

LEARNING MADE EASY



4th Edition

Microsoft®
Excel® Data Analysis
for dummies®
A Wiley Brand



Navigate and analyze
complicated data

Work with external databases,
PivotTables, and PivotCharts

Make the most of the latest
features of Excel 2019

Paul McFedries

Excel expert and bestselling
author of more than 90 books



Excel[®] Data Analysis

4th Edition

by Paul McFedries

**for
dummies[®]**
A Wiley Brand

Excel® Data Analysis For Dummies®, 4th Edition

Published by: **John Wiley & Sons, Inc.**, 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Copyright © 2019 by John Wiley & Sons, Inc., Hoboken, New Jersey

Media and software compilation copyright © 2019 by John Wiley & Sons, Inc. All rights reserved.

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and may not be used without written permission. Microsoft and Excel are registered trademarks of Microsoft Corporation. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit <https://hub.wiley.com/community/support/dummies>.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2018959384

ISBN 978-1-119-51816-7 (pbk); ISBN 978-1-119-51822-8 (ebk); ISBN 978-1-119-51820-4 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents at a Glance

| | |
|--|-----|
| Introduction | 1 |
| Part 1: Getting Started with Data Analysis | 5 |
| CHAPTER 1: Learning Basic Data-Analysis Techniques | 7 |
| CHAPTER 2: Working with Data-Analysis Tools | 31 |
| CHAPTER 3: Introducing Excel Tables | 53 |
| CHAPTER 4: Grabbing Data from External Sources | 75 |
| CHAPTER 5: Scrub-a-Dub-Dub: Cleaning Data | 97 |
| CHAPTER 6: Analyzing Table Data with Functions | 119 |
| Part 2: Analyzing Data with PivotTables and PivotCharts | 133 |
| CHAPTER 7: Creating and Using PivotTables | 135 |
| CHAPTER 8: Performing PivotTable Calculations | 155 |
| CHAPTER 9: Building PivotCharts | 177 |
| Part 3: Discovering Advanced Data Analysis Tools | 197 |
| CHAPTER 10: Tracking Trends and Making Forecasts | 199 |
| CHAPTER 11: Analyzing Data with Statistics | 219 |
| CHAPTER 12: Analyzing Data with Descriptive Statistics | 237 |
| CHAPTER 13: Analyzing Data with Inferential Statistics | 253 |
| Part 4: The Part of Tens | 271 |
| CHAPTER 14: Ten Things You Ought to Know about Statistics | 273 |
| CHAPTER 15: Ten Ways to Analyze Financial Data | 285 |
| CHAPTER 16: Ten Ways to Raise Your PivotTable Game | 295 |
| Appendix: Glossary of Data Analysis and Excel Terms | 309 |
| Index | 319 |

Table of Contents

| | |
|---|---------------|
| INTRODUCTION | 1 |
| About This Book | 1 |
| What You Can Safely Ignore | 2 |
| Foolish Assumptions | 3 |
| Icons Used in This Book | 3 |
| Beyond the Book | 4 |
| Where to Go from Here | 4 |
| PART 1: GETTING STARTED WITH DATA ANALYSIS | 5 |
| CHAPTER 1: Learning Basic Data-Analysis Techniques | 7 |
| What Is Data Analysis, Anyway? | 8 |
| Cooking raw data | 8 |
| Dealing with data | 8 |
| Building data models | 9 |
| Performing what-if analysis | 9 |
| Analyzing Data with Conditional Formatting | 9 |
| Highlighting cells that meet some criteria | 10 |
| Showing pesky duplicate values | 12 |
| Highlighting the top or bottom values in a range | 13 |
| Analyzing cell values with data bars | 14 |
| Analyzing cell values with color scales | 16 |
| Analyzing cell values with icon sets | 16 |
| Creating a custom conditional formatting rule | 17 |
| Editing a conditional formatting rule | 20 |
| Removing conditional formatting rules | 22 |
| Summarizing Data with Subtotals | 23 |
| Grouping Related Data | 24 |
| Consolidating Data from Multiple Worksheets | 26 |
| Consolidating by position | 27 |
| Consolidating by category | 28 |
| CHAPTER 2: Working with Data-Analysis Tools | 31 |
| Working with Data Tables | 31 |
| Creating a basic data table | 32 |
| Creating a two-input data table | 34 |
| Skipping data tables when calculating workbooks | 36 |
| Analyzing Data with Goal Seek | 37 |
| Analyzing Data with Scenarios | 39 |
| Create a scenario | 39 |
| Apply a scenario | 41 |

| | |
|---|-----------|
| Edit a scenario | 41 |
| Delete a scenario | 42 |
| Optimizing Data with Solver | 42 |
| Understanding Solver | 43 |
| The advantages of Solver | 43 |
| When should you use Solver? | 43 |
| Loading the Solver add-in. | 45 |
| Optimizing a result with Solver | 46 |
| Adding constraints to Solver | 48 |
| Save a Solver solution as a scenario | 51 |
| CHAPTER 3: Introducing Excel Tables | 53 |
| What Is a Table and Why Should I Care? | 53 |
| Building a Table | 55 |
| Getting the data from an external source | 56 |
| Converting a range to a table | 56 |
| Basic table maintenance | 58 |
| Analyzing Table Information | 59 |
| Displaying simple statistics | 59 |
| Adding a column subtotal | 60 |
| Sorting table records. | 62 |
| Filtering table records. | 64 |
| Clearing a filter. | 66 |
| Turning off AutoFilter | 66 |
| Applying a predefined AutoFilter | 66 |
| Applying multiple filters | 69 |
| Applying advanced filters | 69 |
| CHAPTER 4: Grabbing Data from External Sources | 75 |
| What's All This About External Data? | 75 |
| Exporting Data from Other Programs | 77 |
| Importing External Data into Excel | 78 |
| Importing data from an Access table | 78 |
| Importing data from a Word table | 79 |
| Introducing text file importing. | 80 |
| Importing a delimited text file | 81 |
| Importing a fixed-width text file | 82 |
| Importing data from a web page | 83 |
| Importing an XML file | 85 |
| Querying External Databases | 87 |
| Defining a data source | 88 |
| Querying a data source | 91 |
| It's Sometimes a Raw Deal | 95 |

