Visual®

Read Less-Learn More®

Paul McFedries

Microsoft® Excel® PivotTables and PivotCharts

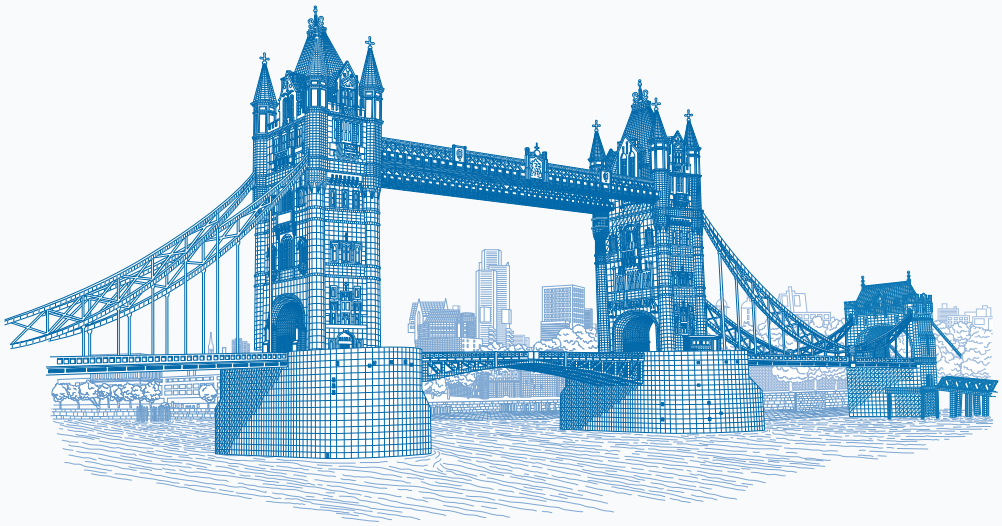
2nd Edition

- Companion Web site features example files to illustrate the lessons

Your visual blueprint™ for creating dynamic spreadsheets

Excel® PivotTables and PivotCharts

Your visual blueprint™ for creating dynamic spreadsheets, 2nd Edition



by Paul McFedries

WILEY
Wiley Publishing, Inc.

Excel® PivotTables and PivotCharts: Your visual blueprint™ for creating dynamic spreadsheets, 2nd Edition

Published by
Wiley Publishing, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256

www.wiley.com

Published simultaneously in Canada

Library of Congress Control Number: 2010926851

ISBN: 978-0-470-59161-1

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Trademark Acknowledgments

Contact Us

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 www.wiley.com/techsupport.

The Tower Bridge

Truly a 19th century architectural wonder, London's Tower Bridge took 432 construction workers eight years to build. Architect Horace Jones designed the massive bascule bridge—a type of counterweighted drawbridge—and at its completion in 1894, it was the largest and most sophisticated of its kind ever built.

Discover more about London's historic architecture in *Frommer's England 2010* (ISBN 978-0-470-47070-1), available wherever books are sold or at www.Frommers.com.

Disclaimer

In order to get this information to you in a timely manner, this book was based on a pre-release version of Microsoft Office 2010. There may be some minor changes between the screenshots in this book and what you see on your desktop. As always, Microsoft has the final word on how programs look and function; if you have any questions or see any discrepancies, consult the online help for further information about the software.

Sales

Contact Wiley
at (877) 762-2974
or (317) 572-4002.

Credits

Executive Editor
Jody Lefevere

Project Editor
Kristin DeMint

Technical Editor
Namir Shammas

Copy Editor
Kim Heusel

Editorial Director
Robyn Siesky

Editorial Manager
Cricket Krengel

Business Manager
Amy Knies

Senior Marketing Manager
Sandy Smith

Vice President and Executive Group Publisher
Richard Swadley

Vice President and Executive Publisher
Barry Pruett

Project Coordinator
Lynsey Stanford

Graphics and Production Specialists
Carrie Cesavice
Joyce Haughey
Andrea Hornberger
Jennifer Mayberry

Quality Control Technician
Lauren Mandelbaum

Proofreading and Indexing
Penny Stuart
Potomac Indexing, LLC

Media Development Project Manager
Laura Moss

Media Development Assistant Project Manager
Jenny Swisher

Media Development Associate Producer
Marilyn Hummel

Screen Artist
Jill A. Proll

Illustrator
Cheryl Grubbs

Special Help
Rebekah Worthman

About the Author

Paul McFedries is a full-time technical writer. Paul has been authoring computer books since 1991 and he has more than 70 books to his credit. Paul's books have sold more than three million copies worldwide. These books include the Wiley titles *Teach Yourself VISUALLY Excel 2010; Excel 2010 Visual Quick Tips; Teach Yourself VISUALLY Windows 7;* and *Teach Yourself VISUALLY Office 2008 for Mac*. Paul is also the proprietor of Word Spy (www.wordspy.com and twitter.com/wordspy), a Web site that tracks new words and phrases as they enter the language. Paul invites you to drop by his personal Web site at www.mcfedries.com or to follow him on Twitter at twitter.com/paulmcf.

Author's Acknowledgments

It goes without saying that writers focus on text, and I certainly enjoyed focusing on the text that you'll read in this book. However, this book is more than just the usual collection of words and phrases. A quick thumb-through of the pages will show you that this book is also chock full of images, from sharp screen shots to fun and informative illustrations. Those images sure make for a beautiful book, and that beauty comes from a lot of hard work by Wiley's immensely talented group of designers and layout artists. They are all listed in the Credits section on the previous page, and I thank them for creating another gem. Of course, what you read in this book must also be accurate, logically presented, and free of errors. Ensuring all of this was an excellent group of editors that included project editor Kristin DeMint, copy editor Kim Heusel, and technical editor Namir Shammas. Thanks to all of you for your exceptional competence and hard work. Thanks, as well, to acquisitions editor Jody Lefevere for asking me to write this book.

How to Use This Visual Blueprint Book

Who This Book Is For

This book is for advanced computer users who want to take their knowledge of this particular technology or software application to the next level.

The Conventions in This Book

1 Steps

This book uses a step-by-step format to guide you easily through each task. Numbered steps are actions you must do; bulleted steps clarify a point, step, or optional feature; and indented steps give you the result.

2 Notes

Notes give additional information — special conditions that may occur during an operation, a situation that you want to avoid, or a cross reference to a related area of the book.

3 Icons and Buttons

Icons and buttons show you exactly what you need to click to perform a step.

4 Extra or Apply It

An Extra section provides additional information about the preceding task — insider information and tips for ease and efficiency. An Apply It section takes the code from the preceding task one step further and allows you to take full advantage of it.

5 Bold

Bold type shows text or numbers you must type.

6 Italics

Italic type introduces and defines a new term.

7 Courier Font

`Courier font` indicates the use of scripting language code such as statements, operators, or functions, and code such as objects, methods, or properties.

