Chapter 3

Data Visualization

CONTENTS

ANALYTICS IN ACTION

Cincinnati Zoo & Botanical Garden[1]

The Cincinnati Zoo & Botanical Garden, located in Cincinnati, Ohio, is one of the oldest zoos in the United States. To improve decision making by becoming more data-driven, management decided they needed to link the various facets of their business and provide nontechnical managers and executives with an intuitive way to better understand their data. A complicating factor is that when the zoo is busy, managers are expected to be on the grounds interacting with guests, checking on operations, and dealing with issues as they arise or anticipating them. Therefore, being able to monitor what is happening in real time was a key factor in deciding what to do. Zoo management concluded that a data-visualization strategy was needed to address the problem.

Because of its ease of use, real-time updating capability, and iPad compatibility, the Cincinnati Zoo decided to implement its data-visualization strategy using IBM's Cognos advanced data-visualization software. Using this software, the Cincinnati Zoo developed the set of charts shown in Figure 3.1 (known as a data dashboard) to enable management to track the following key measures of performance:

- Item analysis (sales volumes and sales dollars by location within the zoo)
- Geoanalytics (using maps and displays of where the day's visitors are spending their time at the zoo)
- Customer spending

[1]The authors are indebted to John Lucas of the Cincinnati Zoo & Botanical Garden for providing this application.

FIGURE 3.1 Data Dashboard for the Cincinnati Zoo

- Cashier sales performance
- Sales and attendance data versus weather patterns
- Performance of the zoo's loyalty rewards program

An iPad mobile application was also developed to enable the zoo's managers to be out on the grounds and still see and anticipate occurrences in real time. The Cincinnati Zoo's iPad application, shown in Figure 3.2, provides managers with access to the following information:

- Real-time attendance data, including what types of guests are coming to the zoo (members, nonmembers, school groups, and so on)
- Real-time analysis showing which locations are busiest and which items are selling the fastest inside the zoo
- Real-time geographical representation of where the zoo's visitors live

Having access to the data shown in Figures 3.1 and 3.2 allows the zoo managers to make better decisions about staffing levels, which items to stock based on weather and other conditions, and how to better target advertising based on geodemographics.

The impact that data visualization has had on the zoo has been substantial. Within the first year of use, the system was directly responsible for revenue growth of over $500,000, increased visitation to the zoo, enhanced customer service, and reduced marketing costs.

FIGURE 3.2 The Cincinnati Zoo iPad Data Dashboard

The first step in trying to interpret data is often to visualize it in some way. Data visualization can be as simple as creating a summary table, or it could require generating charts to help interpret, analyze, and learn from the data. Data visualization is very helpful for identifying data errors and for reducing the size of your data set by highlighting important relationships and trends.

Data visualization is also important in conveying your analysis to others. Although business analytics is about making better decisions, in many cases, the ultimate decision maker is not the person who analyzes the data. Therefore, the person analyzing the data has to make the analysis simple for others to understand. Proper data-visualization techniques greatly improve the ability of the decision maker to interpret the analysis easily.

In this chapter we discuss some general concepts related to data visualization to help you analyze data and convey your analysis to others. We cover specifics dealing with how to design tables and charts, as well as the most commonly used charts, and present an overview of some more advanced charts. We also introduce the concept of data dashboards and geographic information systems (GISs). Our detailed examples use Excel to generate tables and charts, and we discuss several software packages that can be used for advanced data visualization.

The chapter appendix available in the MindTap Reader covers the use of Analytic Solver (and Excel Add-in) for data visualization.

3.1 Overview of Data Visualization

Decades of research studies in psychology and other fields show that the human mind can process visual images such as charts much faster than it can interpret rows of numbers. However, these same studies also show that the human mind has certain limitations in its ability to interpret visual images and that some images are better at conveying information than others. The goal of this chapter is to introduce some of the most common forms of visualizing data and demonstrate when each form is appropriate.

Microsoft Excel is a ubiquitous tool used in business for basic data visualization. Software tools such as Excel make it easy for anyone to create many standard examples of data visualization. However, as discussed in this chapter, the default settings for tables and charts created with Excel can be altered to increase clarity. New types of software that are dedicated to data visualization have appeared recently. We focus our techniques on Excel in this chapter, but we also mention some of these more advanced software packages for specific data-visualization uses.