| | | |
|-------------------|---|-----|
| CHAPTER 5: | Scrub-a-Dub-Dub: Cleaning Data | 97 |
| | Editing Your Imported Workbook | 97 |
| | Deleting unnecessary columns | 98 |
| | Deleting unnecessary rows | 99 |
| | Resizing columns | 99 |
| | Resizing rows | 99 |
| | Erasing unneeded contents in a cell or range | 100 |
| | Formatting numeric values | 100 |
| | Copying worksheet data | 101 |
| | Moving worksheet data | 101 |
| | Replacing data in fields | 102 |
| | Cleaning Data with Text Functions | 103 |
| | The CLEAN function | 103 |
| | The CONCAT function | 104 |
| | The EXACT function | 105 |
| | The FIND function | 105 |
| | The LEFT function | 106 |
| | The LEN function | 107 |
| | The LOWER function | 107 |
| | The MID function | 107 |
| | The NUMBERTOVALUE function | 108 |
| | The PROPER function | 109 |
| | The REPLACE function | 109 |
| | The RIGHT function | 110 |
| | The SEARCH function | 111 |
| | The SUBSTITUTE function | 111 |
| | The TEXT function | 112 |
| | The TEXTJOIN function | 112 |
| | The TRIM function | 112 |
| | The UPPER function | 113 |
| | The VALUE function | 113 |
| | Converting text function formulas to text | 114 |
| | Using Validation to Keep Data Clean | 114 |
| CHAPTER 6: | Analyzing Table Data with Functions | 119 |
| | The Database Functions: Some General Remarks | 120 |
| | Retrieving a Value from a Table | 121 |
| | Summing a Column's Values | 122 |
| | Counting a Column's Values | 123 |
| | Averaging a Column's Values | 125 |
| | Determining a Column's Maximum and Minimum Values | 126 |
| | Multiplying a Column's Values | 127 |
| | Deriving a Column's Standard Deviation | 128 |
| | Calculating a Column's Variance | 130 |

| | |
|--|-----|
| PART 2: ANALYZING DATA WITH PIVOTTABLES AND PIVOTCHARTS | 133 |
| CHAPTER 7: Creating and Using PivotTables | 135 |
| Understanding PivotTables | 135 |
| Exploring PivotTable Features | 137 |
| Building a PivotTable from an Excel Table | 138 |
| Creating a PivotTable from External Data | 141 |
| Refreshing PivotTable Data | 143 |
| Refreshing PivotTable data manually | 144 |
| Refreshing PivotTable data automatically | 144 |
| Adding Multiple Fields to a PivotTable Area | 145 |
| Pivoting a Field to a Different Area | 146 |
| Grouping PivotTable Values | 147 |
| Grouping numeric values | 147 |
| Grouping date and time values | 148 |
| Grouping text values | 148 |
| Filtering PivotTable Values | 149 |
| Applying a report filter | 149 |
| Filtering row or column items | 151 |
| Filtering PivotTable values | 151 |
| Filtering a PivotTable with a slicer | 152 |
| CHAPTER 8: Performing PivotTable Calculations | 155 |
| Messing around with PivotTable Summary Calculations | 156 |
| Changing the PivotTable summary calculation | 156 |
| Trying out the difference summary calculation | 158 |
| Applying a percentage summary calculation | 160 |
| Adding a running total summary calculation | 162 |
| Creating an index summary calculation | 164 |
| Working with PivotTable Subtotals | 166 |
| Turning off subtotals for a field | 166 |
| Displaying multiple subtotals for a field | 167 |
| Introducing Custom Calculations | 168 |
| Formulas for custom calculations | 168 |
| Checking out the custom calculation types | 169 |
| Understanding custom calculation limitations | 169 |
| Inserting a Custom Calculated Field | 171 |
| Inserting a Custom Calculated Item | 173 |
| Editing a Custom Calculation | 175 |
| Deleting a Custom Calculation | 176 |

| | |
|--|-----|
| CHAPTER 9: Building PivotCharts | 177 |
| Introducing the PivotChart | 177 |
| Understanding PivotChart pros and cons | 178 |
| Taking a PivotChart tour | 179 |
| Understanding PivotChart limitations | 180 |
| Creating a PivotChart | 181 |
| Creating a PivotChart from a PivotTable | 181 |
| Embedding a PivotChart on a PivotTable's worksheet | 181 |
| Creating a PivotChart from an Excel table | 183 |
| Working with PivotCharts | 185 |
| Moving a PivotChart to another sheet | 185 |
| Filtering a PivotChart | 186 |
| Changing the PivotChart type | 188 |
| Adding data labels to your PivotChart | 189 |
| Sorting the PivotChart | 190 |
| Adding PivotChart titles | 191 |
| Moving the PivotChart legend | 193 |
| Displaying a data table with the PivotChart | 194 |

