/1/2001 1 3/1/2011 1 3/1/2012 1 3/1/2008 2 3/1/2006 2 3/1/2006 2 3/1/2007 2 3/1/2011 2 3/12011 2 3/12011 2 3/12011 2 3/12011	5.10 5.10 4.90 4.50 4.50 4.20 4.10 8.40 8.20 8.10 7.70 7.70 7.70 7.30 7.20	Currency Accounting Draw Percentage Paration Scientific Test Scientific Test Curtom	Sample unemployment Decimal playees: 2 0 Use 1000 Separator () Hegative numbers: 1234.10 (1234.10 (1234.10)	~				
17 Belgium 18 Denmark 19 Denmark 20 Denmark	2 3/1/2008 3 3/1/2012 3 3/1/2010 3 3/1/2011	6.90 7.70 7.60 7.50	Number is used for general display of numbers. Currency and Accounting offer specialized formating for monetary value.						
21 Denmark 22 Denmark 23 Denmark	3 3/1/2009 3 3/1/2005 3 3/1/2006 Sheet1 Sheet2 (5.70 5.40 4.30			OK Cancel				

Figure 9-34. Applying number format to unemployment column

- 8. Then select the Insert tab and click Map and Launch Power Map.
- **9.** Click the Refresh Data icon on the ribbon. This is important for applying the formatting changes.
- 10. Confirm that country is the geographic field by clicking Next, as shown in Figure 9-35.



Figure 9-35. Defining geographic field

11. Click Map Labels on the ribbon, then click the unemployment field to select the metric to be plotted and drag the legend box to the left, as shown in Figure 9-36.



Figure 9-36. Selecting unemployment metric

12. Change the aggregation of the unemployment field to no aggregation by clicking the arrow after unemployment in the Height window and selecting No Aggregation, as shown in Figure 9-37.



Figure 9-37. Changing unemployment aggregation to no aggregation

13. Right-click year and select Set as Time, as shown in Figure 9-38.



Figure 9-38. Setting the year as time field

Notice that a time control appears at the bottom of the map window, and the time appears in a box in the upper left, as shown in Figure 9-39. A time animation can be played by clicking the Play button on the time control at the bottom of the screen.



Figure 9-39. Map set up for time animation

The speed of play can be adjusted by clicking the gear icon on the upper right and clicking Scene Options to change the scene duration time and effects, as shown in Figure 9-40. The Speed slider at the bottom of the pane can be adjusted to slow down or speed up the transitions.



Figure 9-40. Setting scene duration

Summary

This chapter has demonstrated how to load data into Power Query, how to shape the data in the Query Editor, and then how to load it in Power Map to create a visual story about the data. Some of the capabilities of Power Map were demonstrated, such as creating 2D and 3D visualizations, as well as creating animations to show variations over time.

CHAPTER 10

Statistical Calculations

Excel provides a variety of built-in tools for performing statistical analysis. This chapter highlights some of those tools, including basic charting and the Excel Analysis ToolPak. The difference between descriptive and inferential statistics is discussed, and Excel statistical functions are introduced.

Recommended Analytical Tools in 2013

Excel 2013 tries to anticipate your needs by recommending conditional formatting, charts, tools, totals, and Pivot Tables. When data is highlighted, or when Ctrl+Q is pressed, with any cell in the data range selected, a pop-up window with various formatting options is displayed, as shown in Figure 10-1. Moving the cursor over the conditional formatting options will show a preview of that conditional formatting on the data.



Figure 10-1. Recommended format and analytics options
Clicking Totals in the pop-up window will cause the Totals menu to be displayed, which is shown in Figure 10-2.



Figure 10-2. Totals and statistical options

Scrolling right to highlight % Total will produce the results shown in Figure 10-3, in which the percent of the total is shown to the right of each item.

FLE	HOME IN	SERT PAG	SE LAYOUT	FORMULA	S DATA	REVIE	W W	NEW P	OWER QUE	RY POW	ERPIVOT	r FINERA	FIII.	EB.	-		53 6	∑ AutoS	um -	A- 4		ei - 🏳
D Bac	opy -	Calibri	- 11	• A A		41.	@ Wra	ap Text	Ge	neral		1 E		-	E	E L		Fill -		ZY		
Paste	ormat Painter	BIU	• 🖽 •	<u>⊘</u> - <u>A</u> -	E = 3	任任	🗒 Me	rge & Centi	s • \$	- % *	·8 -8	Conditiona	 Format a: Table * 	Cell Styles *	Insert	Delete Fo	ormat	& Clear -	- F	int & Fin	id 8t ect ≠	
Clipbo	oard 15		Font	r,		Align	ment		5	Number	5		Styles			Cells			Editin	9		^
C12	· : 3	× 2 1	fr 97																			~
										1 (14)			1.00		-			- Ti - S				1
A 1	B	C	D	E	F	G	н	1	1	K	L	M	N	_	0	Р	Q	F	R	S	т	0 -
2																						
3		84	5.23%																			
4	1	75	4.67%																			
5		79	4.92%																			
6		92	5.73%																			
7		83	5.17%																			
8		65	4.05%																			
9		78	4.86%																			
10	_	66	4.11%																			
1	_	93	5.79%																			
12	_	97	6.04%																			
3		88	5.48%																			
4	-	76	4.73%																			
		87	5.4270																			
7		92	5.7370																			
8		68	4 23%																			
19		72	4.48%																			
0		62	3,86%																			
21		78	4.86%																			
22		86	5.35%																			
3																						
	1	1					-	_			-											
0.0	Sheet1	Sheet2	÷		_					_			4				_					E.

Figure 10-3. Percent totals

Customizing the Status Bar

Notice that the status bar at the bottom of the Excel window shows statistics related to the highlighted data, such as the average, count, and sum. You can customize the values shown by right-clicking the status bar and selecting the items you want displayed, as shown in Figure 10-4.



Figure 10-4. Customize Status Bar menu

If Minimum and Maximum, for example, are selected on the Customize Status Bar menu, those values will be added to the status bar, as shown in Figure 10-5.

	Cut	Calibri	- 11	- A A	==	- W/-	₿w	rap Text	Ge	ineral				P				AutoSum	ZY	#	
te 💉	Format Paint	er B I	u - 🗉 -	<u>⊳</u> - <u>∧</u> -	==		E 🗒 M	erge & Center	- 5	- % •	10 -00 -00	Conditional I Formatting *	Format as Table =	Cell Styles •	Insert *	Delete F	ormat	Clear*	Sort & F Filter = S	rind & ielect =	
Cript	• :	XV	fx 65				printeri			HUBIOLI			iques -			CERS			ming		
A	B	с	D	E	F	G	н	1	J	к	L	м	N		0	p	Q	R	s	т	1
		8	1																		
		7:	9																		-
		9.	2																		-
		6	5																		
		6	5																		
		9	7																		
	-	8	3			_															
FORM	ATTING	CHARTS	TOTALS TAB	ILES SI	PARKLINES																
				1																	
Data	Bars Colo	Icon Set	Greater To	op 10%	Clear																
Condit	tional Format	ting uses rules	to highlight inte	resting dat	а.																

Figure 10-5. Status bar with Min and Max added

Inferential Statistics

Sampling involves examining a small portion of a large population in order to draw conclusions about the entire population. This is called inferential statistics. For a sample to be representative of the larger population, the members of the sample must be selected randomly, meaning that every member of the population has the same chance of being included as any other member. The objective is to draw reliable conclusions about the population from the sample. Determining the sample size is critical. The margin of error for a sample is calculated as one divided by the square root of the sample size. Thus, as the sample size increases, the margin of error decreases.

Generally, a sample is used because it is not economically feasible to survey an entire population. One of the promises of big data is that if the data is available, an entire population can be processed for more accurate results.

Review of Descriptive Statistics

Descriptive statistics summarize data with a few key calculated values. Figure 10-6 shows a normal distribution or, bell, curve. The frequency with which any value occurs is represented by the height of the curve. For example, if a coin is tossed ten times over multiple trials, the average number of heads will be five. This is the result that will occur most frequently and is represented by the highest point of the curve. The next most common occurrence will be six heads and four tails or six tails and four heads. Rarely would there be nine heads and one tail.



Figure 10-6. Bell curve

Calculating Descriptive Statistics

The most common calculated statistical values are mean, median, and mode.

The mean, the average or arithmetic mean, is computed by adding all the values and dividing by the number of values.

The middle value is called the median—the data point at which 50% of the values fall above it and 50% of the values fall below it.

The most frequent value is called the mode. In a bell curve, the median, mode, and mean are equal.

Measures of Dispersion

Many distributions don't fit neatly into a bell curve, however, so we need measures of dispersion. The variance measures how far a set of numbers is spread out. A small variance indicates that the data points tend to be very close to the mean and, hence, to each other, while a high variance indicates that the data points are very spread out around the mean and from each other.

One measure of dispersion is the standard deviation, which is calculated by computing the deviation of each value from the arithmetic mean and squaring them. The arithmetic mean of these squared values is calculated, and the square root taken to calculate the standard deviation is represented by the Greek letter sigma.

The 68-95-99 rule is used to remember what percentage of values fall within a range around the mean of a normal distribution. In other words, 68.27% of the values are within the range of plus or minus one standard deviation; 95.45% of the values are in a range of plus or minus two standard deviations; and 99.73% of the values are within a range of plus or minus three standard deviations. This is sometimes called the Three Sigma Rule of Thumb. Remember that this rule only applies to a normal distribution. Notice the percentages shown in the illustration in Figure 10-6 that correspond to this rule.

For example, in an analysis of home values, if all home sales fall within a fairly narrow range, such as \$400,000 to \$600,000, there would be a relatively small variance and standard deviation. However, if a couple of high-end homes sell for more than \$1,000,000, there would be a larger variance and standard deviation. That is why the graphical approaches described in the remainder of this chapter are useful for showing relationships among values.

Excel Statistical Functions

Excel provides numerous functions to calculate statistical values. Here are a few of the more common ones.

Average(range): Calculates the arithmetic mean of a range of data or a list of discrete values. It ignores empty cells or cells containing text.

Averagea(range): Same as average, except that it assigns a value of zero to cells containing text.

Averageif(range, condition): Returns an arithmetic mean based on a conditional statement enclosed in quotes. It does not include empty cells or cells containing text in calculations.

Averageifs(range, condition1, condition2): Returns an arithmetic mean based on multiple conditional statements. It does not include empty cells or cells containing text in calculations.

Median(range): Calculates the media or middle value.

Mode(range): Calculates the most frequent value, excluding empty cells and cells containing text not counted.

Charting Data

Excel provides several ways of viewing data graphically. It is important to select the best graphical tool to accurately portray the data, as described in the list below.

- Column and bar charts plot a category variable on one axis and a numeric variable on the other.
- An XY scatter chart shows the relationship between two numeric values.
- Frequency distributions deal with one variable only—it shows how many instances there are of each value. It groups individual measurements into classes or bins. In its most common form, the variable's values are on the horizontal, or x, axis and the frequency of instances on the vertical, or y, axis
- A histogram is a graphical representation of the distribution of numerical data.

Excel Analysis ToolPak

This section will cover how to use the Excel Analysis ToolPak, which is included with Excel. The Analysis ToolPak can be used to perform sophisticated statistical calculations that would be very time-consuming to program in by hand.