Effective Design Techniques

One of the most helpful ideas for creating effective tables and charts for data visualization is the idea of the **data-ink ratio**, first described by Edward R. Tufte in 2001 in his book *The Visual Display of Quantitative Information*. The data-ink ratio measures the proportion of what Tufte terms "data-ink" to the total amount of ink used in a table or chart. Data-ink is the ink used in a table or chart that is necessary to convey the meaning of the data to the audience. Non-data-ink is ink used in a table or chart that serves no useful purpose in conveying the data to the audience.

Let us consider the case of Gossamer Industries, a firm that produces fine silk clothing products. Gossamer is interested in tracking the sales of one of its most popular items, a particular style of women's scarf. Table 3.1 and Figure 3.3 provide examples of a table and chart with low data-ink ratios used to display sales of this style of women's scarf. The data used in this table and figure represent product sales by day. Both of these examples are similar to tables and charts generated with Excel using common default settings. In Table 3.1, most of the grid lines serve no useful purpose. Likewise, in Figure 3.3, the horizontal lines in the chart also add little additional information. In both cases, most of these lines can be deleted without reducing the information conveyed. However, an important piece of information is missing from Figure 3.3: labels for axes. Axes should always be labeled in a chart unless both the meaning and unit of measure are obvious.

TABLE 3.1 Example of a Low Data-Ink Ratio Table

Scarf Sales by Day			
Day	Sales	Day	Sales
1	150	11	170
2	170	12	160
3	140	13	290
4	150	14	200
5	180	15	210
6	180	16	110
7	210	17	90
8	230	18	140
9	140	19	150
10	200	20	230

Table 3.2 shows a modified table in which all grid lines have been deleted except for those around the title of the table. Deleting the grid lines in Table 3.1 increases the data-ink ratio because a larger proportion of the ink used in the table is used to convey the information (the actual numbers). Similarly, deleting the unnecessary horizontal lines in Figure 3.4 increases the data-ink ratio. Note that deleting these horizontal lines and removing (or reducing the size of) the markers at each data point can make it more difficult to determine the exact values plotted in the chart. However, as we discuss later, a simple chart is not the most effective way of presenting data when the audience needs to know exact values; in these cases, it is better to use a table.

In many cases, white space in a table or a chart can improve readability. This principle is similar to the idea of increasing the data-ink ratio. Consider Table 3.2 and Figure 3.4. Removing the unnecessary lines has increased the "white space," making it easier to read both the table and the chart. The fundamental idea in creating effective tables and charts is to make them as simple as possible in conveying information to the reader.

FIGURE 3.3 Example of a Low Data-Ink Ratio Chart

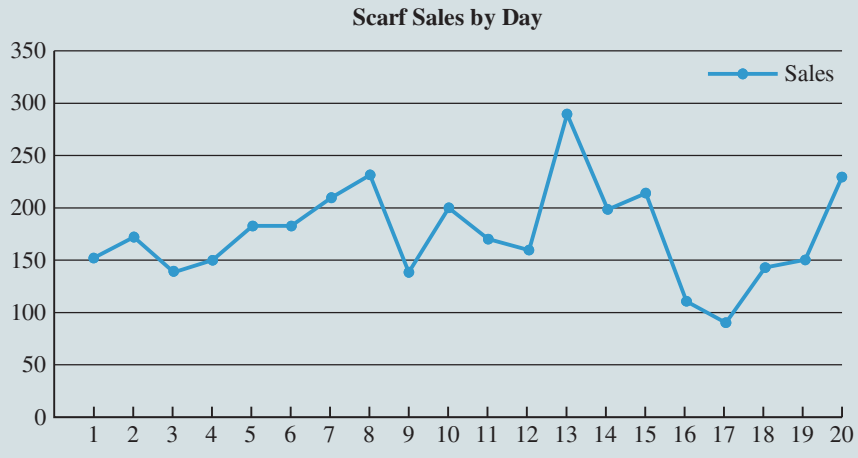
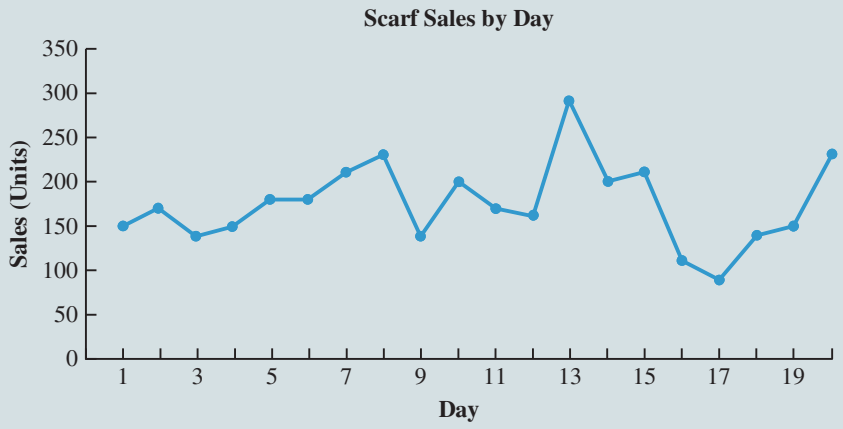