PART 3: DISCOVERING ADVANCED DATA ANALYSIS TOOLS

| | |
|---|-----|
| CHAPTER 10: Tracking Trends and Making Forecasts | 199 |
| Plotting a Best-Fit Trend line | 200 |
| Calculating Best-Fit Values | 202 |
| Plotting Forecasted Values | 204 |
| Extending a Linear Trend | 206 |
| Extending a linear trend using the fill handle | 206 |
| Extending a linear trend using the Series command | 207 |
| Calculating Forecasted Linear Values | 208 |
| Plotting an Exponential Trend Line | 210 |
| Calculating Exponential Trend Values | 211 |
| Plotting a Logarithmic Trend Line | 213 |
| Plotting a Power Trend Line | 215 |
| Plotting a Polynomial Trend Line | 216 |
| CHAPTER 11: Analyzing Data with Statistics | 219 |
| Counting Things | 220 |
| Counting numbers | 220 |
| Counting nonempty cells | 221 |
| Counting empty cells | 221 |
| Counting cells that match criteria | 221 |
| Counting cells that match multiple criteria | 222 |
| Counting permutations | 223 |
| Counting combinations | 224 |

| | |
|--|------------|
| Averaging Things | 225 |
| Calculating an average | 225 |
| Calculating a conditional average | 225 |
| Calculating an average based on multiple conditions | 226 |
| Calculating the median | 227 |
| Calculating the mode | 227 |
| Finding the Rank | 228 |
| Determining the Nth Largest or Smallest Value | 230 |
| Calculating the nth highest value | 231 |
| Calculating the nth smallest value | 231 |
| Creating a Grouped Frequency Distribution | 232 |
| Calculating the Variance | 233 |
| Calculating the Standard Deviation | 234 |
| Finding the Correlation | 235 |
| CHAPTER 12: Analyzing Data with Descriptive Statistics | 237 |
| Loading the Analysis ToolPak | 238 |
| Generating Descriptive Statistics | 240 |
| Calculating a Moving Average | 243 |
| Determining Rank and Percentile | 245 |
| Generating Random Numbers | 247 |
| Creating a Frequency Distribution | 250 |
| CHAPTER 13: Analyzing Data with Inferential Statistics | 253 |
| Sampling Data | 254 |
| Using the t-Test Tools | 256 |
| Performing a z-Test | 259 |
| Determining the Regression | 261 |
| Calculating the Correlation | 264 |
| Calculating the Covariance | 265 |
| Using the Anova Tools | 267 |
| Performing an f-test | 268 |
| PART 4: THE PART OF TENS | 271 |
| CHAPTER 14: Ten Things You Ought to Know about Statistics | 273 |
| Descriptive Statistics Are Straightforward | 274 |
| Averages Aren't So Simple Sometimes | 274 |
| Standard Deviations Describe Dispersion | 275 |
| An Observation Is an Observation | 276 |
| A Sample Is a Subset of Values | 276 |
| Inferential Statistics Are Cool but Complicated | 277 |

| | |
|--|------------|
| Probability Distributions Aren't Always Confusing | 278 |
| Uniform distribution | 279 |
| Normal distribution | 280 |
| Parameters Aren't So Complicated | 281 |
| Skewness and Kurtosis Describe a Probability Distribution's Shape | 281 |
| Confidence Intervals Seem Complicated at First, but Are Useful | 282 |
| CHAPTER 15: Ten Ways to Analyze Financial Data | 285 |
| Calculating Future Value | 286 |
| Calculating Present Value | 287 |
| Determining Loan Payments | 288 |
| Calculating a Loan Payment's Principal and Interest | 289 |
| Calculating Cumulative Loan Principal and Interest | 289 |
| Finding the Required Interest Rate | 290 |
| Determining the Internal Rate of Return | 291 |
| Calculating Straight-Line Depreciation | 292 |
| Returning the Fixed-Declining Balance Depreciation | 292 |
| Determining the Double-Declining Balance Depreciation | 293 |
| CHAPTER 16: Ten Ways to Raise Your PivotTable Game | 295 |
| Turn the PivotTable Fields Task Pane On and Off | 296 |
| Change the PivotTable Fields Task Pane Layout | 296 |
| Display the Details Behind PivotTable Data | 298 |
| Apply a PivotTable Style | 300 |
| Create a Custom PivotTable Style | 301 |
| Preserve PivotTable Formatting | 303 |
| Rename the PivotTable | 304 |
| Turn Off Grand Totals | 305 |
| Reduce the Size of PivotTable Workbooks | 306 |
| Use a PivotTable Value in a Formula | 306 |
| APPENDIX: GLOSSARY OF DATA ANALYSIS AND EXCEL TERMS | 309 |
| INDEX | 319 |

Introduction

The world is bursting at the seams with data. It's on our computers, it's in our networks, it's on the web. Some days, it seems to be in the very air itself, borne on the wind. But here's the thing: no one actually cares about data. A collection of data — whether it resides on your PC or some giant server somewhere — is really just a bunch of numbers and text, dates and times. No one cares about data because data doesn't *mean* anything. Data isn't cool. You know what's cool? *Knowledge* is cool. *Insight* is cool.