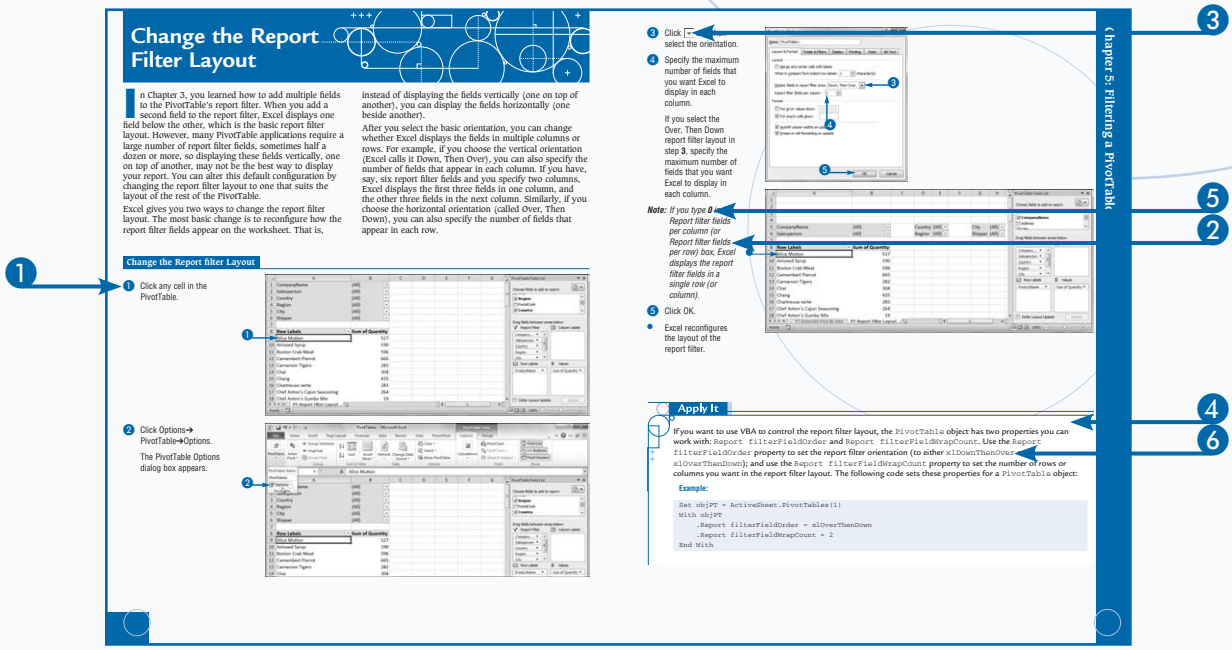


TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

Understanding Data Analysis

The PivotTables and PivotCharts that you learn about in this book are part of the larger category of *data analysis*. You can get the most out of these tools if you have a broader understanding of what data analysis is, what its benefits are, and what other tools are available to you.

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. 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.

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 it in some meaningful way. This 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.

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 and you get it there in one of two ways: You enter it by hand or you import it from an external source. You can then either leave the data as a regular range, or you can convert it into a table for easier data manipulation.

Data Entry

In many data analysis situations, the required data must be entered into the worksheet manually. For example, if you want to determine a potential monthly mortgage payment, you must first enter values such as the current interest rate, the principal, and the term. Manual data entry is suitable for small projects only, because entering hundreds or even thousands of values is time consuming and can lead to errors.

Imported Data

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. See Appendixes B and C for more about these tools.

Table

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 flat-file database and enables you to apply a number of database-specific analysis techniques to the table. To learn how to do this, see Chapter 2.

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 functions, as described in Appendix A.

Formulas

A *formula* is a set of symbols and values that perform some kind of calculation and produce a result. All Excel formulas have the same general structure: an equals sign (=) followed by one or more *operands* — which can be a value, a cell reference, a range, a range name, or a function name — separated by one or more *operators*, which are the symbols that combine the operands in some way, such as the plus sign (+) and the multiplication sign (*). For example, the formula `=A1+A2` adds the values in cells A1 and A2.

Functions

A *function* is a predefined formula that is built in to Excel. Each function takes one or more inputs — called *arguments,* such as numbers or cell references — and then returns a result. Excel offers hundreds of functions and you can use them to compute averages, determine the future value of an investment, compare values, and much more.

What-If Analysis

One of the most common data analysis techniques is *what-if analysis*, where you set up worksheet models to analyze hypothetical situations. The what-if part comes from the fact 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.

Data Tables

A *data table* is a range of cells where one column consists of a series of values, called *input cells*. You can then apply each of those inputs to a single formula, and Excel displays the results for each case. For example, you can use a data table to apply a series of interest rate values to a formula that calculates the monthly payment for a loan or mortgage.

Goal Seek

You use Excel's Goal Seek tool when you want to manipulate one formula component — called the *changing cell* — in such a way that the formula produces a specific result. For example, in a *break-even analysis*, you determine the number of units of a product that you must sell for the profit to be 0. Given a formula that calculates profit, you can use Goal Seek to determine the break-even point.

Solver

You use Excel's Solver tool when you want to manipulate multiple formula components — called the *changing cells* — in such a way that the formula produces the optimal result. For example, you can use Solver to tackle the so-called *transportation problem*, where the goal is to minimize the cost of shipping goods from several product plants to various warehouses around the country.

Scenarios

A *scenario* is a collection of input values that you plug into formulas within a model to produce a result. The idea is that you make up scenarios for various situations — for example, best-case, worst-case, and so on — and Excel's Scenario Manager saves each one. Later you can apply any of the saved scenarios, and Excel automatically applies all the input values to the model.

Introducing the PivotTable

Tables and external databases can contain hundreds or even thousands of records. Analyzing that much data can be a nightmare without the right kinds of tools. To help you, Excel offers a powerful data analysis tool called a *PivotTable*. This tool enables you to summarize hundreds of records in a concise tabular format. You can then manipulate the layout of — or *pivot* — the table to see different views of your data.

This book teaches you everything you need to know (and, indeed, just about everything there *is* to know) about PivotTables. You learn how to create them, edit them, pivot them, format them, calculate with them, and much more. You can get more out of the rest of the book if you take a few minutes now to get acquainted with some PivotTable background and basics.

Database Analysis

To understand PivotTables, you need to see how they fit in with Excel's other database-analysis features. Database analysis has three levels of complexity: lookup and retrieval, criteria and table functions, and multiple variables. As you move from one level to another, the need for PivotTables becomes apparent.

Lookup and Retrieval

The simplest level of database analysis involves the basic lookup and retrieval of information. For example, if you have a database that lists the company sales reps and their territory sales, you can use a data form (or even Excel's Find feature) to search for a specific rep and to look up the sales in that rep's territory.

Criteria and Table Functions

The next level of database analysis complexity involves more sophisticated lookup and retrieval systems in which you apply criteria to work with a subset of the data. You can then use this subset to apply subtotals and Excel's table functions (such as the `DSUM()` function, which sums those table cells that meet some specified criteria). For example, suppose that each sales territory is part of a larger region and you want to know the total sales in the eastern region. You can either subtotal by region or set up your criteria to match all territories in the eastern region and use `DSUM()` to get the total. To get more specific information, such as the total eastern region sales in the second quarter, you just add the appropriate conditions to your criteria.