Enabling the Excel Analysis ToolPak

Before use, it must be activated.

To enable the Excel Analysis ToolPak, follow these steps:

- 1. Click the File tab to see the Backstage View.
- 2. Click Options at the bottom of the left pane.

- **3.** Select Add-Ins on the left pane.
- 4. At the bottom of the window, select Excel Add-Ins and click Go, as shown in Figure 10-7.

	Excel O	ptions	?	
General	View and manage Microsoft Office	Add-ins.		
Proofing	Add-ins			
Save	Name 🛧	location	Type	-
Language.	Active Application Add-ins	200000	1)00	1
Language	Microsoft Office PowerPivot for Excel 2013	C:_dd-in\PowerPivotExcelClientAddIn dll	COM Add-in	
Advanced	Microsoft Power Map	C:\review for Excel\ExcelPlugInShellExt.dll	COM Add-in	
	Microsoft Power Map for Excel	C:\ Excel Add-in\EXCELPLUGINSHELL.DLL	COM Add-in	
Customize Ribbon	Microsoft Power Query for Excel	C:\Windows\System32\mscoree.dll	COM Add-in	
Quick Access Toolbar	Power View	C:\Add-in\AdHocReportingExcelClient.dll	COM Add-in	
Add-Ins	Inactive Application Add-ins			
Truct Center	Analysis ToolPak	C:\fice15\Library\Analysis\ANALYS32.XLL	Excel Add-in	
indst Center	Analysis ToolPak - VBA	C:\e15\Library\Analysis\ATPVBAEN.XLAM	Excel Add-in	
	Date (XML)	C:\icrosoft shared\Smart Tag\MOFL.DLL	Action	
	Euro Currency Tools	C:\ice\Office15\Library\EUROTOOL.XLAM	Excel Add-in	
	Inquire	C:\ft Office\Office15\DCF\NativeShim.dll	COM Add-in	
	Microsoft Actions Pane 3		XML Expansion Pack	k
	Solver Add-in	C:\ffice15\Library\SOLVER\SOLVER.XLAM	Excel Add-in	
	Team Foundation Add-in	C:\erver\11.0\amd64\TFSOfficeAdd-in.dll	COM Add-in	
	Document Related Add-ins			
	Add-in: Microsoft Office PowerPiv	vot for Excel 2013		_
	Publisher: Microsoft Corporation			
	Compatibility: No compatibility informat	tion available		
	Location: C:\Program Files\Microsol	ft Office\Office15\ADDINS\PowerPivot Excel Ad	ld-in\	
	Description: Microsoft Office PowerPiv	vot for Excel 2013		
	Manage: Excel Add-ins	io		
] []			OK Car	nce

Figure 10-7. Excel Add-Ins screen

5. Select Analysis TookPak, as shown in Figure 10-8, and click OK.

	Add-Ins	?	×
Add-Ins available: Analysis ToolPak Analysis ToolPak Euro Currency To Solver Add-in	- VBA pols	OK Cano <u>B</u> rows A <u>u</u> tomat	el :e
	~		
Analysis ToolPak Provides data anal engineering analy	ysis tools for st sis	atistical an	d

Figure 10-8. Add-Ins pop-up

After activation, the Analysis ToolPak can now be accessed by clicking Data Analysis on the right side of the Data tab in Excel.

A Simple Example

The following example will show how to use the Excel Analysis ToolPak to analyze some sample test scores.

EXERCISE 10-1. USING THE ANALYSIS TOOLPAK

Follow these steps to analyze the test score:

1. Click Data Analysis on the right side of the Data tab. When the Data Analysis windows appears, select Descriptive Statistics, as shown in Figure 10-9.

FILE H	• <> IOME	NSERT	PAGE L	AYOUT	FORM	IULAS	DATA	REVIEW	VIEW	POV	Boi VER QUER	ok1 - Erce Y PO	WERPINOT				1						Neil	// /P
C From Acce	SS [From Sou Get Exter	Other rces - Conal Data	Existing	Refrest All -	Con Prop Edit Connectio	nections serties Links ms	21 ZA ZI Sort	Filter	Res Res Ad	ar ipply vanced	Text to Columns	Flash Fill C	Remove Ouplicates 1	Data /alidation Data 1	Consoli cols	date Wha Analy	? E t-lf Relati sis *	onships	Group U	Ingroup Su Outline	btotal	Pata Ani	Analysis alysis	,
C3	* 1	\times	/ fx	84																				
A	8	с	D	E	1	G	н	1	1	к	L	м	N	0	P	Q	R	s	T	U	V	W	х	F
2		84																						
4 5		75														2 4								
7		83							_	Analysis	Tools	4	Jata Ana	lysis										
8 9		65 78								Anova	Two-Facto	r Without	Replication	n	^	OK								
10		66 93								Covaria	ance ative Statist	ics			- 12	Help								
12		97 88								Expon F-Test	Ivvo-Sampl	e for Varia	ances			Deb								
14		76 87							_	Histog	ram Average													_
16		92								Rando	m Number	Generatio	in		v									
18		68																						
20		62																						
21 22		78																-						-
23																								
25																								
26 27																								
28	Sh	eet1	(+)											i										
READY	30		T	_	-	_			-				-	AVER	AGE 80.3		COUNT: 20	SUM: 1	606 F		<u> </u>	1+		85%

Figure 10-9. Data Analysis tools

2. Click OK. In the Descriptive Statistics dialog box, specify the input range and the output range and select summary statistics, as shown in Figure 10-10, and click OK.

Input		
Input Range:	c3:c22	ОК
Grouped By:	Columns	Cancel
		Help
✓ Labels in first row		
Output options		
	a3 🎼	
Output Range:	95 Hits	
• New Worksheet Ply:		
🔘 New <u>W</u> orkbook		
✓ Summary statistics		
Confidence Level for M	ean: 95 %	
Kth Largest:	1	
Kth Smallest	1	

Figure 10-10. Descriptive Statistics dialog

The results are shown in Figure 10-11.

FILE H	• ↔	INSERT	PAGE LA	YOUT	FORM	IULAS	DATA	REVIEW	VIEW	POWER C	Book1 - E	xcel POWERPIV	от								7 1	E - Nei	8 1 - F
G From Acce From Web	From Sour Get Extern	Other Exi rces • Conn nal Data	isting	Refresh All •	Con Prop Edit Connectio	perties Links ons	21 AZ	Filter Sort & Fil	📡 Clear 🐑 Reapply 🏷 Advance ter	Tex Colu	to Flash	Remove	Data s Validation Data	Consolid Tools	ate What Analys	? t-lf Relati	onships	Group Ur	outline	btotal 5	Ib Data	Analysis	
M3	• 1	XV	fx																				
a A	8	c	D	£		G	н	1	J	ĸ	L 1	A N	0	P	Q	R	5	T	U	٧	W	х	1 24
2										-													
1		84				8	4																
		75								_													
		79				Mean	80.10526																
		92				Standard	2.376659																
		65				Mode	92																
		78				Standard	10.35952																
0		66				Sample V	107.3216																
1		93				Kurtosis	-0.99237																
2		97				Skewnes:	-0.17972																
3		88				Range	35																
4		76				Minimum	62																
5		87				Maximum	97																
6		92				Sum	1522																
		85				Count	19																
		08																					
		62																					
1		78																					
2		86																					
3																							
4																							
5																							
6																							
7																							
4 3	She	net1 (Ð											11									
-	-			_	_		_	_		_	_	_	_	- LSL	_	_	_					-	100

Figure 10-11. Descriptive Statistics output

3. To plot a histogram using the Data Analysis ToolPak, go to the Data tab and click Data Analytics. Select Histogram, as shown in Figure 10-12.

Data Analysis		? ×
<u>A</u> nalysis Tools		OK
Anova: Single Factor	^	JA
Anova: Two-Factor With Replication		Cancel
Anova: Two-Factor Without Replication		
Correlation		Halp
Covariance		Help
Descriptive Statistics		
Exponential Smoothing		
F-Test Two-Sample for Variances		
Fourier Analysis		
Histogram	~	

Figure 10-12. Selecting Histogram from the Data Analysis dialog

4. Define the range and select chart output, as shown in Figure 10-13, and click OK.

	Histogram		?
put			OK
nput Range:	\$C\$3:\$C\$22	1	
in Range:		1	Cancel
/ Labels			Help
<u>O</u> utput Range: New Worksheet <u>Ply:</u> New <u>W</u> orkbook Pareto (sorted histograming)	SJS3		

Figure 10-13. Histogram dialog

The results are shown in Figure 10-14. This operation sets up bins, or categories, for classifying the data, which are used for plotting the histogram.

FILE	ち・ ぐ HOME	- 🛄	PAGE L	AYOUT	FORM	IULAS	DATA	REVIEW	VIEW	POW	chap10es VER QUE	ample1.dr	ox - Excel									1	Nei	
From	Access [Web From Text Sou Get Exter	o Other inces •	Existing Connections	Refresh All •	Con Proj Edit	perties Links ons	2↓ AZA Z↓ Sort	Filte	The Cle The Real Adda	ar pply vanced	Text to Column	Flash Fill (Remove Duplicates	Data Validation Data	Consolida Tools	ate What Analys	-If Relati	ionships	Group U	Outline	btotal	Data Ana	Analysis itysis	~
J3	•	\times	$\sqrt{f_x}$	Bin																				~
A	8	с	D	E		G	н	1	J	К	L	м	N	0	P	Q	R	5	т	U	٧	W	х	1
1																								
3		84				8	4		Bin	Frequency	/													
4		75	5						62	1	1													
5		79				Mean	80.10526		70.75	-	8						Hist	ograr	n		_			
6		93				Standard	2.376659		79.5												-			
		61	5			Median	79		88.25	-											-			
9		75	2			Standard	10 35962	-	more	-		-			C C			-			-			
10		64	5			Sample V	107.3216								B 4	-								
11		95	5			Kurtosis	-0.99237								£ 2					E Freque	ency			
12		93	7			Skewness	-0.17972								0				-					
13		88	8			Range	35								6	2 70.7	5 79.5	88.25	More					
14		76	5			Minimum	62										Bin							
15		83	7			Maximum	97													-				
16		93	2			Sum	1522													C	sart Area			
17		85	5			Count	19														_			
18		68																						
19		1.																						
20		71																						
22		24																						
23																								
24																								
25																								
26																								
27																								
28	Ch.	net1	æ	_	_		_							-	[A]									
DEADY	SI	1122	Ð	_	_	_	_	_	_	_	_	_	_		115040	6.22.6	0.00	CURE.	2105	aan 980			_	
READT		-					_				_		_		AVERAG	6555 (200MI: 12	SUME	21303					▼ 6576

Figure 10-14. Histogram plot

Other Analysis ToolPak Functions

The Analysis ToolPak includes other functions. Some of the most useful ones are listed following:

- The Anova tools provide different types of variance analysis. They test the randomness of the selection of two samples.
- The correlation coefficient is a value between -1 and +1 that measures how strongly two variables are related to each other. A +1 indicates a perfect positive correlation, meaning that the two variables move together. A -1 indicates a perfect negative correlation, meaning that the two variables move in opposition to each other. A positive correlation does not prove cause and effect. The Excel CORREL function can also be used to calculate correlation.
- Covariance is a measure of how much two random variables change together.
- Exponential smoothing is like a moving average used to forecast the next value in a time series.