So how do you turn data into knowledge? How do you tweak data to generate insight? You need to organize that data, and then you need to clean it, sort it, filter it, run calculations on it, and summarize it. In a word, you need to *analyze* the data.

Now for the good news: If you have (or can get) that data into Excel, you have a giant basket of data analysis tools at your disposal. Excel really seems to have been made with data analysis in mind, because it offers such a wide variety of features and techniques for organizing, manipulating, and summarizing just about anything that resides in a worksheet. If you can get your data into Excel, Excel will help you turn that data into knowledge and insight.

This book takes you on a tour of Excel's data-analysis tools. You learn everything you need to know to make your data spill its secrets and to uncover your data's hidden-in-plain-sight wisdom. Best of all, if you already know how to perform the basic Excel chores, you don't need to learn any other fancy-schmancy Excel techniques to get started in data analysis. Sweet? You bet.

About This Book

This book contains 16 chapters (and a bonus appendix), but that doesn't mean that you have to, as the King says gravely in *Alice's Adventures in Wonderland*, "Begin at the beginning and go on till you come to the end: then stop." If you've done a bit of data-analysis work in the past, please feel free to dip into the book

wherever it strikes your fancy. The chapters all present their data analysis info and techniques in readily digestible, bite-sized chunks, so you can certainly graze your way through this book.

However, if you're brand-spanking new to data analysis — particularly if you're not even sure what data analysis even *is* — no problem: I'm here to help. To get your data analysis education off to a solid start, I highly recommend reading the book's first three chapters to get some of the basics down cold. From there, you can travel to more advanced territory, safe in the knowledge that you've got some survival skills to fall back on.

What You Can Safely Ignore

This book consists of several hundred pages. Do I expect you to read every word on every page? Yes, I do. Just kidding! No, of course I don't. Entire sections — heck, maybe even entire *chapters* — might contain information that's not relevant to what you do. That's fine and my feelings won't be hurt if you skim through (or — who's kidding whom? — skip over) those parts of the book.

If time (or attention) is short, what else might you want to ignore? Okay, in many places throughout the book I provide step-by-step instructions to complete some task. Each of those steps includes some bold type that gives you the basic instruction. In many cases, however, below that bold text I offer supplementary information to flesh out or extend or explain the bold instruction. Am I just showing off how much I know about all this stuff? Yes, sometimes. Do you have to read these extended instructions? Nope. Read the bold stuff, for sure, but feel free to skip the details if they seem unnecessary or unimportant.

This book also contains a few sidebars that are marked with the Technical Stuff icon. These sidebars contain extra information that's either a bit on the advanced side or goes into heroic, often obscure, detail about the topic at hand. Do you need to read these sidebars? Not at all. Does that make them a waste of page real estate? I don't think so, because they're useful for folks interested in delving into the minutiae of data analysis. If that's not you, ignore away.

If your time is very limited (or you're just aching to get tonight's binge-watching started), you can also ignore the information contained in this book's Tip sidebars. Yes, these tidbits offer easier and faster ways to get things done, so skipping them to save time now might cost you more time in the long run, but, hey, it's a judgment call.

Foolish Assumptions

This book is for people who are new (or relatively new) to Excel data analysis. That doesn't mean, however, that the book is suitable to people who have never used a PC, Microsoft Windows, or even Excel. So first I assume not only that you have a PC running Microsoft Windows but also that you've had some experience with both. (For the purposes of this book, that just means you know how to start and switch between programs.) I also assume that your PC has a recent version of Excel installed. What does "recent" mean? Well, this book is based on Excel 2019, but you should be fine if you're running Excel 2016 or even Excel 2013.

As I said before, I do *not* assume that you're an Excel expert, but I do assume that you know at least the following Excel basics:

- » Creating, saving, opening, and switching between workbooks.
- » Creating and switching between worksheets.
- » Finding and running commands on the Ribbon.
- » Entering numbers, text, dates, times, and formulas into worksheet cells.
- » Working with Excel's basic worksheet functions.

Icons Used in This Book

Like other books in the *For Dummies* series, this book uses icons, or little margin pictures, to flag things that don't quite fit into the flow of the chapter discussion. Here are the icons that I use:



REMEMBER

This icon marks text that contains some things that useful or important enough that you'd do well to store the text somewhere safe in your memory for later recall.



TECHNICAL
STUFF

This icon marks text that contains some for-nerds-only technical details or explanations that you're free to skip.



TIP

This icon marks text that contains a shortcut or easier way to do things, which I hope will make your life — or, at least, the data analysis portion of your life — more efficient.



WARNING

This icon marks text that contains a friendly but unusually insistent reminder to avoid doing something. You have been warned.

Beyond the Book

- » **Examples:** This book's sample Excel workbooks can be found by searching the book's title online at www.dummies.com or at my website: <https://www.mcfedries.com/>.
- » **Cheat Sheet:** To locate this book's Cheat Sheet, go to www.dummies.com and, again, search for *Excel Data Analysis For Dummies*. See the Cheat Sheet for info on Excel database functions, Boolean expressions, and important statistical terms.
- » **Updates:** If this book has any updates after printing, they will be posted to this book's page at www.dummies.com.

Where to Go from Here

If you're just getting your feet wet with Excel data analysis, flip the page and start perusing the first chapter.

If you have some experience with Excel data analysis or you have a special problem or question, use the Table of Contents or the index to find out where I cover that topic and then turn to that page.