Multiple Variables

The next level of database analysis applies a single question to multiple variables. For example, if the company in the preceding example has four regions, you might want to see separate totals for each region broken down by quarter. One solution would be to set up four different criteria and four different `DSUM()` functions. But what if there were a dozen regions? Or a hundred? Ideally, you need some way of summarizing the database information into a sales table that has a row for each region and a column for each quarter. This is exactly what PivotTables do and, as you see with Excel's PivotTable Wizard in Chapter 2, you can create your own PivotTables with just a few mouse clicks.

What PivotTables Do

PivotTables help you analyze large amounts of data by performing three operations: grouping the data into categories, summarizing the data using calculations, and filtering the data to show just the records you want to work with.

Grouping

A PivotTable is a powerful data-analysis tool in part because it automatically groups large amounts of data into smaller, more manageable categories. For example, suppose you have a data source with a Region field where each cell contains one of four values: East, West, North, and South. The original data may contain thousands of records, but if you build your PivotTable using the Region field, the resulting table has just four rows — one each for the four unique Region values in your data.

You can also create your own grouping after you build your PivotTable. For example, if your data has a Country field, you can build the PivotTable to group together all the records that have the same Country value. When you have done that, you can further group the unique Country values into continents: North America, South America, Europe, and so on. See Chapter 4 to learn how to group PivotTable values.

Summarizing

In conjunction with grouping data according to the unique values in one or more fields, Excel also displays summary calculations for each group. The default calculation is Sum, which means for each group, Excel totals all the values in some specified field. For example, if your data has a Region field and a Sales field, a PivotTable can group the unique Region values and display the total of the Sales values for each one. Excel has other summary calculations, including Count, Average, Maximum, Minimum, and Standard Deviation.

Even more powerful, a PivotTable can display summaries for one grouping broken down by another. For example, suppose your sales data also has a Product field. You can set up a PivotTable to show the total Sales for each Product, broken down by Region.

Filtering

A PivotTable also enables you to view just a subset of the data. For example, by default the PivotTable's groupings show all the unique values in the field. However, you can manipulate each grouping to hide those that you do not want to view; see Chapter 4 for more. Each PivotTable also comes with a report filter — see the section "Explore PivotTable Features" later in this chapter — that enables you to apply a filter to the entire PivotTable. For example, suppose your sales data also includes a Customer field. By placing this field in the PivotTable's report filter, you can filter the PivotTable report to show just the results for a single Customer.

PivotTable Limitations

PivotTables come with certain limitations and restrictions that you need to be familiar with. See the section "Explore PivotTable Features" later in this chapter for explanations of the PivotTable terminology used here:

- The maximum number of row fields is 1,048,576. (If you are using a version of Excel prior to 2007, the maximum number is 65,536.)
- The maximum number of column fields is 16,384. (If you are using a version of Excel prior to 2007, the maximum number is 256.)
- The maximum number of page fields is 256.
- The maximum number of data fields is 256.
- The maximum number of unique items that can appear in a row, column, or page field is 1,048,576. (If you are using Excel 2003, the maximum number is 32,500; if you are using a version of Excel prior to 2003, the maximum number is 8,000.)
- The size and number of PivotTables are limited by how much available memory your system has.

Learn PivotTable Benefits

If Excel comes with so many powerful data analysis tools and features, why do you need to learn how to build and work with PivotTables? The short answer is that PivotTables are a useful weapon to add to your data-analysis arsenal. The long answer is that PivotTables are worth learning because they come with not just one or two, but a long list of benefits.

PivotTables are easy to build and maintain; they perform large and complex calculations amazingly fast; you can quickly and easily update them to account for new data; PivotTables are dynamic, so components can be easily moved, filtered, and added to; they are fully customizable so you can build each report the way you want; and, finally, PivotTables can use most of the formatting options that you can apply to regular Excel ranges and cells.

PivotTables Save Time

These days, most people have far too much to do and far too little time in which to do it. Computers are supposed to help us with this problem by reducing the amount of time spent on routine tasks, such as adding up rows of numbers. Some computer features have the opposite effect — e-mail, for example, takes up increasing amounts of time — but PivotTables is not one of these features. The chore PivotTables are designed to replace — cross-tabulating massive amounts of data — is inherently time consuming. But PivotTables, by virtue of being easy to use, lightning fast, and readily updated, reduce that time to a mere fraction of what it was, resulting in true time savings.

Easy

Perhaps the most important benefit of PivotTables is that they do not come with a daunting learning curve. After you understand the basic features, you can use Excel's Summarize with PivotTable command to build a simple PivotTable report with as little as five or six mouse clicks; see the section "Explore PivotTable Features" later in this chapter. Even the most complex PivotTables are not much harder to build because the PivotTable Tools tab offers everything you need to configure and format a PivotTable.

Fast

The average PivotTable must do quite a bit of work when it generates its report: It must analyze hundreds or even thousands of records, each of which may have a dozen or more fields; extract the unique values from one or more fields; calculate the data summary for each unique item; and then lay everything out on the worksheet. Amazingly, for all but the largest data sources, this entire process usually only takes a second or two.

Updateable

PivotTables are often used in situations where the original data changes. When that happens, the PivotTable can become out of date. However, each PivotTable "remembers" the original data upon which the report was based. This means that when a PivotTable is out of date, you do not need to re-create the report from scratch. Instead, you can run the Refresh Data command, which instantly updates the PivotTable with the latest data. You can even set up your PivotTable to refresh its data automatically. For the details on refreshing PivotTables, see Chapter 3.

PivotTables Are Flexible

One of the traits that makes a PivotTable a powerful data analysis tool is its flexibility. For example, when you create a PivotTable, the resulting report is not set in stone. Instead, you can move components from one part of the PivotTable to another, filter the results, add and remove data, and more. Another aspect of the flexibility of PivotTables is their versatility, which means that you can create them from more than just Excel ranges and tables.

Dynamic

Every PivotTable is a dynamic creation that you can reconfigure to produce the kind of report you need. Specifically, most of the fields that you add to the PivotTable you can also move from one part of the report to another. This is called *pivoting* the data and it causes Excel to reconfigure the PivotTable and recalculate the results. Excel produces the updated PivotTable immediately, so you can use this feature as needed, making PivotTables even more powerful and useful. See Chapter 4 to learn how to pivot data.

Manipulable

You can easily and quickly manipulate your PivotTable layout to get the results you are looking for. For example, you can always add new fields to any part of the PivotTable, usually with just a few mouse clicks, and you can easily remove any fields that you no longer need. Also, as you learned in the previous section, you can group and filter the PivotTable results to work with just the data you need.