Using a Pivot Table to Create a Histogram

A Pivot Table can also be used to create a more customized histogram.

EXERCISE 10-2. A CUSTOMIZED TEST SCORE HISTOGRAM EXAMPLE

Create a histogram that shows test score distributions by following these steps:

- 1. Highlight the data to be plotted. Select the Insert tab and click PivotTable.
- 2. Confirm the range and put the Pivot Table in the same worksheet with the location of F3, as shown in Figure 10-15.

HC PivotTable Recc Pin Ta	DME INSI mmended rotTables ables	ERT PAGE	LAYOUT s Online Pictures Rustrations	FORM	ULAS (Store My Apps Apps	ATA	REVIEW V	chap EW POWE 네네 - 플 - 위 네시 - 쇼스 - 레 네 - 는 - Charts	R QUERY PivotCl	POWERP	VOT Maj s Power	Lin Map	e Colum Sparklin	Win/ Loss	Slicer Timeline Filters	Hyperlink	A Test	T Equation Ω Symbol Symbols	Neil -
C2	- : [X	c Jx	score	F	F	G	н	T. I	a i	к	1	м	N	0	P	0	R	s	ти
1		-		-	-					-									, ,
2 3 4 5 6 7 8 9 9 10 11 12 13 14 15 16 17	5	2000 - 20						Choose the © Select Is Use as (Choose wh) New V © Data Io Choose wh) Add th	e data that y a table or n ble/Range: n external d thoose Con innection n iere you war Vorksheet ng Worksheet ration: F3 ether you w his data to t	Create + ou want to : snge Sheet2ISCS ata source nection at the PivotT et et	ivot i ab nalyze 2:5C522 ible report ce multiple el	to be plac tables OK	ed Cance						
19 20 21 22 23		68 72 62 78 86																	
· · ·	Sheet1	Sheet2	۲									1							

Figure 10-15. Create PivotTable dialog

X	8 5	· C3 ·	∃× ≖		chap10e	xample1au	ig21.xlsx - E	xcel				PIVOTTAB	LE TOOLS				? 🗉	- 5 3
	E HO	OME IN	SERT PAGE	LAYOUT F	ORMULAS DATA	REVIEW	N VIEV	V POV	VER QUERY	POWERS	AN TOVIS	VALVZE	DESIGN					Neil • O
PivotT Pivot	able Nam Table2 ptions * rotTable	e Active	Field: Dril Id Settings Dow Active Field	v ↑ +∃ I Drill +∃ m Up + d	→ Group Selection Image of the selection Image o	Insert Slicer T	Insert imeline Co Filter	Filter	Refresh Chan • Sor Data	ge Data urce *	Clear • Select • Move Pix Action	votTable	院 Fields, Ite 前入 OLAP Too 中間 Relations Calcul	ms, & Sets * ols * hips ations	PivotChart R	ecommended PrivotTables	Field +/- Field Show	Field saders
F3		• : :	$\times \checkmark f_x$	Row Labe	els													,
4	A	8	с	DE	F	G	н	1	1	к	L	M	N	0	P [A)		
1																Pivot	able Fields	+ >
2			scores		-											Choose fi	elds to add to report:	- da -
3			84		Row Labels 👻													
4			75		62											✓ score	6 ()	
5			79		65											MORE TA	BLES	
7			92		60													
0			60		72													
0			78		75													
10			66		76													
11			93		78													
12			97		79													
13			88		83													
14			76		84											Drag field	ls between areas belo	W/S
15			87		85											-		1.10.10.10
16			92		86											T PLIE	NS II COI	LUMINS
17			85		87													
18			68		88													
19			72		92											= 0000	C	100
20			62		93											=1 KOW	5 Z VAL	UES
21			78		97											scores	•	
22			86		Grand Total													
23																		
		Sheet	Sheet2	÷											•	Defer	Layout Update	UPDATE
READ	r			-											E	I 🔘 🛛		-+ 100

3. Drag scores to the Rows box, as shown in Figure 10-16.

Figure 10-16. Test scores as rows

FILE HOI	ME II	NSERT PAGE	LAYOUT	ORMULAS	DATA	A REVIEW	V VIEW	PO	WER C	2UERY	POWERS	NVOT	ANALYZE	DESIGN					N	el - 10
NotTable Name	a Active	Field:	A =	→ Group Sele	ection			12	L F	3	E.	🐺 Clear -	12	R Fields, It	tems, & Sets *	10			t_ effe	1
wotTable2	scores		1 T	Ungroup		and to	- Y	927		-C	田間	E Select		IT OLAP T	ools *	123	3		E	
Options *	To Fie	Id Settings Dea	Un Un	Group Fiel	d	Slicer T	insert imeline Co	Filter	Ket	resh Cha	nge Data	E Move	PivotTable	Relation	ships	PrvotCr	hart Kec	ommended Held	+/- Held	
PivotTable		Active Fiel	d	Group		onces 11	Filter		·	Dati	a	Act	ions	Calor	ulations		Tool	is S	show	
F7 .	* 1	$\times \checkmark f_x$	68																	
A	В	C	DE	F	5	G	н	1		J	κ	ι	М	N	0	Р				
																		PivotTable Fi	elds	* >
		scores																Choose fields to add to	o report	8.
r .		84		Row Lab	els -	-												Choose menus to more to	o report.	24
6		75		62	Cali	bri - 11	* A A	5 - 9	6 .									✓ scores		
		79		65	в	1 = 0	· A · FR	2	19 5									MORE TABLES		
		92		66	-					_								mone more		
		83		68	0.	Carri		-												
		65		72	46	Coby														
6		78		75		Eormat Cell	5													
0		66		76	B	Refresh														
		93		78		Sort		-												
2		97		79		Filter														
3		88		83		riger		-										Dran fields hetween a	neas helows	
1		/0		84	1	Subtotal "sc	ores"											Diag news outricer a	reas provin	
5		07		85		Expand/Coll	Japse)	-										▼ FILTERS	III COLUM	INS
7		95		97	翻	Group														
0		69		99	ál)	Unaroup														
0		72		92	-	Maria												-		
0		62		93	12.1	Move												= ROWS	Σ VALUES	5
1		78		97	×	Remove "so	ores"											scores 💌		
2		86		Grand To	0	Field Setting	gs	1												
3						PivotTable (Options													
		- r			101	Hide Field L	ist	-										Defect and Hode	as 1	

4. Right-click the PivotTable column and select Group, as shown in Figure 10-17.

Figure 10-17. Selecting Group from the context menu

5. By default, the Grouping dialog box sets the lower range to the lowest score and the upper range to the highest score, with a default interval of 10, as shown in Figure 10-18. Click OK.

Grouping) ? ×
Auto	
✓ <u>Starting</u> at:	62
✓ Ending at:	97
<u>B</u> y:	10
ОК	Cancel

Figure 10-18. Grouping dialog

6. Drag the scores to the Value box. The result indicating the frequency for each grouping category is shown in Figure 10-19. Notice that the categories are approximately what the grading categories would be: 92 to 100 would be an A, 82 to 91 would be a B, etc.

08	. .	∋× ≠		chap10	example1aug21.xlsx -	Excel				PIVOTTAB	LE TOOLS	1.			0	· • •	5 X
FILE	HOME	INSERT PAGE	LAYOUT F	FORMULAS DATA	REVIEW VIE	W POV	VER QUERY	POWER	PIVOT	ANALYZE	DESIGN					Ne	14 P
PivotTable Na	me: Active	e Field:		→ Group Selection		1.2	1	(III)	🐺 Clear •		R Fields	s, Items, & Sets -				-	
PivotTable2	score	5	(T) =	Congroup		9 TT	LØ	E	E Select		The OLAP	Tools -	1				
PivotTable	• Fo Fi	ield Settings Dow	n Up-	Group Field	Slicer Timeline C	Filter	Refresh Ch	iange Data Source * ta	Move i	PivotTable	Relati	ionships Noutations	PivotCha	rt Reco Pin Tools	votTables List Bi	*/- Field ittons Header Show	5
F7	• 1	× × fr	92-101														~
11		JA							v	w 11		N	0	171			
- A	D	C	U E	F	6	n		1	ĸ	L	M	N	0	-1-	PivotTable F	olds	+ X
2		coroc													i ivotrabie i	icius	
2		scores 84		Row Labels	Count of scores										Choose fields to add t	o report:	Q٠ -
4		75		62-71	count of scores												
5		79		72-81	6										Scores		
6		92		82-91	6										MORE TABLES		
7		83		92-101	4												
8		65		Grand Total	20												
9		78															
10		66															
11		93															
12		97															
13		88															
14		76													Drag fields between a	reas below:	
15		87													T FILTERS	III COLUMN	NS
16		92															
17		85															
18		68															
19		72													= ROWS	E VALUES	
20	-	62													10000	Count of co	
21		78													2010	Count of sc	cres -
22		86															
25															0.00		
()	Shee	t1 Sheet2	۲					4						Þ	Defer Layout Upd	ste	JPDATE
READY														Ħ	面 門	1	+ 100%
							-	-		-	_						

Figure 10-19. Grouping categories

7. To plot a histogram, click the Insert tab and select Column Chart, 2D Column. The result, which plots the frequency for each category, is shown in Figure 10-20.

FILE H	o • <² · ∃ HOME INS	X = SERT PAGE	LAYOUT FORM	chap1 MULAS DATA	0example1aug21.xtsx - A REVIEW VIE	Excel W POWER	QUERY	POWERPIVOT	ANALYZ	PIVOTCHART I DESIGN	OOLS FORM	TAT			r tel - 9 Nel - 5
Chart Name: Chart 1 Options PivotChart	Active Field:	ttings Down	↑ *∃ Expand I Drill Up • ∃ Collapse Field	Field Insert Slicer	Insert Filter Timeline Connections Filter	Refresh Chai Data	nge Data (ource *	Clear Move • Chart Actions	Fields, Items, & Sets *	OLAP Relat Tools =	3 ionships	Field Fiel List Butto Show/Hide	ld ns =		
Chart 1	• = 2	< v fx													
A h	В	с	D E	F	G	н	1	J K	L	М	N	0	1 -	DivetChart F	alda - X
1		scores												PivotChart F	leids
3		84		Row Labels	Count of scores									Choose fields to add t	o report:
4		75		62-71	4										
5		79		72-81	6									Scores	
6		92		82-91	6										
7		83		92-101	4										
8		65		Grand Total	20										
9		78										2			
10		66			Co	unt of scores						+			
11		93						Tota							
12		97						1018				-			
13		88			7 .										
14		76			6									Drag fields between i	reas below:
15		87			5		_							T FILTERS	III LEGEND (SERIES
16		92			4										
17		85			3		_				Total				
18		68			2	_			5 P	_					
19		72			1	_								AXIS (CATEG	Σ VALUES
20		62			0									scores T	Count of scores T
21		78				62-71	72-8	1 1	2-91	92-101					
22		80			500	res *									
2.5		_													
C 3	Sheet1	Sheet2	+				1 4						Þ	Defer Layout Upd	UPDATI
READY													Ħ	💷 😐	+ 10

Figure 10-20. Histogram

- 8. The chart can be tweaked by clicking a column and pressing Ctrl+1 to open the Format Data Series pane on the right side of the screen. Adjust the Gap Width slider to zero to remove the space between columns.
- **9.** The title can be modified by clicking the title, which now says Total, and changing it to "Frequency Distribution of Test Scores," as shown in Figure 10-21.