Either way, happy analyzing!

1

Getting Started with Data Analysis

IN THIS PART . . .

Understand data analysis and get to know basic analysis features such as conditional formatting and subtotals.

Discover Excel's built-in tools for analyzing data.

Learn how to build Excel tables that hold and store the data you need to analyze.

Find quick and easy ways to begin your analysis using simple statistics, sorting, and filtering.

Get practical stratagems and common-sense tactics for grabbing data from extra sources.

Explore techniques for cleaning and organizing the raw data you want to analyze.

- » Learning about data analysis
- » Analyzing data by applying conditional formatting
- » Adding subtotals to summarize data
- » Grouping related data
- » Combining data from multiple worksheets

Chapter **1**

Learning Basic Data-Analysis Techniques

You are awash in data. Information multiplies around you so fast that you wonder how to make sense of it all. I know, you say, I can paste the data into Excel. That way, you've at least got the data nicely arranged in the worksheet cells, and you can add a little formatting to make things somewhat palatable. That's a fine start, but you're often called upon to do more with your data than make it merely presentable. Your boss, your customer, or perhaps just your curiosity requires you to divine some inner meaning from the jumble of numbers and text that litter your workbooks. In other words, you need to *analyze* your data to see what nuggets of understanding you can unearth.

This chapter gets you started down that data-analysis path by exploring a few straightforward but very useful analytic techniques. After discovering what data analysis entails, you investigate a number of Excel data-analysis techniques, including conditional formatting, data bars, color scales, and icon sets. From there, you dive into some useful methods for summarizing your data, including subtotals, grouping, and consolidation. Before you know it, that untamed wilderness of a worksheet will be nicely groomed and landscaped.

What Is Data Analysis, Anyway?

That's an excellent question! Here's an answer that I unpack for you as I go along: *Data analysis* is the application of tools and techniques to organize, study, reach conclusions and sometimes also make predictions about a specific collection of information.

For example, a sales manager might use data analysis to study the sales history of a product, determine the overall trend, and produce a forecast of future sales. A scientist might use data analysis to study experimental findings and determine the statistical significance of the results. A family might use data analysis to find the maximum mortgage it can afford or how much it must put aside each month to finance retirement or the kids' education.

Cooking raw data

The point of data analysis is to understand information on some deeper, more meaningful level. By definition, *raw data* is a mere collection of facts that by themselves tell you little or nothing of any importance. To gain some understanding of the data, you must manipulate the data in some meaningful way. The purpose of manipulating data can be something as simple as finding the sum or average of a column of numbers or as complex as employing a full-scale regression analysis to determine the underlying trend of a range of values. Both are examples of data analysis, and Excel offers a number of tools — from the straightforward to the sophisticated — to meet even the most demanding needs.

Dealing with data

The “data” part of data analysis is a collection of numbers, dates, and text that represents the raw information you have to work with. In Excel, this data resides inside a worksheet, which makes the data available for you to apply Excel's satisfyingly large array of data-analysis tools.

Most data-analysis projects involve large amounts of data, and the fastest and most accurate way to get that data onto a worksheet is to import it from a non-Excel data source. In the simplest scenario, you can copy the data — from a text file, a Word table, or an Access datasheet — and then paste it into a worksheet. However, most business and scientific data is stored in large databases, and Excel offers tools to import the data you need into your worksheet. I talk about all this in more detail later in the book.

After you have your data in the worksheet, you can leave it as a regular range and still apply many data-analysis techniques to the data. However, if you convert the range into a *table*, Excel treats the data as a simple database and enables you to apply a number of database-specific analysis techniques to the table.

Building data models

In many cases, you perform data analysis on worksheet values by organizing those values into a *data model*, a collection of cells designed as a worksheet version of some real-world concept or scenario. The model includes not only the raw data but also one or more cells that represent some analysis of the data. For example, a mortgage amortization model would have the mortgage data — interest rate, principal, and term — and cells that calculate the payment, principal, and interest over the term. For such calculations, you use formulas and Excel's built-in worksheet functions.

Performing what-if analysis

One of the most common data-analysis techniques is *what-if analysis*, for which you set up worksheet models to analyze hypothetical situations. The “what-if” part means that these situations usually come in the form of a question: “What happens to the monthly payment if the interest rate goes up by 2 percent?” “What will the sales be if you increase the advertising budget by 10 percent?” Excel offers four what-if analysis tools: data tables, Goal Seek, Solver, and scenarios, all of which I cover in this book.

Analyzing Data with Conditional Formatting

Many Excel worksheets contain hundreds of data values. You could try to make sense of such largish sets of data by creating complex formulas and wielding Excel's powerful data-analysis tools. However, just as you wouldn't use a steam-roller to crush a tin can, sometimes these sophisticated techniques are too much tool for the job. For example, what if all you want are answers to simple questions such as the following:

- » Which cell values are less than 0?
- » What are the top 10 values?
- » Which cell values are above average, and which are below average?

These simple questions aren't easy to answer just by glancing at the worksheet, and the more numbers you're dealing with, the harder it gets. To help you "eyeball" your worksheets and answer these and similar questions, Excel lets you apply *conditional formatting* to the cells. This is a special format that Excel applies only to cells that satisfy some condition, which Excel calls a *rule*. For example, you could apply formatting to show all the negative values in a red font, or you could apply a filter to show only the top 10 values.