TABLE 3.2 Increasing the Data-Ink Ratio by Removing Unnecessary Gridlines

Scarf Sales by Day

Day	Sales	Day	Sales
1	150	11	170
2	170	12	160
3	140	13	290
4	150	14	200
5	180	15	210
6	180	16	110
7	210	17	90
8	230	18	140
9	140	19	150
10	200	20	230

FIGURE 3.4 Increasing the Data-Ink Ratio by Adding Labels to Axes and Removing Unnecessary Lines and Labels



NOTES + COMMENTS

1. Tables have been used to display data for more than a thousand years. However, charts are much more recent inventions. The famous 17th-century French mathematician, René Descartes, is credited with inventing the now familiar graph with horizontal and vertical axes. William Playfair invented bar charts, line charts, and pie charts in the late 18th century, all of which we will discuss in this chapter. More recently, individuals such as William Cleveland, Edward R. Tufte, and Stephen Few have introduced design techniques for both clarity and beauty in data visualization.
2. Many of the default settings in Excel are not ideal for displaying data using tables and charts that communicate effectively. Before presenting Excel-generated tables and charts to others, it is worth the effort to remove unnecessary lines and labels.

TABLE 3.3 Table Showing Exact Values for Costs and Revenues by Month for Gossamer Industries

	Month						
	1	2	3	4	5	6	Total
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues ($)	64,124	66,128	67,125	48,178	51,785	55,687	353,027

3.2 Tables

The first decision in displaying data is whether a table or a chart will be more effective. In general, charts can often convey information faster and easier to readers, but in some cases a table is more appropriate. Tables should be used when the

1. reader needs to refer to specific numerical values.
2. reader needs to make precise comparisons between different values and not just relative comparisons.
3. values being displayed have different units or very different magnitudes.

When the accounting department of Gossamer Industries is summarizing the company's annual data for completion of its federal tax forms, the specific numbers corresponding to revenues and expenses are important and not just the relative values. Therefore, these data should be presented in a table similar to Table 3.3.

Similarly, if it is important to know by exactly how much revenues exceed expenses each month, then this would also be better presented as a table rather than as a line chart as seen in Figure 3.5. Notice that it is very difficult to determine the monthly revenues and costs in Figure 3.5. We could add these values using data labels, but they would clutter the figure. The preferred solution is to combine the chart with the table into a single figure, as in Figure 3.6, to allow the reader to easily see the monthly changes in revenues and costs while also being able to refer to the exact numerical values.