	· <> - []		VOLIT FOR	cha NULAS DATA	p10example1pi	vitableads	x - Excel	NIERV	DOMEDDIVOT	ANALY	PIVOTCHART TO	EOPMAT		? E - 8 3
Paste For Clipboar	y * nat Painter d rs	B I U -	- <u>11</u> - A* = - <u>2</u> - A int		Alignment	Wrap Text Merge & I	Center -	General \$ - % Num	• • • • • • • • • • • • • • • • • • •	Conditional Formatting *	Format as Cel Table - Style Styles		elete Format	Z AutoSum · 27 III © Fill - Sort & Find & Clear · Filter · Select · Editing
Chart 1	- : >	$\checkmark f_x$	=SERIES(Shee	t21\$6\$3,Sheet2	1\$F\$4:\$F\$8,5	heet2!\$G	\$4:\$6\$8,1)							
A A	В	C D	E	F	G	н	1	J	к	L	M	4 0	5	Format Data Series * *
3 4		84 75		Row Labels ¥ 62-71	Count of 84 4									SERIES OPTIONS *
5 6 7		79 92 83		72-81 82-91 92-101	6 5 4		Count of 84							 SERIES OPTIONS
8 9		65 78		Grand Total	19	_		Freq	uency Dist	tribution o	of Test Score	s		Plot Series On (B) Erimany Axis
10 11 12		93 97						1		1	0			Series Qverlap
13 14		88 76					1				-	. To	cal	Gap Width
15 16 17		92												
18 19		68 72					62-	71	72-81	82-91	92-10	1		
20 21 22		62 78 86												
23	Sheet1	Sheet2	(+)					1.4						
READY														■ ■ ■+ 100*

Figure 10-21. Histogram with title

Scatter Chart

It is possible to see the visual relationship between two variables by plotting a scatter chart and then allowing the program to calculate the R-squared value. R-squared is a statistical measure of how close the data are to the fitted regression line—the best-fitting straight line through the points. It is also known as the coefficient of determination. R-squared values range from 0 to 100. An R-squared value of 100 means that all movements of one value are completely explained by movements of the other variable. X is the independent variable, and y is the dependent variable.

EXERCISE 10-3. GRADE-ABSENCE RELATIONSHIP SCATTER CHART EXAMPLE

This example explores whether there is a relationship between class absences and numeric grades, with 1 being the lowest grade and 4 the highest. The grade data is shown in Figure 10-22.

X) FILE	н н	• c≯ • [IOME IN	💷 💌	PAGE LA	YOUT F	FORMUL	AS DATA	REVI	ew view	chap10e POWEI	example1pivot R QUERY	table.xisx - Ex POWERPIVO	ccel T			1				? 🗹	- 8 ×
Î	X cu		Calibr		* 11 *	A" A"	==	21-	📴 Wrap T	ext	General					**	<	∑ AutoSun	· AT	46	
Paste	Co de	ру -	BZ	U -	11. 0	- A -	===	42.42	E Merne	Center *	5 - %	, +2 -22	Condition	al Format a	s Cell	Insert Delet	e Format	Fill -	Sort 8	k Find &	
*	For For	mat Painter			-	-			and the second second				Formattin	g* Table*	Styles *	• •	•	Clear *	Filter	* Select *	
	Clipboa	rd r	9 (c)	Fe	ont	r ₂		Align	ment	15	Num	ber 5		Styles		Cell	s		Editing		^
D19		• 1	XV	fx																	~
2	A	в	C	D	E	F	G	н	1	J	K	L	M	N	0	P	Q	R	S	т	U 🔺
1																					
2																					
3																					
4																					
5		absences	grade																		
6		8	1																		
/			1								-										
8		0	2																		
30			2																		
11			3																		
12		2	4																		
13		1	4																		
14		7	1																		
15		1	4																		
16																					
17																					
18					-																
19																					
20																					
21																					
22																					
25																					
		Sheet1	She	et2 S	heet3 S	heet4	Sheet5	(i)						4							•
READ	6			l.				-	11.14									## III	.		+ 100%

Figure 10-22. Grade and absence data

Follow these steps:

1. Highlight the data, click the Insert tab, and select Scatter (X, Y), as shown in Figure 10-23.

1 8 5	• 👌 🔳) > =	-					chap10ex	mple1pivott	ableadsx - Exce	i.							? 🗉	- 8 >
FILE	IOME INSE	ERT PA	GE LAYOUT	FORMUL	AS DATA	REVIEW	VIEW	POWER	QUERY P	OWERPIVOT									Neil 4 O
PivotTable Rei	commended T ivotTables Tables	Table Pic	tures Online Pictures Illustration		Store 🖸 My Apps 📼 💽 Apps	Recomme Chart	ended 🔊 -	■・火· M・hi 上・ arts	PivotChart	Power View Reports Pou	Map • •	Line Colur Sparkl	mn Win/ Loss	Slicer Tir Filter	meline H	lyperlink Links	4 π Text	Equation + Symbol Symbols	
85	• : ×	V .	f _X =She	et4!A5				Insert Se	atter (X, Y)	or Bubble Cha	rt								
A 1	В	с	DE	F	G	н	1	relations	hip between	snow the sets of values.	м	N	0	P	Q	R	s	т	U [
1 2 3 4								Click the types of available the icons	arrow to see scatter and b and pause th to see a pre-	the different ubble charts te pointer on view in your									
5	absences g	rade						Gocomo	n										
6	8	1																	
8	6	2																	
9	5	2																	
10	4	3																	
11	3	3																	
13	1	4																	
14	7	1																	
15	1	4																	
16																			
17																			
19																		-	
20																			
21																			
22																			-
23																			
()	Sheet1	Sheet2	Sheet3	Sheet4	Sheet5	۲					1	4							Þ
READY					_	-				AVERAGE 3.4		COUNT: 22	MIN: 1	MAX:8 SU	IM: 69	# M			+ 1005

Figure 10-23. Insert Scatter (X, Y) or Bubble Chart option

2. Select the Scatter option, as shown in Figure 10-24.

Protectable Recommended Table Recommended	FLE H	HOME INSE	RT PA	÷ GE LAYOUT	FORMUL	AS DATA	REVIEW	VIEW	chap10 POWE	lexample1pivo R QUERY	ttable.xlsx - POWERPIV	Excel OT							? 🖭	- 8 Nei - 10	×
Chart 1 I I I State N O P Q R S T U 1 A B C D E F G H I <	PivotTable Re	commended T PivotTables Tables	able Pict	ures Online Pictures Illustrations		Store (My Apps - Apps	Recomm Ch	nended arts Chi	Scatte	k • PivotCha	rt Power View	Map • Power Map	Line Col	umn Win/ Loss	Slicer Tit	meline H	typerlink Links	4 π 6 Text • s	iuation + mbol ymbols		^
A B C D E F G H I Souther N O P Q R S T U 1 -	Chart 1	• : ×	V .	fx					10.	1800	hd										v
4 -	A A 1 2 3 .	8	c	DE	F	G	н	1	• Si • U • VI Bi of	catter se this chart ty Compare at le slues or pairs o Show relations I values	vpe to: ast two sets of data. ships betwee	of en sets	N	0	P	Q	R	S	T	U	1
8 6 2 9 5 2 10 4 3 11 3 3 12 2 4 13 1 4 14 7 1 15 1 4 16 - - 17 - - 18 - - 19 - - 10 - - 18 - - 19 - - 10 - - 10 - - 16 - - 17 - - 18 - - 19 - - 20 - - 21 - - 22 - - 23 - - 4 - - 21 - - 22 - - 23 - - 4 - - 4 - - 5 - - 22 - - 23 -	4 5 6 7	absences g 8 7	rade 1 1				-			se it when: The data repre leasurements.	esents separa	ite		L							
9 5 2 11 3 3 11 3 3 12 2 13 1 14 7 15 1 16 - 17 - 18 - 19 - 10 - 10 - 11 - 12 2 14 7 13 - 14 7 15 - 16 - 17 - 18 - 19 - 10 - 10 - 11 - 12 2 20 - 21 - 22 - 23 - 14 - 15 - 16 - 17 - 18 - 19 - 20 - 21 - 22 - 23 - 24 - 25 - 26	8	6	2				4.5		28 M	ore scatter Cr	harts										
11 3 3 12 2 4 13 1 14 7 15 1 16 1 17 1 18 - 19 - 10 - 10 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 19 - 10 - 10 - 11 - 12 3 14 - 15 - 16 - 16 - 17 - 18 - 19 - 19 - 10 - 10 - 11 2 12 3 13 - 14 - 15 - 16 - 17 - 18 - <td>9</td> <td>4</td> <td>2</td> <td></td> <td></td> <td></td> <td>4</td> <td>1</td> <td>•</td> <td></td> <td></td> <td></td> <td></td> <td>_</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	9	4	2				4	1	•					_							
12 2 4 13 1 4 14 7 15 1 16 1 17 1 18 1 19 0 20 0 21 0 22 3 4 5 6 7 8 1 19 1 20 1 21 1 22 1 23 1 24 1 25 1 26 1 27 1 28 1 29 1 20 1 21 1 22 1 23 1 24 1 25 1 26 1 27 1 28 1 29 1 29 1 20 1 21 1 22 1 23 1 24 1 25 1 26 1 27 1	11	3	3																		
13 1 4 14 7 1 15 1 16 1 17 1 18 0 19 0 21 0 22 3 23 0 24 0 25 0 21 0 22 0 23 0 24 0 25 0 26 0 27 0 28 0 29 0 20 0 21 0 22 0 23 0 24 0 25 0 26 0 27 0 28 0 29 0 20 0 21 0 22 0 23 0 24 0 25 0 26 0 27 0 28 0 29 0 29 0 20 0 20 0 <td>12</td> <td>2</td> <td>4</td> <td></td>	12	2	4																		
14 7 1 15 1 16 1 16 1 17 1 18 0 19 0 19 0 19 0 19 0 19 0 19 0 10 0 10 0 10 0 10 0 10 0 10 0 10 0 10 0 10 0 10 0 10 0 10 0 11 2 12 1 12 1 13 1 14 1 15 1 10 1 10 1 10 1 10 1 10 1 11 1 12 1 13 1 14 1 15 1 16 1 17 1 18 1 19 1 <tr< td=""><td>13</td><td>1</td><td>4</td><td></td><td></td><td></td><td>1.</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr<>	13	1	4				1.							1							
15 1 4 16 17 18 19 20 20 21 22 23 4 5 Sheet1 Sheet2 Sheet3 Sheet4 Sheet5 Sheet5 \odot \odot i	14	7	1																		
10 0 0 1 2 3 4 5 6 7 8 9 10 0 1 2 3 4 5 6 7 8 9 10 0 1 2 3 4 5 6 7 8 9 11 10 1 2 3 4 5 6 7 8 9 12 1 1 1 1 1 1 1 1 1 12 1 1 1 1 1 1 1 1 12 1 1 1 1 1 1 1 1	15	1	4				1.5														
0 0 <td>10</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td></td>	10						-														
0 1 2 3 4 5 6 7 8 9 20 2 3 4 5 6 7 8 9 21 2 3 4 5 6 7 8 9 22 3 4 5 6 7 8 9 23 5 6 7 8 9 24 6 6 7 8 9 25 6 7 8 9 26 7 8 9 27 7 8 9 27 7 8 9 28 7 8 9 29 7 8 9 29 7 8 9 20 7 8	18						0.5														
20 22 23 24 25 26 26 25 26 26 26 26 26 26 26 26 26 26 26 26 26	19						0	1	2	3 4	5	6	7 8	9							
21 22 23 24 25 26 26 26 26 26 26 26 26 26 26 26 26 26	20																				
22 23 ↔ Sheet1 Sheet2 Sheet3 Sheet4 Sheet5 Sheet6 ⊕ : €	21																				
23 → Sheet1 Sheet2 Sheet3 Sheet5 Sheet5 ④ : €	22																				
Sheet1 Sheet2 Sheet3 Sheet4 Sheet5 Sheet6 (23																				
	4 3	Sheet1	Sheet2	Sheet3	Sheet4	Sheet5	Sheet6	(-			-			1						1
	Colord declars				-			0				STRACE 2 4	COUNT 2		1111 0 01	14.60	000 (001	II)			

Figure 10-24. Select Scatter option

3. Click the + sign at the upper right of the chart and select Axis Titles, as shown in Figure 10-25.



Figure 10-25. Select Axis Titles

4. Click the Axis Title on the horizontal or x axis and change it to Absences. Change the label on the vertical, or y, axis to Grades and change the chart title to "Relationship between Grade and Absences," as shown in Figure 10-26.