Versatile

If you could create PivotTables only from an Excel range or table, they would still be enormously useful. However, Excel has made PivotTables versatile enough to handle many other types of data. You can create them from Access tables, Word tables, text files, Web pages, XML data, and from tables in powerful database systems such as SQL Server and Online Analytical Processing (OLAP) servers. See Chapter 10 to build advanced PivotTables; see Chapter 11 to build a PivotTable from an OLAP Cube.

PivotTables Suit Your Needs

Although many of the PivotTables that you create will be for your own use, you are also likely to set up PivotTables for other people to view, either on-screen, on paper, or even on the Web (see Chapter 3). In these more public situations, you will usually want to set up your PivotTable so that it looks its best. To that end, Excel has given PivotTables a number of features that enable you to customize and format them as needed.

Customizing

Each PivotTable comes with a number of options that you can use to customize both the report as a whole and individual PivotTable components. For example, you can hide items, sort the data, and customize the report printout. You can also customize the calculations used in the report, either by changing to one of Excel's built-in calculations or by defining custom calculations. For more about custom calculations, see Chapter 8.

Formatting

After you have the PivotTable result you want, you can spend time dressing up the report to make the data easier on the eyes. Fortunately, most of the cells in a PivotTable act as regular Excel cells. This means you can format them in the same way by changing the font, applying colors and borders, using numeric and date formats, and much more. See Chapter 5 to learn about customizing your PivotTable fields.

Learn When to Use PivotTables

One of the keys to using Excel's data-analysis tools is knowing which tool to use under which circumstance. If you want to glean one or two facts about your data, then a formula or two is often all you need. For more elaborate needs, especially ones where you need to build a worksheet version of some real-world concept, a data model is required. If you want to "interrogate" your data by plugging various values into a formula and comparing the results, a data table is best. If you are looking for a particular or optimal result, use Goal Seek for simple models or Solver for more complex models.

PivotTables, too, are best used only in certain scenarios. Those where a PivotTable is your best data-analysis tool — or at least a worthwhile one to consider — depend on one of three factors: the structure of the underlying data, the analysis you require, and your (or your manager's) reporting needs.

Data Structure

More than any other factor, the structure of your data determines whether a PivotTable is a good data analysis choice. Certain types of data simply cannot be analyzed in a PivotTable, while other data sets would produce largely useless results. In general, the best data structure for PivotTables is one where the data exists in a tabular format with consistent and repeated data, such as those found in databases of transactions. For more detailed information on setting up your data for a PivotTable report, see Chapter 2.

Tabular Data

Your data is a good candidate for a PivotTable analysis if it exists in tabular format. This means that the data is arranged in a row-and-column structure, with the same number of columns used in each row. If your data is scattered around the worksheet and cannot be rearranged into tabular format, you cannot build a PivotTable from it.

Consistent and Repeated Data

You should consider a PivotTable analysis if your tabular data also has consistent and repeated values. *Consistent values* means that each column contains the same type of data in the same format. For example, one column contains only customer names, another contains only order dates, and yet another contains only invoice amounts. *Repeated values* means that at least one column contains only a limited number of values that repeat throughout the records. For example, a Region column may contain just four values — East, West, North, and South, for example — that are repeated over hundreds or thousands of records.

Transactional Data

The perfect type of data to benefit from a PivotTable analysis is *transactional* data that records frequent, consistent exchanges of information. Common examples of transactional data include customer orders, accounts receivable data, experiment results, inventory totals, product sales, survey answers, and production schedules. This transactional data creates the same data structure for each record, has consistent data, and has repeated values in at least one field, all which makes this kind of data ideal for a PivotTable approach.

Analysis Required

When deciding whether to build a PivotTable from your data, think about the type of analysis you require. What is your goal? What do you need to know? What secret do you suspect is hidden within all that data? Generally, building a PivotTable is a good idea if you are seeking one or more of the following as part of your analysis of the data: a list of unique values in a field, a summary of a large amount of data, the relationships between two or more fields, and the trend of the data over time.

Unique Values

When faced with a huge amount of data, you may find that one of the first things you want from that data is a list of the unique values in some field. For example, in a database of thousands of orders, you may simply want to know which customers placed orders. The PivotTable is your best choice here because extracting a list of the unique values that occur in a field is one of the things that PivotTables do best.

Summary

Analyzing data often means summarizing it in some way: totaling it, counting it, finding the average or maximum value, and so on. Excel has worksheet functions, subtotals, and other tools for this kind of analysis, but none is suitable for summarizing large amounts of data, particularly if you want to view the results in a compact report. To do that, you must build a PivotTable.

Relationships

One of the biggest problems you face when confronted with a large data set is determining the relationships that exist between one field and another. Which customers are buying which products? How do product defects vary by manufacturing plant? PivotTables are ideal for this kind of analysis because they can break down the values in one field with respect to another. For example, you can display the total sales generated by each of your salespeople, and then break that down by customer, country, product, category, and so on.

Trends

If your data includes a field with date or time values, you may be interested to see how a particular field varies over time. This *trend analysis* can be extremely useful, and Excel has several powerful tools to help you see the trend. However, a PivotTable is an excellent choice if you want to summarize one field and break it down according to the date or time values. How do sales vary throughout the year? How do manufacturing defects vary throughout the day or week?

Reporting Needs

The final aspect to consider when deciding whether to analyze your data with a PivotTable is your reporting needs. In other words, what do you want to end up with? Choose the PivotTable route if you want to end up with a report that is flexible and can easily handle frequent changes.

Flexibility

Build a PivotTable to analyze your data if you want the flexibility to change the report quickly and easily. If you need to switch the layout — for example, to switch from a vertical layout to a horizontal one — you can pivot any field with a click and drag of the mouse. If you need to view subsets of the results, you can filter the report based on the values in a particular field.

Frequent Changes

Choose a PivotTable if you think your underlying data will change frequently. You can easily update the PivotTable to use the latest data, so your report is always accurate and up to date. It is also easy to change the structure of the PivotTable by adding a new field that has been inserted into the data, so you can always incorporate new data.

Explore PivotTable Features

You can get up to speed with PivotTables very quickly after you learn a few key concepts. You need to understand the features that make up a typical PivotTable, particularly the four areas — row, column, data, and page — to which you add fields from your data.

You also need to understand some important PivotTable terminology that you will encounter throughout this book, including terms such as *source data*, *pivot cache*, and *summary calculation*.

A Report Filter

Displays a drop-down list that contains the unique values from a field. When you select a value from the list, Excel filters the PivotTable results to include only the records that match the selected value.

B Column Area

Displays horizontally the unique values from a field in your data.

C Row Area

Displays vertically the unique values from a field in your data.