Now suppose that you wish to display data on revenues, costs, and head count for each month. Costs and revenues are measured in dollars, but head count is measured in number of employees. Although all these values can be displayed on a line chart using multiple

FIGURE 3.5 Line Chart of Monthly Costs and Revenues at Gossamer Industries

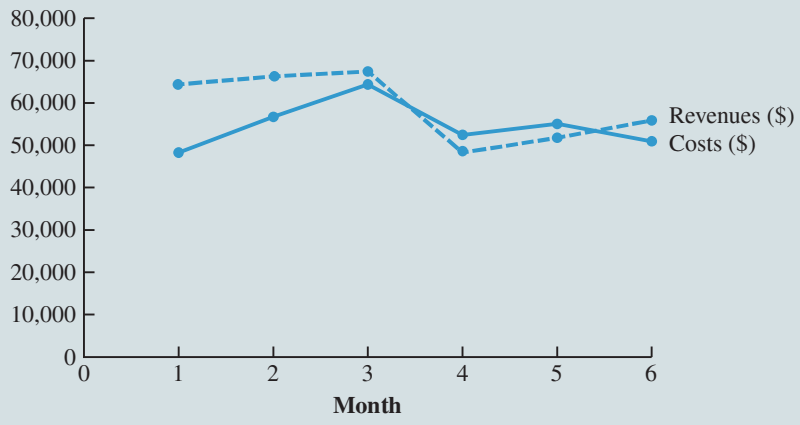
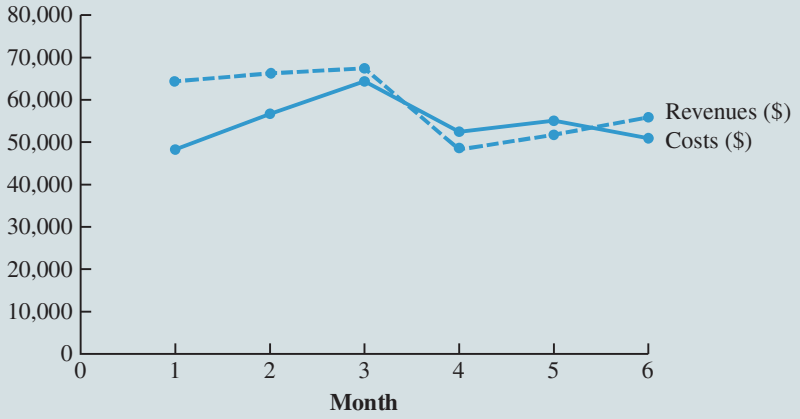


FIGURE 3.6 Combined Line Chart and Table for Monthly Costs and Revenues at Gossamer Industries



	Month						
	1	2	3	4	5	6	Total
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues ($)	64,124	66,128	67,125	48,178	51,785	55,687	353,027

vertical axes, this is generally not recommended. Because the values have widely different magnitudes (costs and revenues are in the tens of thousands, whereas head count is approximately 10 each month), it would be difficult to interpret changes on a single chart. Therefore, a table similar to Table 3.4 is recommended.

Table Design Principles

In designing an effective table, keep in mind the data-ink ratio and avoid the use of unnecessary ink in tables. In general, this means that we should avoid using vertical lines in a table unless they are necessary for clarity. Horizontal lines are generally necessary only for separating column titles from data values or when indicating that a calculation has taken place. Consider Figure 3.7, which compares several forms of a table displaying Gossamer's costs and revenue data. Most people find Design D, with the fewest grid lines, easiest to read. In this table, grid lines are used only to separate the column headings from the data and to indicate that a calculation has occurred to generate the Profits row and the Total column.

In large tables, vertical lines or light shading can be useful to help the reader differentiate the columns and rows. Table 3.5 breaks out the revenue data by location for nine cities

TABLE 3.4 Table Displaying Head Count, Costs, and Revenues at Gossamer Industries

	Month						
	1	2	3	4	5	6	Total
Head count	8	9	10	9	9	9	
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues ($)	64,124	66,128	67,125	48,178	51,785	55,687	353,027

FIGURE 3.7 Comparing Different Table Designs

Design A:

	Month						
	1	2	3	4	5	6	Total
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues ($)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits ($)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design B:

	Month						
	1	2	3	4	5	6	Total
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues ($)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits ($)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design C:

	Month						
	1	2	3	4	5	6	Total
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues ($)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits ($)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design D:

	Month						
	1	2	3	4	5	6	Total
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues ($)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits ($)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

and shows 12 months of revenue and cost data. In Table 3.5, every other column has been lightly shaded. This helps the reader quickly scan the table to see which values correspond with each month. The horizontal line between the revenue for Academy and the Total row helps the reader differentiate the revenue data for each location and indicates that a calculation has taken place to generate the totals by month. If one wanted to highlight the differences among locations, the shading could be done for every other row instead of every other column.