Highlighting cells that meet some criteria

A *conditional format* is formatting that Excel applies only to cells that meet the criteria you specify. For example, you can tell Excel to apply the formatting only if a cell's value is greater or less than some specified amount, between two specified values, or equal to some value. You can also look for cells that contain specified text, dates that occur during a specified time frame, and more.

When you set up your conditional format, you can specify the font, border, and background pattern. This formatting helps to ensure that the cells that meet your criteria stand out from the other cells in the range. Here are the steps to follow:

- 1. Select the range you want to work with.**

Just select the data values you want to format. You don't have to (in fact, you shouldn't) select any surrounding data.

- 2. Choose Home ⇨ Conditional Formatting.**

- 3. Choose Highlight Cells Rules and then select the rule you want to use for the condition.**

You have six rules to play around with:

- **Greater than:** Applies the conditional format to cells that have a value larger than a value that you specify.
- **Less than:** Applies the conditional format to cells that have a value smaller than a value that you specify.
- **Between:** Applies the conditional format to cells that have a value that is greater than or equal to a minimum value that you specify and less than or equal to a maximum value that you specify.
- **Equal To:** Applies the conditional format to cells that have a value that is the same as a value that you specify.
- **Text that Contains:** Applies the conditional format to cells that include the text that you specify.

- **A Date Occurring:** Applies the conditional format to cells that have a date value that meets the condition that you specify (such as Yesterday, Last Week, or Next Month).

(There's a seventh rule here — Duplicate Values — that I cover later in this chapter.) A dialog box appears, the name of which depends on the rule you click in Step 3. For example, Figure 1-1 shows the dialog box for the Greater Than rule.

FIGURE 1-1:
The Greater Than
dialog box with
some highlighted
values.

| | A | B | C | D | E | F | G | H | I | J | K |
|----|--|------|------|------|------|------|------|------|------|------|------|
| 1 | GDP — % Annual Growth Rates (Source: The World Bank) | | | | | | | | | | |
| 2 | Country Name | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
| 3 | Afghanistan | 13.7 | 3.6 | 21.0 | 8.4 | 6.1 | 14.4 | 3.9 | 2.7 | 1.3 | 2.4 |
| 4 | Albania | 5.9 | 3.8 | 3.4 | 3.7 | 2.6 | 1.4 | 1.0 | 1.8 | 2.2 | 3.4 |
| 5 | Algeria | 3.4 | 2.4 | 1.6 | 3.6 | 2.9 | 3.4 | 2.8 | 3.8 | 3.8 | 3.3 |
| 6 | American Samoa | 2.0 | -2.6 | -4.2 | 0.4 | 0.3 | -4.4 | -2.8 | 0.9 | 1.2 | -2.6 |
| 7 | Andorra | 0.0 | -8.6 | -3.7 | -5.4 | -4.6 | -1.6 | 0.4 | 2.3 | 0.8 | 1.2 |
| 8 | Angola | 22.6 | 13.8 | 2.4 | 3.4 | 3.9 | 5.2 | 6.8 | 4.8 | 3.0 | -0.7 |
| 9 | Antigua and Barbuda | 9.3 | 0.0 | -2.1 | 7.3 | 3.1 | 3.5 | 0.1 | 5.1 | 1.1 | 5.3 |
| 10 | Arab World | 4.6 | 5.8 | -1.1 | 3.1 | 2.1 | 2.1 | 0.1 | 1.1 | 1.1 | 3.2 |
| 11 | Argentina | 9.0 | 4.1 | -1.1 | 3.1 | 2.1 | 2.1 | 0.1 | 1.1 | 1.1 | -2.2 |
| 12 | Armenia | 13.7 | 6.9 | -1.1 | 3.1 | 2.1 | 2.1 | 0.1 | 1.1 | 1.1 | 0.2 |
| 13 | Australia | 3.7 | 3.7 | -1.1 | 3.1 | 2.1 | 2.1 | 0.1 | 1.1 | 1.1 | 2.8 |
| 14 | Austria | 3.6 | 1.5 | -1.1 | 3.1 | 2.1 | 2.1 | 0.1 | 1.1 | 1.1 | 1.5 |
| 15 | Azerbaijan | 25.0 | 10.8 | 9.4 | 4.9 | 0.1 | 2.2 | 5.8 | 2.0 | 1.1 | -3.1 |
| 16 | Bahamas, The | 1.4 | -2.3 | -4.2 | 1.5 | 0.6 | 3.1 | -0.6 | -1.2 | -3.1 | 0.2 |

Greater Than

Format cells that are GREATER THAN:

5 with Light Red Fill with Dark Red Text

OK Cancel

4. Type the value to use for the condition.

You can also click the button that appears to the right of the text box and then select a worksheet cell that contains the value. Also, depending on the operator, you might need to specify two values.

5. Use the drop-down list to select the formatting to apply to cells that match your condition.

If you're feeling creative, you can make up your own format by selecting the Custom Format command.

6. Click OK.

Excel applies the formatting to cells that meet the condition you specified.



TIP

Excel enables you to specify multiple conditional formats. For example, you can set up one condition for cells that are greater than some value, and a separate condition for cells that are less than some other value. You can apply unique formats to each condition. Follow the same steps to configure the new condition.

Showing pesky duplicate values

You use conditional formatting mostly to highlight numbers greater than or less than some value, or dates occurring within some range. However, you can also use conditional formatting to look for duplicate values in a range. Why would you want to do that? The main reason is that many range or table columns require unique values. For example, a column of student IDs or part numbers shouldn't have duplicates.