	• 0	• 🗉 🕅	m••			chap1	0example1pi	vottable.xls	- Excel			CH	ART TOOLS						? 🖽	- 6	×
FILE	HOME	INSERT	PAGE LAYO	DUT F	ORMULAS	DATA	REVIEW	VIEW	POWER QU	IERV	POWERPIVOT	DESIGN	N FORM	TAN						Neil -	7
Add Chart Q	luick	Change					····.			Î		···	• • •	Switch Row	V Select	Change Chart Turne	Move				
Chart Layou	uts	Colors					Chart S	yles						Dat	ta	Туре	Location				^
Chart 1	*	X	fr																		~
			Ja	E	E	6	1	1	1 X 1	K			N	0		0	P	e			-
1		-	U	E		0	п	-	,	ĸ		PM1	14	0		4	N	3	-	0	۱Ħ
2																					
3																					
4		-					0				-lal		11		200						
5	abse	nces grade						Re	lationship	betw	een Grade a	and Abs	ences		+						
6		8 1					45														
0	-	6 3					4														
0		5 2					35								7						
10		4 3																			
11		3 3					19.95			1	T.				-						
12		2 4					ge .								5						
13		1 4					- G 2					•									
14		7 1					15														
15		1 4					1						•	•							
16	1						0.5								_						
17							0						7								
18											Absenses										
19											Passenses				-						
20																					
21																					
22																					
						-															Ŧ
<).	SI	neet1 Sh	eet2 She	et3 Sł	neet4 Sh	neet5 SI	neet6	۲					4							•	
READY									a		AVER	AGE 3.45	COUNT: 22	MIN: 1	MAX: 8 S	SUM: 69			-	-+ 10	0%



5. Right-click the plot area and select Add Trendline..., as shown in Figure 10-27.



Figure 10-27. Select Add Trendline...

6. In the Format Trendline pane on the right, scroll down and select Display Equation on Chart and Display R squared value on chart, as shown in Figure 10-28. The formula for the line, including the slope and y intercept, is shown. Note that the R-squared value of 0.9603 (shown in decimal form) indicates a high correlation. This is an inverse relationship, which means the more absences, the lower the grade.

	5	• 🔿		• • •			chi	p10example1p	ivottable.d	sx - Excel			C	HART TOOLS		1		7 10 - 8 :
FILE	н	OME	INSERT	PAGE LA	YOUT I	ORMULAS	DATA	REVIEW	VIEW	POWER	R QUERY	POWERPIVO	T DESIG	N FORM	AT			Neil - C
Paste	Cu Co For Clipboa	t py * mat Pain rd	ter B	I <u>U</u> - Fo	- - B - <u>A</u>	A* A*	= = = = =	◎・ 伝 伝 目 Alignmen	Wrap Test Merge &	t Center - ra	General \$ - % Nu	• *.8	Condition	al Format as 1 - Table - Styles	Cell Styles -	Insert Delete Cells	Format	∑ AutoSum - A v ⇒ Fill - Sort & Find & Clear - Filter - Select - Editing
Chart	t 1	7 1	X	/ fx														
1 2 3	A	В	c	D	E	F	G	н	1	J	K	L	М	N	0	P	(*	Format Trendline
5 6		absen	es grade	1					Re	ations	hip betv	veen Grad	e and Abs	ences		+		
7 8			6	2				45	-	-								O Polynomial Order 2
10 11			4 3	3				3 \$ 25				Concession of the local division of the loca						Moving Period 2
12			2					2 2 2 15						-				Trendline Name
14 15 16			1	1				1					γ = -0.48 R ²	801x 4.6126 - 0.9603	ő			Automatic Linear (grade) Qustom
17 18								0	0 1	2	3	4 ! Absenses	5 6	7	8	9		Eorward 0.0 period
19 20 21								_				0						Set Intercept 0.0
22 23																		Display <u>E</u> quation on chart Display <u>R</u> -squared value on chart
		She	et1 SI	neet2 Sh	eet3 S	heet4	Sheet5	Sheet6	۲		: •					1	Þ	
READY	9 1																	■ ■ ■+ 100

Figure 10-28. Chart showing the formula and R-squared value

Summary

This chapter has only begun to scratch the surface of the ability of Excel to calculate statistical values by using the Excel Analysis ToolPak. It has shown how to create visual displays of data, using the built-in charting tools. It has shown how to determine whether there is a relationship between two values. An entire book could be written on these topics.

CHAPTER 11

HDInsight

Chapter 1 described Hadoop as an open source distributed file system that allows processing of vast amounts of data. Microsoft's implementation of Hadoop on its Azure cloud platform is called HDInsight, which is designed to handle large amounts of data in the cloud. HDInsight uses Azure blob storage for storing the data, to make Hadoop available as a service in the cloud. Azure is useful for storing large datasets in the cloud in a cost-effective manner. Microsoft only charges for resources actually used.

Apache Hive supports analysis of large datasets stored in Hadoop. Hive is a data warehouse infrastructure built on top of Hadoop for providing data summarization, query, and analysis. Originally developed by Facebook, it is now an open source Apache project.

This chapter explores the easiest way to export data for analysis from Azure HDInsight into Excel.

Getting a Free Azure Account

Earlier chapters covered how to use the Azure Data Marketplace, which is essentially free for certain databases. The example in this chapter requires an Azure account. A 30-day trial account can be obtained by going to https://azure.microsoft.com/en-us/pricing/free-trial/.

The screen shown in Figure 11-1 will appear.



Figure 11-1. Azure free trial screen as of July 2015

Click Try It now, and you will be prompted to enter your information to set up a free one-month trial.

Importing Hadoop Files into Power Query

This example shows how to run a Hive query on a large dataset, using an HDInsight on Windows Azure Blob Storage. It then demonstrates how to use Power Query to import the data into Excel, so that it can be analyzed using Pivot Tables and Power Map.

Creating an Azure Storage Account

To do this tutorial, you need the Windows Azure account just created.

1. Sign in to the Azure portal

https://manage.windowsazure.com

The screen shown in Figure 11-2 will appear. Log in with your Azure credentials.



Figure 11-2. Microsoft Azure portal

2. After logging in, click +NEW at the lower left-hand corner of the Azure portal screen, as shown in Figure 11-3.

Ele Edit View History Bookmarks Tools Help http://api.nobel1/laureate.json × http://query	.y22&format≈json × New Tab	× Microsoft Azure	×	🧑 Sentiment A	nalysis In Excel × 🛛 🖬 Lo	oad Data from th	e Azure		0	×
A https://manage.windowsazure.com/@profes	sor94704hotmail.onmicrosoft.com#Workspaces/All/da	shboard	Ø ∀ C	Q, Search		☆自	1	ŵ	ø	=
Microsoft Azure 🛛 🗸		CREDIT STATUS			professor_	94704@hotmail.	com 🔎			
ALL ITEMS	all items									
	NAME		TYPE	STATUS	SUBSCRIPTION	LOCATION	Q			
	sfcampmay19	÷	Cloud service	🗸 Running	Free Trial	West US				
	portalvhdsbfq9h3109lvs Default Directory		Storage Account Directory	 ✓ Online ✓ Active 	Free Trial Shared by all Default Di	West US United States				
	sfcampmay19		Virtual machine	V Running	Free Trial	West US				
DB SQL DATABASES										
+ NEW		DELETE				1	0			

Figure 11-3. Azure portal screen

3. Select Data Services ➤ Storage and Quick Create, as shown in Figure 11-4. Enter a unique name for the URL. This example will use testhive. Select a location closest to where you are.

Eile Edit View Higtory Bookmarks Iools Help					- 0 ×
http://api.nobel1/laureate.json × http://query	y.y228:format=json × New Tab	× Microsoft Azure	× 🏩 Sentiment Analysis In Excel 🗴	Load Data from the Azure	× +
A https://manage.windowsazure.com/@profest	sor94704hotmail.onmicrosoft.com#Workspaces/All/	/dashboard	⊙ ∀ C ^a Q, Search	☆ 自 ♥ ↓	ê ⊕ ≡
Microsoft Azure 🛛 🗸		CREDIT STATUS	De pro	ofessor_94704@hotmail.com	· ^
ALL ITEMS	all items				
NEW				۲	
ИЗ СОМРИТЕ	DB SQL DATABASE	📕 QUICK CREATE	URL	0	
Data services			.*.	ore.windows.net	
			West US	~	
			REPLICATION		
			Geo-Redundant	~	
	STREAM ANALYTICS				
	STORSIMPLE MANAGER				
			CREATE STORAGE A	CCOUNT 🕑	
		in the second second			×

Figure 11-4. Setting up a storage account in Azure

4. Click Create Storage Account in the lower right and wait for the account to be created. The resulting screen will appear, as shown in Figure 11-5.

Eile Edit View Higtory B	Bookmarks <u>T</u> ools <u>H</u> elp						- 1	5
http://api.nobel1/laurea	ste.json × http://query.y228	format=json × New Tab	× Storage - Mie	rosoft Azure × 😒 Ser	itiment Analysis In Excel ×	Load Data from the A	kzure × +	
(A https://manage.wi	ndowsazure.com/@professor9470	4hotmail.onmicrosoft.com#Workspace	s/StorageExtension/storage	⊘ ∀ C Q Sear	ch	☆ 自 ♥	+ + +	Ð
	Microsoft Azure 🛛 🧹		CREDIT STATUS		professor_947	704@hotmail.com 🞴		
	ALL ITEMS	storage				_		
	WIE APPS	NAME	STATUS	LOCATION	SUBSCRIPTION	٩		
	VIRTUAL MACHINES	portalvhdsbfg9h3189lvs	→ ✓ Online	West US	Free Trial			
		testhive	V Online	West US	Free Trial			
	B SQL DATABASES							
	STORAGE 2							
	D MEDRA SERVICES							
	Слоне							
	+ NEW			Ú DELETE	1	1A 12 0		

Figure 11-5. New storage account

5. Highlight the new account and click Manage Access Keys at the bottom of the screen. The Manage Access Keys pop-up will appear, as shown in Figure 11-6. Notice the primary and secondary access keys. Clicking the icon to the right of the key will copy it to the clipboard. Click the check mark at the bottom right of the screen, to close the window.