We depart from these guidelines in some figures and tables in this textbook to more closely match Excel's output.

Notice also the alignment of the text and numbers in Table 3.5. Columns of numerical values in a table should be right-aligned; that is, the final digit of each number should be aligned in the column. This makes it easy to see differences in the magnitude of values. If you are showing digits to the right of the decimal point, all values should include the same number of digits to the right of the decimal. Also, use only the number of digits that are necessary to convey the meaning in comparing the values; there is no need to include additional digits if they are not meaningful for comparisons. In many business applications, we report financial values, in which case we often round to the nearest dollar or include two digits to the right of the decimal if such precision is necessary. Additional digits to the right of the decimal are usually unnecessary. For extremely large numbers, we may prefer to display data rounded to the nearest thousand, ten thousand, or even million. For instance, if we need to include, say, $3,457,982 and $10,124,390 in a table when exact dollar values are not necessary, we could write these as 3,458 and 10,124 and indicate that all values in the table are in units of $1,000.

It is generally best to left-align text values within a column in a table, as in the Revenues by Location (the first) column of Table 3.5. In some cases, you may prefer to center text, but you should do this only if the text values are all approximately the same length. Otherwise, aligning the first letter of each data entry promotes readability. Column headings should either match the alignment of the data in the columns or be centered over the values, as in Table 3.5.

Crosstabulation

Types of data such as categorical and quantitative are discussed in Chapter 2.

A useful type of table for describing data of two variables is a **crosstabulation**, which provides a tabular summary of data for two variables. To illustrate, consider the following application based on data from Zagat's Restaurant Review. Data on the quality rating, meal price, and the usual wait time for a table during peak hours were collected for a sample of 300 Los Angeles area restaurants. Table 3.6 shows the data for the first 10 restaurants.

TABLE 3.5 Larger Table Showing Revenues by Location for 12 Months of Data

	Month												
Revenues by Location ($)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
Temple	8,987	8,595	8,958	6,718	8,066	8,574	8,701	9,490	9,610	9,262	9,875	11,058	107,895
Killeen	8,212	9,143	8,714	6,869	8,150	8,891	8,766	9,193	9,603	10,374	10,456	10,982	109,353
Waco	11,603	12,063	11,173	9,622	8,912	9,553	11,943	12,947	12,925	14,050	14,300	13,877	142,967
Belton	7,671	7,617	7,896	6,899	7,877	6,621	7,765	7,720	7,824	7,938	7,943	7,047	90,819
Granger	7,642	7,744	7,836	5,833	6,002	6,728	7,848	7,717	7,646	7,620	7,728	8,013	88,357
Harker Heights	5,257	5,326	4,998	4,304	4,106	4,980	5,084	5,061	5,186	5,179	4,955	5,326	59,763
Gatesville	5,316	5,245	5,056	3,317	3,852	4,026	5,135	5,132	5,052	5,271	5,304	5,154	57,859
Lampasas	5,266	5,129	5,022	3,022	3,088	4,289	5,110	5,073	4,978	5,343	4,984	5,315	56,620
Academy	4,170	5,266	7,472	1,594	1,732	2,025	8,772	1,956	3,304	3,090	3,579	2,487	45,446
Total	64,124	66,128	67,125	48,178	51,785	55,687	69,125	64,288	66,128	68,128	69,125	69,258	759,079
Costs ($)	48,123	56,458	64,125	52,158	54,718	50,985	57,898	62,050	65,215	61,819	67,828	69,558	710,935

TABLE 3.6 Quality Rating and Meal Price for 300 Los Angeles Restaurants

DATA file
Restaurant

Restaurant	Quality Rating	Meal Price ($)	Wait Time (min)
1	Good	18	5
2	Very Good	22	6
3	Good	28	1
4	Excellent	38	74
5	Very Good	33	6
6	Good	28	5
7	Very Good	19	11
8	Very Good	11	9
9	Very Good	23	13
10	Good	13	1

Quality ratings are an example of categorical data, and meal prices are an example of quantitative data.