Unfortunately, scanning such numbers and picking out the repeat values is hard. Not to worry! With conditional formatting, you can specify a font, border, and background pattern that ensures that any duplicate cells in a range or table stand out from the other cells. Here's what you do:

1. **Select the range that you want to check for duplicates.**
2. **Choose Home ⇨ Conditional Formatting.**
3. **Choose Highlight Cells Rules ⇨ Duplicate Values.**

The Duplicate Values dialog box appears. The left drop-down list has Duplicates selected by default, as shown in Figure 1-2. However, if you want to highlight all the unique values instead of the duplicates, select Unique from this list.

4. **Use the right drop-down list to select the formatting to apply to the cells with duplicate values.**

You can create your own format by choosing the Custom Format command.

5. **Click OK.**

Excel applies the formatting to any cells that have duplicate values in the range.

| | A | B | C |
|----|-----------|--------------------------------------|---------|
| 1 | Code | Product Name | Cost |
| 2 | NWTB-1 | Northwind Traders Chai | \$13.50 |
| 3 | NWTCO-3 | Northwind Traders Syrup | \$7.50 |
| 4 | NWTCO-4 | Northwind Traders Cajun Seasoning | \$16.50 |
| 5 | NWTO-5 | Northwind Traders Olive Oil | \$16.01 |
| 6 | NWTJP-6 | Northwind Traders Boysenberry Spread | \$18.75 |
| 7 | NWTJP-6 | Northwind Traders Marmalade | \$60.75 |
| 8 | NWTFDN-7 | Northwind Traders | |
| 9 | NWTS-8 | Northwind Traders | |
| 10 | NWTFDN-14 | Northwind Traders | |
| 11 | NWTCFV-17 | Northwind Traders | |
| 12 | NWTBGM-19 | Northwind Traders | |
| 13 | NWTBGM-21 | Northwind Traders Scones | \$7.50 |
| 14 | NWTB-34 | Northwind Traders Beer | \$10.50 |
| 15 | NWTCM-40 | Northwind Traders Crab Meat | \$13.80 |
| 16 | NWTSO-41 | Northwind Traders Clam Chowder | \$7.24 |
| 17 | NWTB-34 | Northwind Traders Coffee | \$34.50 |
| 18 | NWTCA-48 | Northwind Traders Chocolate | \$9.56 |

FIGURE 1-2: Use the Duplicate Values rule to highlight worksheet duplicates.

Highlighting the top or bottom values in a range

When analyzing worksheet data, looking for items that stand out from the norm is often useful. For example, you might want to know which sales reps sold the most last year, or which departments had the lowest gross margins. To quickly and easily view the extreme values in a range, you can apply a conditional format to the top or bottom values of that range.

You can apply such a format by setting up a *top/bottom rule*, in which Excel applies a conditional format to those items that are at the top or bottom of a range of values. For the top or bottom values, you can specify a number, such as the top 5 or 10, or a percentage, such as the bottom 20 percent. Here's how it works:

1. **Select the range you want to work with.**
2. **Choose Home → Conditional Formatting.**
3. **Choose Top/Bottom Rules and then select the type of rule you want to create.**

You have six rules to mess with:

- **Top 10 Items:** Applies the conditional format to cells that rank in the top *X*, where *X* is a number that you specify (the default is 10).
- **Top 10 %:** Applies the conditional format to cells that rank in the top *X* %, where *X* is a number that you specify (the default is 10).
- **Bottom 10 Items:** Applies the conditional format to cells that rank in the bottom *X*, where *X* is a number that you specify (the default is 10).
- **Bottom 10 %:** Applies the conditional format to cells that rank in the bottom *X* %, where *X* is a number that you specify (the default is 10).
- **Above Average:** Applies the conditional format to cells that rank above the average value of the range.
- **Below Average:** Applies the conditional format to cells that rank below the average value of the range.

A dialog box appears, the name of which depends on the rule that you selected in Step 3. For example, Figure 1-3 shows the dialog box for the Top Ten Items rule.

4. **Type the value to use for the condition.**

You can also click the button that appears to the right of the text box and then select a worksheet cell that contains the value. Note that you don't need to enter a value for the Above Average and Below Average rules.

5. Use the drop-down list to select the formatting to apply to cells that match your condition.
6. Click OK.

Excel applies the formatting to cells that meet the condition you specified.

| | A | B | C | D |
|----|---|------------------|-------------------|-------------------|
| 1 | | Sales Rep | 2017 Sales | 2018 Sales |
| 2 | | Nancy Freehafer | \$996,336 | \$960,492 |
| 3 | | Andrew Cencini | \$606,731 | \$577,983 |
| 4 | | Jan Kotas | \$622,781 | \$967,580 |
| 5 | | Mariya Sergienko | \$765,327 | \$771,399 |
| 6 | | Steven Thorpe | \$863,589 | \$827,213 |
| 7 | | Michael Neipper | \$795,518 | \$669,394 |
| 8 | | Robert Zare | \$722,740 | \$626,945 |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | Andrea Granek | \$540,484 | \$693,762 |
| 15 | | Charles Aster | \$650,733 | \$823,034 |
| 16 | | Karen Reilly | \$509,863 | \$511,569 |
| 17 | | Karen Munson | \$503,699 | \$975,455 |
| 18 | | Vince Voyatzis | \$630,263 | \$599,514 |
| 19 | | Paul Sellars | \$779,722 | \$596,353 |

FIGURE 1-3:
The Top 10 Items
dialog box with
the top 5 values
highlighted.