D Data Area

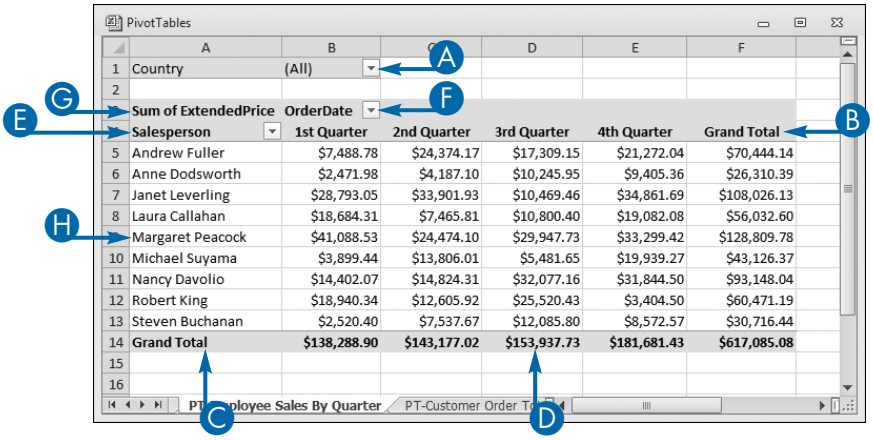
Displays the results of the calculation that Excel applied to a numeric field in your data.

E Row Field Header

Identifies the field contained in the row area. You also use the row field header to filter the field values that appear in the row area.

F Column Field Header

Identifies the field contained in the column area. You also use the column field header to filter the field values that appear in the column area.



	A	B	C	D	E	F
1	Country	(All)				
2						
3	Sum of ExtendedPrice	OrderDate				
4	Salesperson	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter	Grand Total
5	Andrew Fuller	$7,488.78	$24,374.17	$17,309.15	$21,272.04	$70,444.14
6	Anne Dodsworth	$2,471.98	$4,187.10	$10,245.95	$9,405.36	$26,310.39
7	Janet Leverling	$28,793.05	$33,901.93	$10,469.46	$34,861.69	$108,026.13
8	Laura Callahan	$18,684.31	$7,465.81	$10,800.40	$19,082.08	$56,032.60
9	Margaret Peacock	$41,088.53	$24,474.10	$29,947.73	$33,299.42	$128,809.78
10	Michael Suyama	$3,899.44	$13,806.01	$5,481.65	$19,939.27	$43,126.37
11	Nancy Davolio	$14,402.07	$14,824.31	$32,077.16	$31,844.50	$93,148.04
12	Robert King	$18,940.34	$12,605.92	$25,520.43	$3,404.50	$60,471.19
13	Steven Buchanan	$2,520.40	$7,537.67	$12,085.80	$8,572.57	$30,716.44
14	Grand Total	$138,288.90	$143,177.02	$153,937.73	$181,681.43	$617,085.08
15						
16						

G Data Field Header

Specifies both the calculation (such as Sum) and the field (such as Invoice Total) used in the data area.

H Field Items

The unique values for the field added to the particular area.

PivotTable Glossary

PivotTables come with their own terminology, much of which may be unfamiliar to you, even if you have extensive experience with Excel. To learn PivotTables faster, you should understand not only the terms on the previous page, but also the words and phrases that appear in this glossary.

Data

The calculated values that appear within the data area.

Drop Area

A region of the PivotTable onto which you can drop a field from the source data or from another area of the PivotTable. Excel displays each drop area with a blue border.

External Data

Source data that comes from a non-Excel file or database. You can use Microsoft Query to import external data into your Excel worksheet (see Appendix B). Or you can use Excel's other data import tools (see Appendix C).

Labels

The nondata area elements of the PivotTable. The labels include the field buttons, field items, and report filter drop-down list.

Outer Field and Inner Field

When you have multiple fields in the row or column area — see Chapter 3 — Excel places the fields either beside each other in the row area, or one on top of the other in the column area. In either case, the field that is closest to the data area is called the *inner field*, and the field that is farthest from the data area is called the *outer field*.

Pivot

To move a field from one drop area of the PivotTable to another.

Pivot Cache

This is the source data that Excel keeps in memory to improve PivotTable performance.

Source Data

The original data from which you built your PivotTable. The source data can be an Excel range or table, an Access table or query, a Word table, a text file, a Web page, an XML file, SQL Server data, or OLAP server data, among others.

Summary Calculation

The mathematical operation that Excel applies to the values in a numeric field to yield the summary that appears in the data area. Excel offers 11 built-in summary calculations: Sum, Count, Average, Maximum, Minimum, Product, Count Numbers, Standard Deviation (sample), Standard Deviation (population), Variance (sample), and Variance (population). For more, see Chapter 8. You can also create custom calculations (see Chapter 9).

Introducing the PivotChart

When you begin the process of building a PivotTable, Excel actually gives you a choice between building a PivotTable or a PivotChart. In basic terms, a PivotChart is to a PivotTable what a regular chart is to a range. That is, the former is a graphical representation of the latter. So the PivotChart enables you to visualize the PivotTable results by displaying the data area values in chart form.

However, it is also possible to say that a PivotChart is to a regular chart what a PivotTable is to a regular range. In other words, the PivotChart goes far beyond the capabilities of a simple chart because the PivotChart comes with most of the same features that make PivotTables so powerful: You can filter the results to see just the data you need, and you can pivot fields from one area of the PivotChart to another to get the layout you want. See Chapter 9 to learn how to create and work with PivotCharts.

PivotChart Concepts

As you might expect, PivotCharts have a number of elements in common with PivotTables, but there are also some key differences. The following items explain these differences and introduce you to some important PivotChart concepts.

Chart Categories (X-Axis)

Like a PivotTable, a PivotChart automatically groups large amounts of data into smaller, more manageable groups. For example, if you have data with a Category field containing values such as Beverages, Condiments, Confections, and so on, if you build your PivotChart using the Category field, the resulting chart will display one chart category (X-axis value) for each unique Category field value. This is the equivalent of a row field in a PivotTable.

Chart Data Series

Also, as with a PivotTable, you can break down your data in terms of a second field. For example, your data may have an Order Date field. If you add that field to the PivotChart, Excel creates one data series for each unique value in that field. This is the equivalent of a Column field in a PivotTable.

Chart Values (Y-Axis)

You can't have a PivotTable without a data field, and the same is true with a PivotChart. When you add a numeric field for the summary calculation, Excel displays the results as chart values (Y-axis). This is the equivalent of a data field in a PivotTable.

Dynamic PivotCharts

Perhaps the biggest difference between a PivotChart and a regular chart is that each PivotChart is a dynamic object that you can reconfigure as needed, just like a PivotTable. You can pivot fields from one area of the chart to another, you can add fields to different chart areas, and you can place multiple fields in any chart area.

Filtering

Like a PivotTable, you can use the unique values in another field to filter the results that appear in the PivotChart. For example, if your source data has a Country field, you could add it to the PivotChart and use it to filter the chart results to show just those from a specific country. This is the equivalent of a page field in a PivotTable.

Pros and Cons

PivotCharts have advantages and disadvantages, and understanding their strengths and weaknesses will help you decide when and if you should use them. On the positive side, a PivotChart is a powerful data analysis tool because it combines the strengths of Excel's charting capabilities — including most of the options available with regular charts — with the features of a PivotTable. Also, creating a basic PivotChart is just as easy as creating a PivotTable. In fact, if you already have a PivotTable, you can create the equivalent PivotChart with just a couple mouse clicks.