For now, we will limit our consideration to the quality-rating and meal-price variables. A crosstabulation of the data for quality rating and meal price is shown in Table 3.7. The left and top margin labels define the classes for the two variables. In the left margin, the row labels (Good, Very Good, and Excellent) correspond to the three classes of the quality-rating variable. In the top margin, the column labels ($10–19, $20–29, $30–39, and $40–49) correspond to the four classes (or bins) of the meal-price variable. Each restaurant in the sample provides a quality rating and a meal price. Thus, each restaurant in the sample is associated with a cell appearing in one of the rows and one of the columns of the crosstabulation. For example, restaurant 5 is identified as having a very good quality rating and a meal price of $33. This restaurant belongs to the cell in row 2 and column 3. In constructing a crosstabulation, we simply count the number of restaurants that belong to each of the cells in the crosstabulation.

Table 3.7 shows that the greatest number of restaurants in the sample (64) have a very good rating and a meal price in the $20–29 range. Only two restaurants have an excellent rating and a meal price in the $10–19 range. Similar interpretations of the other frequencies can be made. In addition, note that the right and bottom margins of the crosstabulation give the frequencies of quality rating and meal price separately. From the right margin, we see that data on quality ratings show 84 good restaurants, 150 very good restaurants, and 66 excellent restaurants. Similarly, the bottom margin shows the counts for the meal price variable. The value of 300 in the bottom-right corner of the table indicates that 300 restaurants were included in this data set.

TABLE 3.7 Crosstabulation of Quality Rating and Meal Price for 300 Los Angeles Restaurants

	Meal Price				
Quality Rating	$10–19	$20–29	$30–39	$40–49	Total
Good	42	40	2	0	84
Very Good	34	64	46	6	150
Excellent	2	14	28	22	66
Total	78	118	76	28	300

PivotTables in Excel

A crosstabulation in Microsoft Excel is known as a **PivotTable**. We will first look at a simple example of how Excel's PivotTable is used to create a crosstabulation of the Zagat's restaurant data shown previously. Figure 3.8 illustrates a portion of the data contained in the file *Restaurant*; the data for the 300 restaurants in the sample have been entered into cells B2:D301.

To create a PivotTable in Excel, we follow these steps:

DATA file
Restaurant

Step 1. Click the **Insert** tab on the Ribbon
Step 2. Click **PivotTable** in the **Tables** group
Step 3. When the **Create PivotTable** dialog box appears:
Choose **Select a Table or Range**
Enter *A1:D301* in the **Table/Range:** box
Select **New Worksheet** as the location for the PivotTable Report
Click **OK**

The resulting initial PivotTable Field List and PivotTable Report are shown in Figure 3.9.

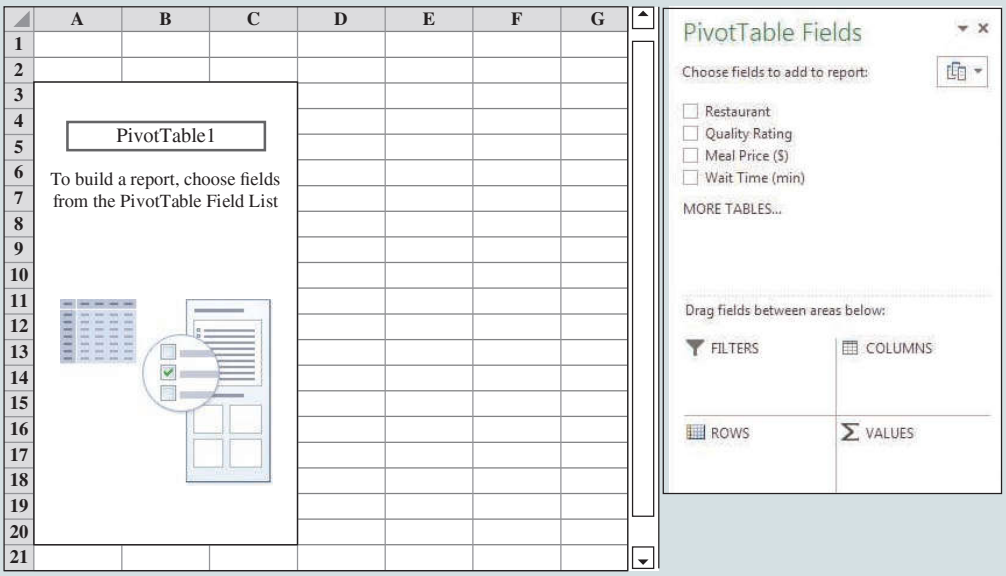
Each of the four columns in Figure 3.8 [Restaurant, Quality Rating, Meal Price ($), and Wait Time (min)] is considered a field by Excel. Fields may be chosen to represent rows, columns, or values in the body of the PivotTable Report. The following steps show how to use Excel's PivotTable Field List to assign the Quality Rating field to the rows, the Meal Price ($) field to the columns, and the Restaurant field to the body of the PivotTable report.