CHAPTER 11 HDINSIGHT



Figure 11-6. Manage Access Keys

Provisioning a Hadoop Cluster

To provision a Hadoop Cluster, follow these steps:

- 1. Click +NEW in the lower left corner of the screen.
- Select Data Services ➤ HDInsight ➤ Hadoop and fill in the values on the right side of the screen, as shown in Figure 11-7. Choose a cluster name. Two nodes should be sufficient for this example. Azure sets the default user name to Admin. You will need to create a password of at least ten characters, including at least one capital letter, one number, and one special character, and enter it twice, as prompted. Be sure to remember the password entered.

CHAPTER 11 ■ HDINSIGHT

The second							
Microsoft Azure 🗸 🗸		CREDIT STATUS		professor_947040	Dyahoo.com 🎴		
ALL ITEMS	storage				- 1		
WEB APPS	NAME	STATUS	LOCATION	SUBSCRIPTION	م		
	sesthive1	V Online	Wett US	Free Trial			
NEW					×		
СОМРИТЕ	SQL DATABASE	📕 HADOOP	CLUSTER NAME				
DATA SERVICES		🖡 HBASE	myhadoopcluste	1 *.azurehdinsight.r	2 1et		
app services	HOINSIGHT	5TORM	CLUSTER SIZE 2 data nodes				
			WITD LISSE NAME	-			
MARKETPLACE			CONFIRM PASSWORD				
	STREAM ANALYTICS	CUSTOM CREATE					
			STORAGE ACCOUNT				
			CONTRACT.				

Figure 11-7. Setting up the HDInsight cluster

3. Click Create HDInsight Cluster at the bottom lower right of the screen. Then the screen shown in Figure 11-8 will appear after the cluster is created, which may take 20 minutes or more.

Eile Edit View Higtory E	Bookmarks Iools Help							-	0	×
(Maintena	ige.windowsazure.com/@professo	r94704yahoo.onmicrosoft.com	#Workspaces/StorageExtension/stora	ige	rch	☆自♡	+	ŵ	ø	=
	Microsoft Azure 🛛 🧹		CREDIT STAT	us and	professor_94704	Byahoo.com 🞴				1
	ALL ITEMS	storage								
		NAME	STATUS	LOCATION	SUBSCRIPTION	Q				
		testhivel	→ ✓ Online	West US	Free Trial					
	STORAGE									
						_				
	+ NEW		P () Manage access manage dome	Ö In delete	1 📥 1 🦺	2 E 				

Figure 11-8. Showing new cluster

4. Click HDInsight from the left pane to see the list of clusters, as shown in Figure 11-9. The cluster that was just created should be shown.

Ele Edit Yiew Higtory Bookmarks Iools Help											-	8	×
A https://manage.windowsazure.com/@professo	or94704yahoo.onmicrosof	ft.com#Workspaces/H	DinsightExtension/list	0	C Q Search			合自	Ø	+	ŵ	ø	=
Microsoft Azure 🛛 🤟	1		CREDIT STAT	JS		professe	or_94704@yat	100.com					
ALL ITEMS	hdinsight												
	NAME	STATUS	CLUSTER TYPE	SUBSCRIPTION NAME	LOCATION	OPERATING SYSTEM	VERSION	Q					
	myhadoopckister1	 Cluster Storage Pr 	Hadoop	Free Trial	West US	Windows Server 2012	31						
STORAGE													
HDINSIGHT													
+ NEW			D Delete			1	1 1 2	≡ 0	ľ				

Figure 11-9. List of HDInsight clusters

5. Select the cluster that was just created and click Query Console at the bottom of the screen. If requested, enter your admin username and password. The screen shown in Figure 11-10 will appear.



Figure 11-10. HDInsight Query Console

CHAPTER 11 HDINSIGHT

- 6. Scroll down to see Solutions With Sample Data.
- 7. Click the Hive Editor tab at the top of the screen. In Hive Editor, enter a query name and accept the default SQL as Select * from hivesampletable to select all records as shown in Figure 11-11.

Dinsight - Microsoft Azure	× Hive Editor ×	+											
https://mytesthive.azureh	dinsight.net/Home/HiveEditor			ଡ ଟ ୯	Q, Search			☆	١.	۵	+	â (Ð
	Use this editor and so	ubmit a hive query	y to your cluster										
	Query Name												
	HTC20												
	1 Select * from hiv	vesampletable											
	1 Select - Troil Hiv	resomptecable											
I	Job Session					Su	bmit						
I	Job Session Query Name	Date	v ki 0	Action	\$	Status	ibmit ¢						
	Job Session Query Name HTC20	Date 6%2015, 112158 AM	v Id ≎ job_1433718445158_0001	Action	¢	Status Completed	bmit ∳						
	Job Session Query Name HTC20	Date 6/8/2015, 11.21.58 AM	v Id ≎ job_1433718445158_0001	Action	¢	Status Completed	<mark>bmit</mark> ∳						

Figure 11-11. SQL query to select all records

- 8. Click Submit.
- **9.** To see the query results, click the query name at the bottom left of the screen, as shown in Figure 11-12.

CHAPTER 11 HDINSIGHT



Figure 11-12. Query results

10. On the HD Query Console screen, click File Browser at the top of the screen. The results are shown in Figure 11-13.

Eile Edit View History Boo	okmarks Iools Help kzure × Hive Editor × / ht	ps://mytesthme/FileBrowser	< +				-	a 🚺	
A https://mytesthive.az	zurehdinsight.net/Home/FileBrowser		⊚ ₹	C] Q Search	公白で	•	ŵ	e =	
		HDInsight Qu	ery Console						
	Getting Started Gallery Hi	ve Editor Job History			Help + Feedback				
	*								
	Show 10 ¥			Search					
	Name portalvhdsbfg9h3189lvs	Type <account></account>	Last Modified	I \$ Size (bytes)	φ				
	Showing 1 to 1 of 1 entries				⊕ 1 ⊖				
		1	Pr	wacy Help	Microsoft @ 2014 Microsoft				

Figure 11-13. List of clusters

11. Click the first item in the Name column on the left and then on your cluster name, to get the results shown in Figure 11-14.

Eile Edit View Higtory Bookmarks Ioo HDInsight - Microsoft Azure × Hi	ls <u>H</u> elp ve Editor ×	https://myhado.	e/FileBrowser ×	Twitter Trend Analysis	×	m 5 Tips T	o Manage Your Wi.	. × +				-	5 ×
A https://myhadoopcluster1.azurehdi	nsight.net/Home/FileBrowser				Ø ∀ C'	Q, Search			☆白	۵	÷	ŵ	⊜ ≡
	Microsoft Azu		sight Que	erv Console									
			ngin au	bry Console									
	Getting Started Gallery	Hive Editor	Job History					Help + Fe	edback				
	* testhive1 \ myhad	oopcluster1			Sea	arch							
	Name		Туре	1 Last Mo	dified		Size (bytes)						
	app-logs		<directory></directory>	7/23/2015, 7	04 20 AM								
	apps		<directory></directory>	7/23/2015, 6	:35:04 AM								
	example		<directory></directory>	7/23/2015, 6	40:15 AM								
	HdiSamples		<directory></directory>	7/23/2015, 6	:40:28 AM								
	hive		<directory></directory>	7/23/2015, 6	40:07 AM								
	mapred		<directory></directory>	7/23/2015, 6	:35:18 AM								
	Portal-Queries		<directory></directory>	4/3/1975, 5	00:00 PM								
	templeton-hadoop		<directory></directory>	7/23/2015, 7	:04:19 AM								
	user		<directory></directory>	7/23/2015, 6	:36:01 AM								
	yam		<directory></directory>	7/23/2015, 6	:35:21 AM								
	Showing 1 to 10 of 10 en	tries					(⊖ 1)	50.				
https://myhadoopcluster1.azurehdinsight.net	/Home/FileBrowser#	v						Microsoft					

Figure 11-14. List of files in cluster

Importing into Excel

There are several ways to import this data into Excel. This section will demonstrate how to download the file to a text file with fixed-length fields.

Follow these steps:

- 1. Go back to the screen shown in Figure 11-12, by clicking Job History at the top of the screen and then clicking the query name. Then click Download File.
- 2. The Job Log for the download is shown in Figure 11-15, which reports that 59,793 rows were downloaded. The number you see may differ.

CHAPTER 11 HDINSIGHT



Figure 11-15. Job Log for downloads

3. Click the blue box at the bottom of the screen labeled Download File. When prompted for the program to use to open the file, select Notepad. The result will appear in Notepad, as shown in Figure 11-16. This appears to be a log file for web site accesses.

					job_14	376585226	39_000	4_stdou	it - No	tepad				 ×
<u>F</u> ile	<u>E</u> dit F <u>o</u> rmat	<u>V</u> iew <u>H</u> elp												
8	18:54:20	en-US	Android Samsung	g SCH-150	e Cal	lifornia	United	States	13.92	84887	0 0			^
23	19:19:44	en-US	Android HTC	Incredi	ble Per	nnsylvania	United	States	NULL	0	0			
23	19:19:46	en-US	Android HTC	Incredi	ble Per	nnsylvania	United	States	1.475	7422	0 1			
23	19:19:47	en-US	Android HTC	Incredi	ble Per	nnsylvania	United	States	0.245	968	0 2			
28	01:37:50	en-US	Android Motoro	la	Droid X Co	lorado	United	States	20.30	95339	1 1			
28	00:53:31	en-US	Android Motoro	la	Droid X Co	lorado	United	States	16.29	81668	0 0			
28	00:53:50	en-US	Android Motoro	la	Droid X Co	lorado	United	States	1.771	5228	0 1			
28	16:44:21	en-US	Android Motoro	la	Droid X Uta	ah United	States	11.675	5987	2	1			
28	16:43:41	en-US	Android Motoro	la	Droid X Uta	ah United	States	36.944	6892	2	0			
28	01:37:19	en-US	Android Motoro	la	Droid X Co	lorado	United	States	28.98	11416	1 0			
30	17:19:36	en-US	RIM OS RIM	9650	Massachuse	tts United	States	3468.5	38966	0	2			
30	17:17:18	en-US	RIM OS RIM	9650	Massachuse	tts United	States	66.853	3378	0	1			
30	17:16:40	en-US	RIM OS RIM	9650	Massachuse	tts United	States	NULL	0	0				
43	00:44:46	en-US	RIM OS RIM	9330	Massachuse	tts United	States	2.3198	876	0	1			
43	00:44:41	en-US	RIM OS RIM	9330	Massachuse	tts United	States	NULL	0	0				
45	21:24:03	en-US	Android Samsun	g SCH-150	e Cal	lifornia	United	States	1.754	7729	1 1			
45	21:09:43	en-US	Android Samsun	g SCH-150	0 II	linois	United	States	857.1	453275	1 0			
45	20:01:50	en-US	Android Samsung	; SCH-150	Ø Nei	w Jersey	United	States	12.41	95326	e e			
49	03:05:50	en-US	Android LG	V\$740	New York	United	States	NULL	0	0				
59	01:23:42	en-US	Android LG	V\$660	Nevada Un:	ited States	0.4996	229	0	0				
59	01:23:45	en-US	Android LG	V\$660	Nevada Un:	ited States	1.1773	128	0	2				
59	01:23:42	en-US	Android LG	VS660	Nevada Un:	ited States	7.7791	862	0	1				
62	03:07:36	en-US	Android LG	VS910	California	United	States	39.499	1038	0	0			
62	03:08:15	en-US	Android LG	VS910	California	United	States	2.3677	288	0	1			
62	03:08:17	en-US	Android LG	VS910	California	United	States	2.5171	87	0	2			
62	03:08:20	en-US	Android LG	VS910	California	United	States	3.7881	73	0	3			
67	03:59:32	en-US	iPhone OS	Apple	iPod Touch	3.1.3	Pennsy	lvania	Unite	d States	19.850022	0	0	
77	14:51:12	en-US	Android LG	VS660	Illinois	United	States	7.1416	43	0	4			
77	14:48:39	en-US	Android LG	VS660	Illinois	United	States	41.237	6733	0	0			
77	14:48:42	en-US	Android LG	VS660	Illinois	United	States	1.4319	403	0	1			
77	14:49:25	en-US	Android LG	VS660	Illinois	United	States	5.8357	112	0	2			
77	14:50:49	en-US	Android LG	VS660	Illinois	United	States	18.002	8619	0	3			
89	22:54:19	en-US	Android Samsung	g SCH-150	0 Te:	xas United	States	77.033	459	0	0			
89	22:54:21	en-US	Android Samsung	g SCH-150	0 Te:	xas United	States	0.4882	386	0	1			
89	22:55:38	en-US	Android Samsung	g SCH-150	0 Te:	xas United	States	46.268	7482	0	2			
89	22:56:27	en-US	Android Samsung	g SCH-150	0 Te:	xas United	States	10.417	6749	0	3			
93	15:26:52	en-US	Android LG	VS910	California	United	States	9.3417	195	1	0			
93	15:26:55	en-US	Android LG	VS910	California	United	States	0.2404	368	1	1			
93	00:08:02	en-US	Android LG	VS910	California	United	States	11.020	7005	0	0			
93	00:08:04	en-US	Android LG	VS910	California	United	States	1.2142	71	0	1			
93	00:12:20	en-US	Android LG	VS910	California	United	States	1.3144	414	0	2			
93	00:12:21	en-US	Android LG	VS910	California	United	States	3.6776	226	0	3			
93	00:15:36	en-US	Android LG	VS910	California	United	States	20.563	3449	0	4			~
<														> .≓