On the negative side, PivotCharts share the same caveats that come with regular charts, particularly the fact that if you do not choose the proper chart type or layout, your data will not be easily understood. Moreover, a PivotChart can quickly become extremely confusing when you have multiple category fields or data series fields. Finally, PivotCharts have inherent limitations that restrict the options and formatting you can apply. For more on PivotChart limitations, see Chapter 9.

PivotChart Features

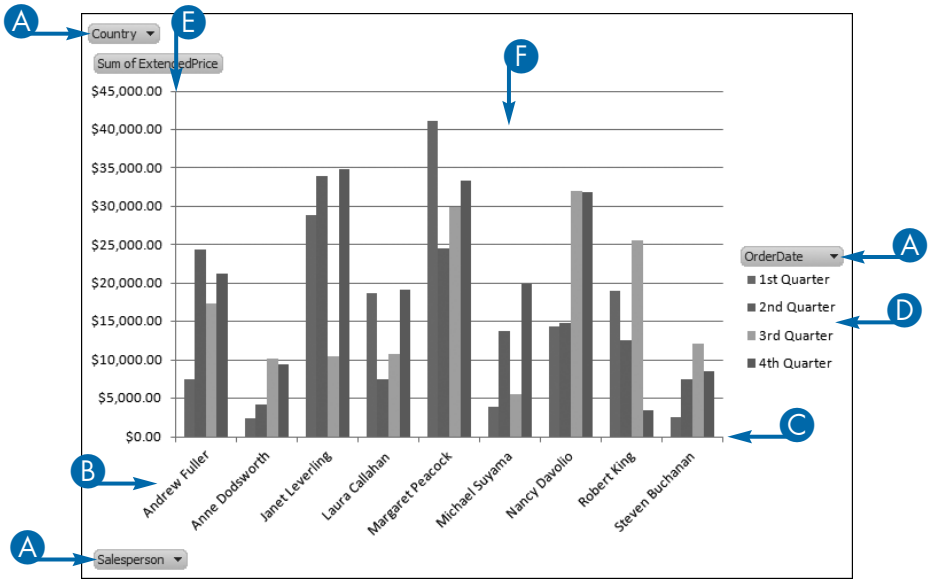
PivotCharts carry over some of the same terminology that you saw earlier for PivotTables, including the concepts of the *report filter*, *data area*, and *field button*. However, PivotCharts also use a number of unique terms such as *category axis* and *series axis* that you need to understand to get the most out of PivotCharts.

A Field Buttons

Displays a drop-down list with unique values from a category field, data series field, or report filter field that you use to filter the PivotChart data.

B Category Items

The unique values from a field that define the chart's categories.



C Category Axis

The chart axis (X-axis) that displays the category items.

D Data Series Items

The unique values from a field that define the chart's data series. The item names appear in the chart legend.

E Series Axis

The chart axis (Y-axis) that displays the values of the data series items.

F Data Area

Displays the charted results of the calculation that Excel applied to a numeric field in your data.

Prepare Your Worksheet Data

The most common method for building a PivotTable is to use data that exists in an Excel worksheet. You can make this task much easier by taking a few minutes to prepare your worksheet data for use in the PivotTable. Ensuring your data is properly prepared will also ensure that your PivotTable contains accurate and complete summaries of the data.

Preparing your worksheet data for use in a PivotTable is not difficult or time consuming. At a minimum, you must ensure that the data is organized in a row-and-column format, with unique headings at the top of each column and accurate and consistent data — all numbers or all text — within each column. You also need to remove blank rows, turn off automatic subtotals, and format the data. In some cases, you may also need to add range names to the data, filter the data, and restructure the data so that worksheet labels appear within a column in the data. You may not need to perform all or even any of these tasks, but you should always ensure that your data is set up according to the guidelines you learn about in this section.

Organize Your Data

In the simplest case, Excel builds a PivotTable from worksheet data by finding the unique values in a specific column of data and *summarizing* — summing or counting — that data based on those unique values. For this to work properly, you need to ensure that your data is organized in such a way that Excel can find those unique values and compute accurate summaries.

Row-and-Column Format

You can perform some Excel tasks on data that is scattered here and there throughout a worksheet, but building a PivotTable is not one of them. To create a PivotTable, your data must be organized in a basic row-and-column format, where each column represents a particular aspect of the data, and each row represents an example of the data. For example, in a parts table, you might have columns for the part name, part number, and cost, and each row would display the name, number, and cost for an individual part.

Unique Column Headings

The first row in your data must contain the headings that identify each column. Excel uses these headings to generate the PivotTable field names, so the headings must be unique and they must reside in a single cell.

Incorporate Labels as Columns

Many worksheets use *labels* — cells that contain descriptive text — as headings to differentiate one section of the worksheet from another. For example, a parts table might have separate sections for each warehouse, and labels such as "East Warehouse" and "West Warehouse" off the side of or above the appropriate section. Unfortunately, this setup prevents you from using the warehouse data as part of the PivotTable — in the page field, for example. To fix this, create a new column with a unique heading, such as "Warehouse," and copy the label value to each row in the section.

East Warehouse

Description	Number	Quantity	Cost	Total Cost	Retail	Gross Margin
Gangley Pliers	D-178	5,700	$10.47	$59,679.00	$ 17.95	71.4%
HCAB Washer	A-201	20,123	$ 0.12	$ 2,414.76	$ 0.25	108.3%
Finley Sprocket	C-098	10,237	$ 1.57	$16,072.09	$ 2.95	87.9%
6" Sonotube	B-111	860	$15.24	$13,106.40	$ 19.95	30.9%

West Warehouse

Description	Number	Quantity	Cost	Total Cost	Retail	Gross Margin
Langstrom 7" Wrench	D-017	755	$18.69	$14,110.95	$ 27.95	49.5%
Thompson Socket	C-321	5,893	$ 3.11	$18,327.23	$ 5.95	91.3%
S-Joint	A-182	3,023	$ 6.85	$20,707.55	$ 9.95	45.3%
LAMF Valve	B-047	6,734	$ 4.01	$27,003.34	$ 6.95	73.3%

Warehouse	Description	Number	Quantity	Cost	Total Cost	Retail	Gross Margin
East	Gangley Pliers	D-178	5,700	$10.47	$59,679.00	$ 17.95	71.4%
East	HCAB Washer	A-201	20,123	$ 0.12	$ 2,414.76	$ 0.25	108.3%
East	Finley Sprocket	C-098	10,237	$ 1.57	$16,072.09	$ 2.95	87.9%
East	6" Sonotube	B-111	860	$15.24	$13,106.40	$ 19.95	30.9%
West	Langstrom 7" Wrench	D-017	755	$18.69	$14,110.95	$ 27.95	49.5%
West	Thompson Socket	C-321	5,893	$ 3.11	$18,327.23	$ 5.95	91.3%
West	S-Joint	A-182	3,023	$ 6.85	$20,707.55	$ 9.95	45.3%
West	LAMF Valve	B-047	6,734	$ 4.01	$27,003.34	$ 6.95	73.3%

Prepare Your Data

To get your data ready for PivotTable analysis, you may also need to run through a few more preparatory chores, including deleting blank rows, ensuring the data is consistent and accurate, and turning off subtotals and the AutoFilter feature.