FIGURE 3.8 Excel Worksheet Containing Restaurant Data

DATA file
Restaurant

	A	B	C	D
1	**Restaurant**	**Quality Rating**	**Meal Price ($)**	**Wait Time (min)**
2	1	Good	18	5
3	2	Very Good	22	6
4	3	Good	28	1
5	4	Excellent	38	74
6	5	Very Good	33	6
7	6	Good	28	5
8	7	Very Good	19	11
9	8	Very Good	11	9
10	9	Very Good	23	13
11	10	Good	13	1
12	11	Very Good	33	18
13	12	Very Good	44	7
14	13	Excellent	42	18
15	14	Excellent	34	46
16	15	Good	25	0
17	16	Good	22	3
18	17	Good	26	3
19	18	Excellent	17	36
20	19	Very Good	30	7
21	20	Good	19	3
22	21	Very Good	33	10
23	22	Very Good	22	14
24	23	Excellent	32	27
25	24	Excellent	33	80
26	25	Very Good	34	9

FIGURE 3.9 Initial PivotTable Field List and PivotTable Field Report for the Restaurant Data



Step 4. In the **PivotTable Fields** task pane, go to **Drag fields between areas below:**
Drag the **Quality Rating** field to the **ROWS** area
Drag the **Meal Price ($)** field to the **COLUMNS** area
Drag the **Restaurant** field to the **VALUES** area
Step 5. Click on **Sum of Restaurant** in the **VALUES** area
Step 6. Select **Value Field Settings** from the list of options
Step 7. When the **Value Field Settings** dialog box appears:
Under **Summarize value field by**, select **Count**
Click **OK**

Figure 3.10 shows the completed PivotTable Field List and a portion of the PivotTable worksheet as it now appears.

To complete the PivotTable, we need to group the columns representing meal prices and place the row labels for quality rating in the proper order:

Step 8. Right-click in cell B4 or any cell containing a meal price column label
Step 9. Select **Group** from the list of options
Step 10. When the **Grouping** dialog box appears:
Enter *10* in the **Starting at:** box
Enter *49* in the **Ending at:** box
Enter *10* in the **By:** box
Click **OK**
Step 11. Right-click on "Excellent" in cell A5
Step 12. Select **Move** and click **Move "Excellent" to End**

The final PivotTable, shown in Figure 3.11, provides the same information as the crosstabulation in Table 3.7.

The values in Figure 3.11 can be interpreted as the frequencies of the data. For instance, row 8 provides the frequency distribution for the data over the quantitative variable of meal price. Seventy-eight restaurants have meal prices of $10 to $19. Column F provides the frequency distribution for the data over the categorical variable of quality.

FIGURE 3.10 Completed PivotTable Field List and a Portion of the PivotTable Report for the Restaurant Data (Columns H:AK Are Hidden)

	A	B	C	D	E	F	G	AL	AM	AN	AO	AP	AQ	AR
1														
2														
3	**Count of Restaurant**	**Columns Labels**												
4	**Row Labels**		**10**	**11**	**12**	**13**	**14**	**15**	**47**	**48**	**Grand Total**				
5	Excellent					1			2	2	66				
6	Good		6	4	3	3	2	4			84				
7	Very Good		1	4	3	5	6	1		1	150				
8	**Grand Total**		**7**	**8**	**6**	**9**	**8**	**5**	**2**	**3**	**300**				
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															