REMEMBER

When you set up your top/bottom rule, select a format that ensures that the cells that meet your criteria stand out from the other cells in the range. If none of the predefined formats suits your needs, you can always choose Custom Format and then use the Format Cells dialog box to create a suitable formatting combination. Use the Font, Border, and Fill tabs to specify the formatting you want to apply, and then click OK.

Analyzing cell values with data bars

In some data-analysis scenarios, you might be interested more in the relative values within a range than the absolute values. For example, if you have a table of products that includes a column showing unit sales, you might want to compare the relative sales of all the products.

Comparing relative values is often easiest if you visualize the values, and one of the easiest ways to visualize data in Excel is to use *data bars*, a data visualization feature that applies colored, horizontal bars to each cell in a range of values; these

bars appear “behind” (that is, in the background of) the values in the range. The length of the data bar that appears in each cell depends on the value in that cell: the larger the value, the longer the data bar.

Follow these steps to apply data bars to a range:

1. **Select the range you want to work with.**
2. **Choose Home ⇨ Conditional Formatting.**
3. **Choose Data Bars and then select the fill type of data bars you want to create.**

You can apply two type of data bars:

- **Gradient fill:** The data bars begin with a solid color and then gradually fade to a lighter color.
- **Solid fill:** The data bars are a solid color.

Excel applies the data bars to each cell in the range. Figure 1-4 shows an example in the Units column.

| | A | B | C |
|----|--|-------|-----------|
| 1 | Product Name | Units | \$ Total |
| 2 | Northwind Traders Almonds | 20 | \$ 200 |
| 3 | Northwind Traders Beer | 487 | \$ 6,818 |
| 4 | Northwind Traders Boysenberry Spread | 100 | \$ 2,500 |
| 5 | Northwind Traders Cajun Seasoning | 40 | \$ 880 |
| 6 | Northwind Traders Chai | 40 | \$ 720 |
| 7 | Northwind Traders Chocolate | 200 | \$ 2,550 |
| 8 | Northwind Traders Chocolate Biscuits Mix | 85 | \$ 782 |
| 9 | Northwind Traders Clam Chowder | 290 | \$ 2,799 |
| 10 | Northwind Traders Coffee | 650 | \$ 29,900 |
| 11 | Northwind Traders Crab Meat | 120 | \$ 2,208 |
| 12 | Northwind Traders Curry Sauce | 65 | \$ 2,600 |
| 13 | Northwind Traders Dried Apples | 40 | \$ 2,120 |
| 14 | Northwind Traders Dried Pears | 40 | \$ 1,200 |
| 15 | Northwind Traders Dried Plums | 75 | \$ 263 |
| 16 | Northwind Traders Fruit Cocktail | 40 | \$ 1,560 |
| 17 | Northwind Traders Gnocchi | 10 | \$ 380 |
| 18 | Northwind Traders Green Tea | 275 | \$ 822 |
| 19 | Northwind Traders Long Grain Rice | 40 | \$ 280 |
| 20 | Northwind Traders Marmalade | 40 | \$ 2,240 |

FIGURE 1-4:
The higher the
value, the longer
the data bar.



TIP

If your range includes right-aligned values, the gradient-fill data bars are a better choice than the solid-fill data bars because even the longest gradient-fill bars fade to white toward the right edge of the cell, so your range values should mostly appear on a white background, making them easier to read.

Analyzing cell values with color scales

Getting some idea about the overall distribution of values in a range is often useful. For example, you might want to know whether a range has many low values and just a few high values. *Color scales* can help you analyze your data in this way. A color scale compares the relative values in a range by applying shading to each cell, where the color reflects each cell's value.

Color scales can also tell you whether your data includes *outliers*: values that are much higher or lower than the others. Similarly, color scales can help you make value judgments about your data. For example, high sales and low numbers of product defects are good, whereas low margins and high employee turnover rates are bad.

1. **Select the range you want to format.**
2. **Choose Home ⇨ Conditional Formatting.**
3. **Choose Color Scales and then select the color scale that has the color scheme you want to apply.**

The color scales come in two varieties: three-color scales and two-color scales. If your goal is to look for outliers, go with a three-color scale because it helps the outliers stand out more. A three-color scale is also useful if you want to make value judgments about your data, because you can assign your own values to the colors (such as positive, neutral, and negative). Use a two-color scale when you want to look for patterns in the data, because a two-color scale offers less contrast.

Excel applies the color scale to each cell in your selected range.

Analyzing cell values with icon sets

Symbols that have common or well-known associations are often useful for analyzing large amounts of data. For example, a check mark usually means that something is good or finished or acceptable, whereas an X means that something is bad or unfinished or unacceptable. Similarly, a green circle is positive, whereas a red circle is negative (think traffic lights). Excel puts these and other symbolic associations to good use with the *icon sets* feature. You use icon sets to visualize the relative values of cells in a range.



REMEMBER

With icon sets, Excel adds a particular icon to each cell in the range, and that icon tells you something about the cell's value relative to the rest of the range. For example, the highest values might be assigned an upward-pointing arrow, the lowest values a downward-pointing arrow, and the values in between a horizontal arrow.