Ensure Accurate Data

One of the most important concepts in data analysis is that your results are only as accurate as your data. This is sometimes referred to, whimsically, as GIGO: Garbage In, Garbage Out. PivotTables are no exception: You can be sure that the summaries displayed in the report are accurate only if you are sure that the values used in the data field column are accurate. This applies to the other PivotTable fields, as well. For example, if you have a column that is supposed to contain just a certain set of values — for example, North, South, East, and West — you need to check the column to make sure there are no typos or extraneous data items.

Turn Off Automatic Subtotals

Excel PivotTables are designed to provide you with numeric summaries of your data: sums, counts, averages, and so on. Therefore, you do not need to use Excel's Automatic Subtotals feature within your data. In fact, Excel will not create a PivotTable from worksheet data that has subtotals displayed. Therefore, you should remove all subtotals from your data. Click inside the data, click Data➔Subtotal, and then click Remove All.

Delete Blank Rows

It is common to include one or more blank rows within a worksheet to space out the data and to separate different sections of the data. This may make the data easier to read, but it can cause problems when you build your PivotTable because Excel includes the blank rows in the PivotTable report. To avoid this, run through your data and delete any blank rows.

Ensure Consistent Data

It is important that each column contains consistent data. First, ensure that each column contains the same kind of data. For example, if the column is supposed to hold part numbers, make sure it does not contain part names, costs, or anything other than part numbers. Second, ensure that each column uses a consistent data type. For example, in a column of part names, be sure each value is text; in a column of costs, make sure each value is numeric.

Turn Off AutoFilter

If you want to use only a subset of the worksheet data in your PivotTable, do not use Excel's AutoFilter feature. If you do, Excel will still use some or all the hidden rows in the PivotTable report, so your results will not be accurate. Instead, you need to use Excel's Advanced Filter feature and have the results copied to a different worksheet location. You can then use the copied data as the source for your PivotTable report.

Use Repeated Data

The power of the PivotTable lies in its ability to summarize huge amounts of data. That summarization occurs when Excel detects the unique values in a field, groups the records together based on those unique values, and then calculates the total (or whatever) of the values in a particular field. For this to work, at least one field must contain repeated data, preferably a relatively small number of repeated items.

Create a Table for a PivotTable Report

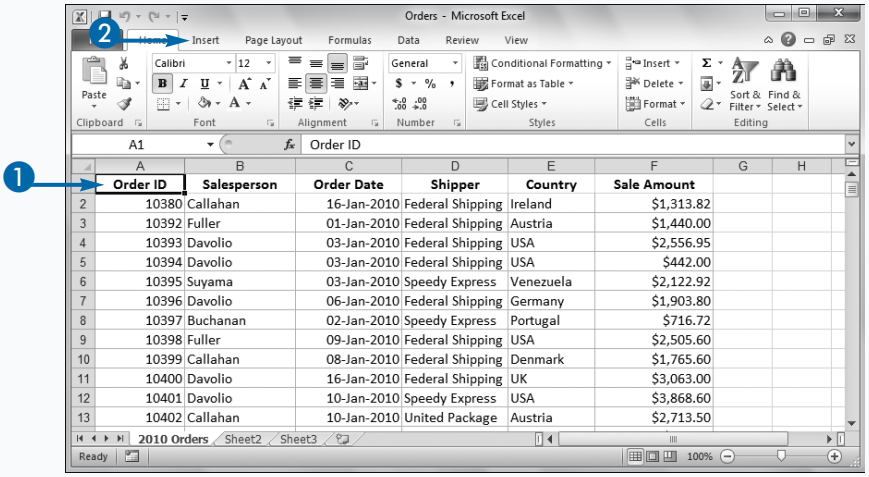
You can make your PivotTable easier to maintain by converting the underlying worksheet data from a regular range to a table. In Excel, a table is a collection of related information with an organizational structure that makes it easy to add, edit, and sort data. In short, a table is a type of database where the data is organized into rows and columns: Each column represents a database field, which is a single type of information, such as a name, address, or phone number; each row represents a database record, which is a collection of associated field values, such as the information for a specific contact. A table differs from a regular Excel range in that Excel offers a set of tools that makes it easier for you to add new records, delete existing records, sort and filter data, and more.

How does a table help you maintain your PivotTables? Using a regular range as the PivotTable source data works well when you insert or delete rows within the range. After the insertions or deletions, you can refresh the PivotTable and Excel automatically updates the report to reflect the changes. However, this does not work if you add new data to the bottom of the range, which is the most common scenario. In this case, you need to rebuild the PivotTable and specify the newly expanded range. You can avoid this extra step by converting your source data range into an Excel table. In this case, Excel keeps track of any new data added to the bottom of the table, so you can refresh your PivotTable at any time.

Create a Table for a PivotTable Report

Note: *This chapter uses the Orders02.xlsx spreadsheet, available at www.wiley.com/go/pivottablesvb2e, or you can create your own sample database.*

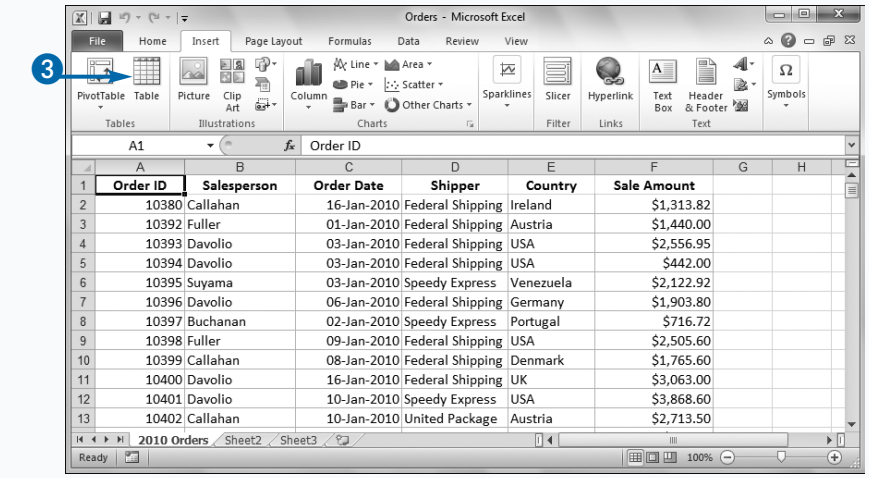
1. Click a cell within the range that you want to convert to a table.
2. Click the Insert tab.