Figure 11-16. Log file for web site accesses

- 4. Do a Save As to save the file to the desktop, to make it easier to access.
- **5.** Open Excel and click the Power Query tab. Select from File and then From Text and Browse to find the downloaded file.
- **6.** The file will be loaded into the Query Editor, as shown in Figure 11-17. Note that Power Query parsed the fixed-length fields to neatly put all the data in separate columns, but there are no descriptive column headings.

XII C	r =				job	_1437658522639_0	0004_stdout - C	Query Editor			- ×
F Close & Load • Close	Home Y01 Refresh Preview	Transform [Y02] Properties Advanced E Query	Add Coll Y03 ditor Co M	International States of the second se	Keep Remov Rows * Rows *	Remove Duplicate:	s 21 21 Split Column	Group By J2 Replace Val	ole Number • ow As Headers • lues	Combine	^
×	V 1	x = Table.T	ransform	olumnTypes(Se	ource.{{"Column1	", Int64.Type}, {"Colum	nn2", type time),	("Column3", type text).	{"Column4",	× 0	
1 Cole	umn1 💌 8 23	Column2 6:54:20 PN 7:19:44 PN	Column3 f en-US f en-US	Column4 Android Android	Column5 Samsung HTC	Column6 SCH-i500 Incredible	Column7 California Pennsylvania	Column8 Co United States 13 United States NU	olumn9 💌 Co 3.9204007 ULL	Query Settings PROPERTIES Name job_1437658522639_0004_stdout	×
4 5	23 23 28	7:19:47 PN 1:37:50 AN	f en-US f en-US	Android Android Android	HTC Motorola	Incredible Drold X Double	Pennsylvania Colorado	United States 0.1 United States 0.1 United States 20	245968 0.3095339	All Properties APPLIED STEPS	
7 8	28	12:53:50 AN 4:44:21 PN	f en-US f en-US	Android Android	Motorola Motorola	Droid X Droid X Droid X	Colorado Utah	United States 1. United States 1. United States 11	7715228 1.6755987	Source X Changed Type	*
9 10 11	28 28 30	4:43:41 PN 1:37:19 AN 5:19:36 PN	f en-US f en-US f en-US	Android Android RIM OS	Motorola Motorola RIM	Droid X Droid X 9650	Utah Colorado Massachusetts	United States 36 United States 28 United States 34	9446892 9811416 968.538966		
12 13 14	30 30 43	5:17:18 PN 5:16:40 PN 12:44:46 AN	1 en-US 1 en-US 1 en-US	RIM OS RIM OS RIM OS	RIM RIM RIM	9650 9650 9330	Massachusetts Massachusetts Massachusetts	United States 66 United States NU United States 2.3	3.8533378 ULL 3198876		
15 16 17	43 45 45	12:44:41 AN 9:24:03 PN 9:09:43 PN	f en-US f en-US f en-US	RIM OS Android Android	RIM Samsung Samsung	9330 SCH-I500 SCH-I500	Massachusetts California Illinois	United States NL United States 1.7 United States 85	ULL 7547729 57.1453275		
18 19	45 49	8:01:50 PN 3:05:50 AN	f en-US f en-US	Android Android	Samsung LG	SCH-I500 V5740	New Jersey New York	United States 12 United States NU	2.4195326 ULL		
20 21 22	59 59 59	1:23:42 AN 1:23:45 AN 1:23:42 AN	/ en-US / en-US / en-US	Android Android Android	LG LG LG	V\$660 V\$660 V\$660	Nevada Nevada	United States 0.4 United States 1.1 United States 7.1	4996229 1773128 7791862	v	

Figure 11-17. File loaded into Power Query

7. To rename critical columns, right-click the column heading, select Rename, and enter a new descriptive name, as shown below.

Old Name	New Name
Column2	Datetime
Column4	OS
Column5	Manufacturer
Coluimn6	Model
Column7	State
Column 8	Country
Column10	Accesses

The result is shown in Figure 11-18.

d i 📇	=			job_1	437658522639_00	004_stdout - Que	ery Editor			- 🗆 🗙
E	Home V01	Transform Add	Column View Y03 Y04			A1 -			E.u.	^ (
Close & Load • Close	Refresh Preview •	Query	Choose Remove Columns Columns • Manage Columns	Keep Remove Rows * Rows * Redu	Remove Duplicates	ZI ZI Split Column	By Jack Print Jack Print	xt * Row As Headers * Values	Append Queries	
×	√ fx	= Table.Rename	eColumns(#"Change	ed Type",{{"Column	7", "state"}, {"Colum	n2", "datetime"}, {"	Column4", "os"},		Ouery Setting	< ×
M- Colu 40	mn1 💌 d 93	atetime Colum Col	nn3 💌 os Android	manufacturer	v model V\$910	✓ state California	country United States	Column9 [Co PROPERTIES	
41	93	12:12:20 AM en-US	Android	LG	V\$910	California	United States	1.3144414	∧ Name	
42	93	12:12:21 AM en-US	Android	LG	V\$910	California	United States	3.6776226	job_1437658522639_	0004_stdout
43	93	12:15:36 AM en-US	Android	LG	V5910	California	United States	20.5633449	All Properties	
44	109	6:55:15 PM en-US	RIM OS	RIM	9330	Massachusetts	United States	1.82519		
45	109	6:55:10 PM en-US	RIM OS	RIM	9330	Massachusetts	United States	728.9452825	APPLIED STEPS	
46	114	4:16:05 AM en-US	Android	Samsung	SCH-I500	Texas	United States	0.7507987	Source	*
47	114	4:16:08 AM en-US	Android	Samsung	SCH-i500	Texas	United States	1.368729	Changed Type	
48	114	4:16:07 AM en-US	Android	Samsung	SCH-i500	Texas	United States	NULL	× Renamed Column	15
49	114	4:16:07 AM en-US	Androld	Samsung	SCH-I500	Texas	United States	0.1978611		
50	141	6:33:24 PM en-US	RIM OS	RIM	9330	Massachusetts	United States	NULL		
51	146	9:30:28 AM en-US	iPhone OS	Apple	iPhone 4.3.x	New York	United States	NULL		
52	146	11:32:01 PM en-US	iPhone OS	Apple	iPhone 4.3.x	New York	United States	NULL		
53	156	7:31:38 PM en-US	iPhone OS	Apple	iPhone 4.3.x	Pennsylvania	United States	5.3385825		
54	156	7:51:59 PM en-US	iPhone OS	Apple	iPhone 4.3.x	Pennsylvania	United States	NULL		
55	156	7:31:31 PM en-US	Phone OS	Apple	iPhone 4.3.x	Pennsylvania	United States	2.4204735		
56	156	2:35:59 AM en-US	iPhone OS	Apple	iPhone 4.3.x	Pennsylvania	United States	17.116298		
57	156	7:31:56 PM en-US	iPhone OS	Apple	iPhone 4.3.x	Pennsylvania	United States	3.0774953		
58	156	6:52:14 PM en-US	iPhone OS	Apple	iPhone 4.3.x	Pennsylvania	United States	NULL		
59	181	8:44:06 PM en-US	Android	HTC	Thunderbolt	Illinois	United States	NULL		
60	182	6:49:22 PM en-US	Android	LG	V\$660	Missouri	United States	2.3695996		
61	182	6:49:22 PM en-US	Android	LG	V\$660	Missouri	United States	0.2170419	~	
67	185	6:31:39 PM en-US	Android	LG	V\$660	Colorado	United States	3.1242108		

Figure 11-18. Power Query Editor after columns are renamed

8. To load the data into a spreadsheet, click Close & Load on the left end of the ribbon and select Close & Load. The data is loaded into a spreadsheet, as shown in Figure 11-19.

CHAPTER 11 HDINSIGHT

FLE	HON	c ^a - ≩ × Me iniser	F PAGE L	AYOUT FORM	chapter11n MULAS DATA	napaug13.xkx - Excel REVIEW VIEW	POWER QUERY PO	WERPIVOT QUE	RY DESIGN		7 🖻 – 8 X Nel - 9
PivotT	able Recom Pivo Tab	nmended Tal	ble Picture	Online Pictures	🚔 Store 🚺	Recommended Charts	PivotChart	Power Map View * Reports Power Map	Line Column Win Los Sparklines	Slicer Timeline	Hyperfink Links Symbol
A1		1 X	$\checkmark f_x$	Column1							¥
	A	8	c	D	E	F	G	н	1	K -	
1 00	lumn1 -	datetime	 Column3 	V OS	manufacturer	r model	🕶 state 💌	country 💌	Column9 💌 access	es 👻 Column1	Workbook Queries * *
2	8	6:54:20 P	M en-US	Android	Samsung	SCH-1500	California	United States	13.9204007	0	
3	23	7:19:44 P	M en-US	Android	HTC	Incredible	Pennsylvania	United States	NULL	0	
4	23	7:19:46 P	M en-US	Android	HTC	Incredible	Pennsylvania	United States	1.4757422	0	job_1437658522639_0004_std2
5	23	7:19:47 P	M en-US	Android	HTC	Incredible	Pennsylvania	United States	0.245968	0	59,793 rows loaded.
6	28	1:37:50 A	M en-US	Android	Motorola	Droid X	Colorado	United States	20.3095339	1	
7	28	12:53:31 A	M en-US	Android	Motorola	Droid X	Colorado	United States	16.2981668	0	
8	28	12:53:50 A	M en-US	Android	Motorola	Droid X	Colorado	United States	1.7715228	0	
9	28	4:44:21 P	M en-US	Android	Motorola	Droid X	Utah	United States	11.6755987	2	
10	28	4:43:41 P	M en-US	Android	Motorola	Droid X	Utah	United States	36.9446892	2	
11	28	1:37:19 A	M en-US	Android	Motorola	Droid X	Colorado	United States	28.9811416	1	
12	30	5:19:36 P	M en-US	RIM OS	RIM	9650	Massachusetts	United States	3468.538966	0	
13	30	5:17:18 P	M en-US	RIM OS	RIM	9650	Massachusetts	United States	66.8533378	0	
14	30	5:16:40 P	M en-US	RIM OS	RIM	9650	Massachusetts	United States	NULL	0	
15	43	12:44:46 A	M en-US	RIM OS	RIM	9330	Massachusetts	United States	2.3198876	0	
16	43	12:44:41 A	M en-US	RIM OS	RIM	9330	Massachusetts	United States	NULL	0	
17	45	9:24:03 P	M en-US	Android	Samsung	SCH-1500	California	United States	1.7547729	1	
18	45	9:09:43 P	M en-US	Android	Samsung	SCH-1500	Illinois	United States	857.1453275	1	
19	45	8:01:50 P	M en-US	Android	Samsung	SCH-1500	New Jersey	United States	12.4195326	0	
20	49	3:05:50 A	M en-US	Android	LG	V\$740	New York	United States	NULL	0	
21	59	1:23:42 A	M en-US	Android	LG	V\$660	Nevada	United States	0.4996229	0	
22	59	1:23:45 A	M en-US	Android	LG	V\$660	Nevada	United States	1.1773128	0	
23	59	1:23:42 A	M en-US	Android	LG	V\$660	Nevada	United States	7.7791862	0	
74					10	10010	California		30 1001030	~ ~	
		Sheet1	Sheet2	•			÷ [4]			Þ	
READY	8						AVERAGE: 17384.623	2 COUNT: 657734	MIN: 0 MAX: 140091	SUM: 4157915099	H III+ 100%

Figure 11-19. Data loaded into spreadsheet

Creating a Pivot Table

To create a Pivot Table to analyze accesses by state, manufacturer, and model, follow these steps:

- 1. Click Insert and Pivot Table. Accept the default of putting it in a new worksheet.
- 2. Drag state to the Columns box, manufacturer and model to the Rows box, and accesses to the Values box. The result, which shows accesses by manufacturer and model, is shown in Figure 11-20.

ILE HOME INSERT	PAGE LAYOUT FORM	Boo IULAS DATA R	k2 - Excel EVIEW VIEW P	OWER QUERY	POWERPINOT	PIVOTTABLE TOO ANALYZE DES	LS RGN		0	Nel - D	
A Cut Copy • Paste Format Painter Clipboard r	ri • 11 • A* I U • ⊡ • 2• A Font	A [*] = = ■ ≫ · • = = = ≪ ·	Wrap Text	General r - \$ - %	• • • • • • • • • • • • • • • • • • •	ditional Format as atting * Table * Styles	Cell In: Styles •	sert Delete Format Cells	AutoSum * A Z Fill * Clear * Filter Editing	Find & Select *	
A3 * : 🗙 🔬	f_x Count of acce	sses									
A Count of accesses Colu	8 C	D	F	G	н	JK	L	M	PivotTable F	elds •>	
Row Labels v California Colorado District Of Columbia Georgia Illinois Massachusetts Missouri Nevada New Jersey New York Pennsylva # 1 1									✓ manufacturer ✓ model ✓ state		
e Apple iPhone 4.3.x iPod Touch 3.1.3			1					2	country Column9 decesses Column11		
Incredible Thunderbolt		10		1					MORE TABLES		
V\$660 V\$740	п	19		5	2	3		1	Trag fields between areas below:		
VS910 Motorola Droid X	11	4								state •	
9330 9650					11 8 3				ROWS manufacturer	∑ VALUES	
Samsung	2		4	1			1	12	Defer Lavout Und	UPDAT	

Figure 11-20. Pivot Table showing accesses by manufactuer and model

Creating a Map in Power Map

This example shows how to map the data using Power Map. Follow these steps:

- 1. With the data loaded into the spreadsheet, click the Insert tab and then Map and Launch Power Map
- 2. Accept the default of country and state as the default geographic fields and click Next, as shown in Figure 11-21.
CHAPTER 11 HDINSIGHT



Figure 11-21. Selecting geographic fields

3. Click Accesses for height and right-click OS and select Set as Category, as shown in Figure 11-22, which shows the map panned over to the United States.



Figure 11-22. Map showing accesses by OS in the United States

A similar display for Europe is shown in Figure 11-23.



Figure 11-23. Map showing access by OS in Europe

If you are finished with the HDInsight cluster you just created, delete it, so that you will not be charged for it.

Summary

This chapter has just scratched the surface of the power of creating and accessing Hadoop clusters in Azure and moving the data into Excel and Power Query for analysis using Pivot Tables and Power Map to provide a graphic display of the data broken out by OS.

Index

A

Analysis ToolPak. See Excel Analysis ToolPak Azure HDInsight, Excel import power query account creation, 226 create Pivot Table, 238 Hadoop Cluster, 229 import data, 234 using Power Map, 239 trial screen, 225 Azure Marketplace browse screen, 49 import data dialog, 53 Real GDP, 51 capita records, 52 import into excel, 53 by slicer, 54

B

Big data agriculture, 8 Apache Hadoop commercial implementations, 13 definition, 12 HDFS, 13 MapReduce, 12 Apache Spark, 14 characteristics, 12 Cholera map, 4 cost savings, 5-6 definition, 1, 11 Google Earth, 5 Google flu trends, 4 HDFS, 11 industrial equipment, 8 industrial internet, 6 life cycle, 12 medicalized smartphone, 7 Microsoft self-service BI, 14

middle managers, 2 natural resource, 1 personal computers, 8 predictive policing, 5 relational database dBASE programs, 11 normalization, 9 projection, 9 selection, 9 role of, 1 storage capacity, 9 time line chart, 2 time-series chart, 3 tracking malaria, 5

C

Charting data, 208 Control Program/Monitor (CP/M) computers, 15

D

Data Analysis Expressions (DAX) analyze sales data, 129 calculated columns, 127 calculated fields/measures, 127 CALCULATE function, 136 key functions, 128 KPI creation, 140 operators, 128 Pivot Table creation, 142 store sales, 138 SUMX function. 135 update formulas, 128 calculated fields, 130 profitability, 132 Database schema, 77 Data forms, 32 Data model calculated columns, 70 calculated fields, 72

INDEX

Data model (cont.) candidate key, 58 composite key, 58 database, 57 enablePowerPivot, 55 excel table creation, 58 field, 57 foreign key, 58 load data, 62 3NF, 58 normalization, 58 **Pivot** Table multiple tables, 67 two tables, 66 primary key, 58 record, 57 related table, 58 relational databases, 57 relationship, 58 source table, 58 table, 57 Data scrubbing, 25 Data validation, 31 Data wrangling, 25 Descriptive statistics bell curve, 206-207 calculation, 207 dialog, 211 measures of dispersion, 207 output, 212 statistical functions in Excel 2013, 208

■ E, F

Excel correct capitalization, 25 as database, 16 data forms, 32 data scrubbing, 25 data tab, import data, 23 data validation, 31 data wrangling, 25 delimited fields, 26 file extensions, 16 formats, 18 JSON format, 22 record selection, 34 remove duplicates, 30 spreadsheet to database, 15 statistical calculations analytical tools, 203-204 graphical tool, 208

statistical functions, 208 status bar menu customization, 205–206 string manipulation, 29 text file format, 18 XML format, 19 XML with attributes, 20 Excel Analysis ToolPak Anova tools, 214 correlation coefficient, 214 covariance, 214 enable, 208 exponential smoothing, 214 test score analysis, 210 Extensible Markup Language (XML), 19

G

Garbage In, Garbage Out (GIGO), 31

I H

Hadoop Cluster, 229 file list, 233–234 lists, 231 new cluster, 230 Query Console, 231 query results, 233 record selection, 232 setting up, 230 Hadoop Distributed File System (HDFS), 11, 13 Histogram Data Analysis ToolPak, 212 dialog, 213 Pivot Table, 214 plot, 213

Inferential statistics, 206

J

JavaScript Object Notation (JSON) format, 22

K, L

Key Performance Indicator (KPI), 140-143

M

Measures of dispersion, 207 MSQuery, 94–98

■ N, O

NewSQL database, 77 NoSQL database, 77 CAP Theorem, 13 characteristics, 13 definition, 13 implementations, 14

P

Pivot Chart, 35, 42 Pivot Table Azure Marketplace (see Azure Marketplace) breakdown creation, 40 changes, 39 creation, 37 definition, 36 in Excel 2013 analysis tool window, 35 preview, 36 Grand Totals, 43 Grouping window, 41 histogram, 214 questions, 37 sales by day of week analyzing, 43 creation. 45 slicers, 47 subtotals, 43 time line, 48 Power Map insert section, 174 installation, 173 laver section, 174 map section, 174 plotting, 173-174 time section, 174 tour section, 174 troubleshooting California crime statistics, 175, 180 2D chart, 184, 193 European unemployment rates, 185 plotting multiple statistics. 180 time animation exercise, 193 unemployment rates, 191 view section, 174 Power Query Data Catalog Search option, 149 importing CSV files from folder, 155 group by, 160

importing JSON data, 162 installation, 145 population data, 149-150 Query Editor, 146 ribbon, 146 S&P 500 stock index, 151 Power View, 99 access database, 111 diagram view, 112 import tables, 112 matrix view, 115 order totals. 114 relationships, 113 reports, 113 add map, 108 bar chart, 105 column chart, 106 considerations, 100 customer and city, 115 design surface, 99 fields, 100 multiple years, 107 orders by employee, 120 orders by product, 122 single spreadsheet fields pane, 103 filters pane, 101 GDP data, 101 summarized data, 100 using titles, 109 viewing data, 104 Proper() function, 25

🔳 Q, R

Query Editor group by, 160 Home ribbon, 146 population data, 150

S

Scatter chart grade-absence relationship, 220 R-squared value, 219, 224. SEQUEL. *See* Structured query language (SQL) Slicers, 47 Structured query language (SQL) aggregate functions, 79 description, 77 equijoin, 80 INDEX

Structured query language (SQL) (*cont.*) extract summary statistics, 89 history, 77 import external database, 80, 86 join condition, 86 MSQuery, 94 NewSQL, 77 NoSQL, 77 outer join, 80 SQL++, 78 subtotals, 79 syntax, 78 total order value report, by employee, 91

T, U, V, W, X, Y, Z

Two dimensional chart addition, 184 creation, 193