3. Click Table.

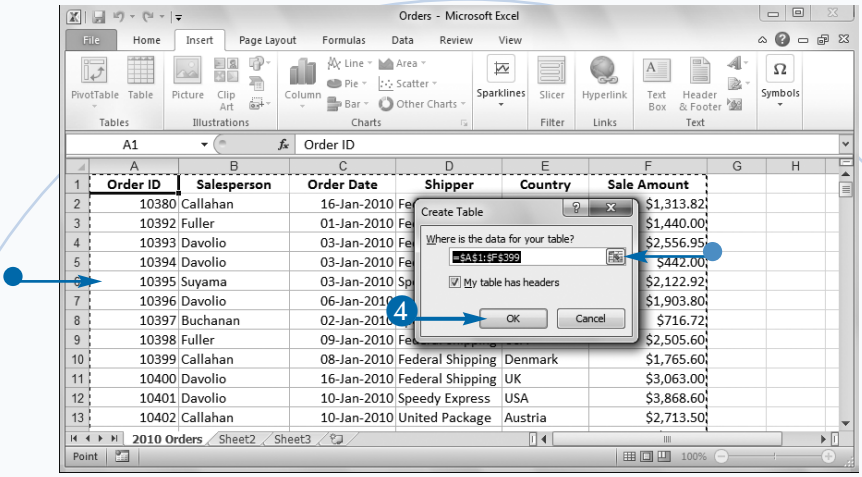
   You can also choose the Table command by pressing Ctrl+T.

   The Create Table dialog box appears.



- Excel selects the range that it will convert to a table.
- If you want to change the range, click here and then click and drag the mouse over the new range.

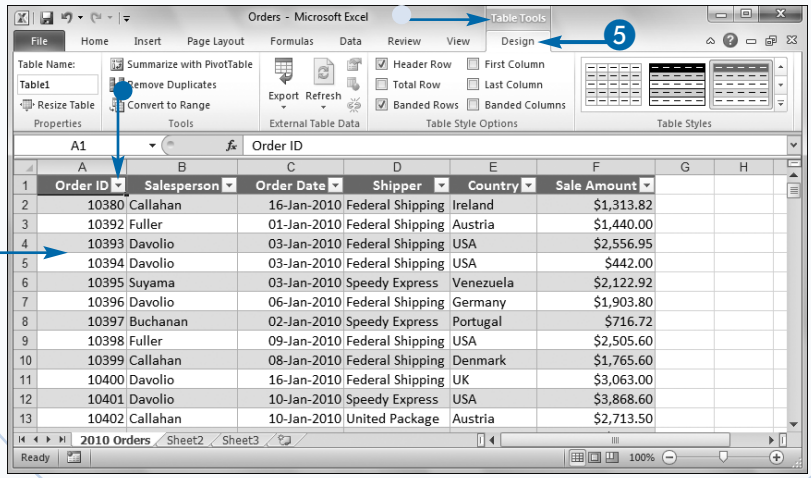
4 Click OK.



Excel converts the range to a table.

- Excel applies a table format to the range.
- The Table Tools contextual tab appears.
- AutoFilter drop-down lists appear in each field heading.

5 Click the Design tab to see Excel's table design tools.



Order ID	Salesperson	Order Date	Shipper	Country	Sale Amount
10380	Callahan	16-Jan-2010	Federal Shipping	Ireland	$1,313.82
10392	Fuller	01-Jan-2010	Federal Shipping	Austria	$1,440.00
10393	Davolio	03-Jan-2010	Federal Shipping	USA	$2,556.95
10394	Davolio	03-Jan-2010	Federal Shipping	USA	$442.00
10395	Suyama	03-Jan-2010	Speedy Express	Venezuela	$2,122.92
10396	Davolio	06-Jan-2010	Federal Shipping	Germany	$1,903.80
10397	Buchanan	02-Jan-2010	Speedy Express	Portugal	$716.72
10398	Fuller	09-Jan-2010	Federal Shipping	USA	$2,505.60
10399	Callahan	08-Jan-2010	Federal Shipping	Denmark	$1,765.60
10400	Davolio	16-Jan-2010	Federal Shipping	UK	$3,063.00
10401	Davolio	10-Jan-2010	Speedy Express	USA	$3,868.60
10402	Callahan	10-Jan-2010	United Package	Austria	$2,713.50

Extra

To add a record to the end of the table, click inside the table, press Ctrl+End to move to the last field in the last record, and then press Tab. To add a record within the table, right-click the record above which you want to insert the new record, and then click Insert→Table Rows Above. Type your data for the first field and then press Tab to move to the second field. For the rest of the fields, type your data and press Tab to move to the next field.

After you create a table, Excel's Design tab offers a number of tools that enable you to format the table. For example, use the Table Styles gallery to change the overall format applied to the table. You can also use the check boxes in the Table Style Options group to turn formatting such as Banded Rows on and off.

Build a PivotTable from an Excel Table

If the data you want to cross-tabulate exists as an Excel table, you can use the Summarize with PivotTable command to easily build a PivotTable report based on your data. You need only specify the location of your source data and then choose the location of the resulting PivotTable.

Excel creates an empty PivotTable in a new worksheet or in the location you specified. Excel also displays PivotTable Field List, which contains four areas with the following labels: Report Filter, Column Labels, Row Labels, and Values. To complete the PivotTable, you must populate some or all of these areas with one or more fields from your data.

When you add a field to the Row Labels, Column Labels, or Report Filter area, Excel extracts the unique values from the field and displays them in the PivotTable in the row, column, or page field, respectively. For example, if you add the Salesperson field to the Row Labels area, Excel updates the PivotTable's row area to display the unique salesperson names as headings that run down the leftmost column of the report. Similarly, if you add the Country field to the Report Filter area, Excel updates the PivotTable's page field to display the unique country names.

When you add a field to the Values area, Excel performs calculations based on the numeric data in the field. The default calculation is sum, so if you add, for example, the Sale Amount field to the Values area, Excel sums the Sale Amount values. How Excel calculates these sums depends on the fields you have added to the other areas. For example, if you add just the Salesperson field to the row area, Excel displays the sum of the Sale Amount values for each salesperson. You can also use other calculations such as Average and Count (see Chapter 8).

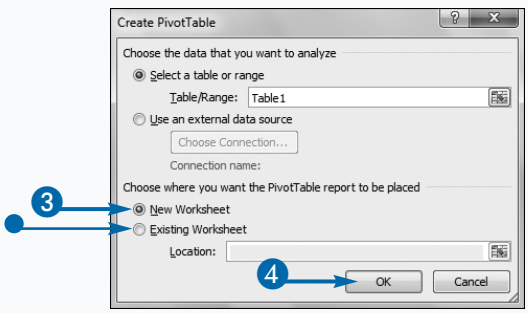
Build a PivotTable from an Excel Table

1. Click a cell within the table that you want to use as the source data.
2. Click Design➔Summarize with PivotTable.

The Create PivotTable dialog box appears.

3. Select the New Worksheet option.

- If you want to place the PivotTable in an existing location, select the Existing Worksheet option, and then use the Location range box to select the worksheet and cell where you want the PivotTable to appear.



4. Click OK.