FIGURE 3.11 Final PivotTable Report for the Restaurant Data

	A	B	C	D	E	F	G	H	I
1									
2									
3	**Count of Restaurant**	**Column Labels**							
4	**Row Labels**	**10–19**	**20–29**	**30–39**	**40–49**	**Grand Total**			
5	Good	42	40	2		84			
6	Very Good	34	64	46	6	150			
7	Excellent	2	14	28	22	66			
8	**Grand Total**	**78**	**118**	**76**	**28**	**300**			
9									
10									
11									
12									
13									
16									
15									
16									
17									
18									
19									
20									
21									

A total of 150 restaurants have a quality rating of Very Good. We can also use a PivotTable to create percent frequency distributions, as shown in the following steps:

Step 1. To invoke the **PivotTable Fields** task pane, select any cell in the pivot table
Step 2. In the **PivotTable Fields** task pane, click the **Count of Restaurant** in the **VALUES** area
Step 3. Select **Value Field Settings . . .** from the list of options
Step 4. When the **Value Field Settings** dialog box appears, click the tab for **Show Values As**
Step 5. In the **Show values as** area, select **% of Grand Total** from the drop-down menu
Click **OK**

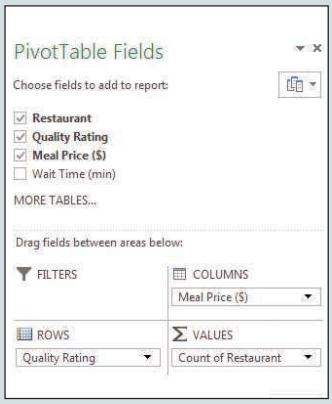
Figure 3.12 displays the percent frequency distribution for the Restaurant data as a PivotTable. The figure indicates that 50% of the restaurants are in the Very Good quality category and that 26% have meal prices between $10 and $19.

PivotTables in Excel are interactive, and they may be used to display statistics other than a simple count of items. As an illustration, we can easily modify the PivotTable in Figure 3.11 to display summary information on wait times instead of meal prices.

Step 1. To invoke the **PivotTable Fields** task pane, select any cell in the pivot table
Step 2. In the **PivotTable Fields** task pane, click the **Count of Restaurant** field in the **VALUES** area
Select **Remove Field**
Step 3. Drag the **Wait Time (min)** to the **VALUES** area
Step 4. Click on **Sum of Wait Time (min)** in the **VALUES** area
Step 5. Select **Value Field Settings…** from the list of options
Step 6. When the **Value Field Settings** dialog box appears:
Under **Summarize value field by**, select **Average**
Click **Number Format**
In the **Category:** area, select **Number**
Enter *1* for **Decimal places:**
Click **OK**
When the **Value Field Settings** dialog box reappears, click **OK**

FIGURE 3.12 Percent Frequency Distribution as a PivotTable for the Restaurant Data

	A	B	C	D	E	F	G
1							
2							
3	**Count of Restaurant**	**Column**					
4	**Row Labels**	**Labels 10–19**	**20–29**	**30–39**	**40–49**	**Grand Total**	
5	Good	14.00%	13.33%	0.67%	0.00%	28.00%	
6	Very Good	11.33%	21.33%	15.33%	2.00%	50.00%	
7	Excellent	0.67%	4.67%	9.33%	7.33%	22.00%	
8	**Grand Total**	**26.00%**	**39.33%**	**25.33%**	**9.33%**	**100.00%**	
9							
10							
11							
12							
13							
16							
15							
16							
17							
18							



The completed PivotTable appears in Figure 3.13. This PivotTable replaces the counts of restaurants with values for the average wait time for a table at a restaurant for each grouping of meal prices ($10–19, $20–29, $30–39, and $40–49). For instance, cell B7 indicates that the average wait time for a table at an Excellent restaurant with a meal price of $10–19 is 25.5 minutes. Column F displays the total average wait times for tables in each quality rating category. We see that Excellent restaurants have the longest average waits of 35.2 minutes and that Good restaurants have average wait times of only 2.5 minutes. Finally, cell D7 shows us that the longest wait times can be expected at Excellent restaurants with meal prices in the $30–39 range (34 minutes).

*You can also filter data in a PivotTable by dragging the field that you want to filter to the **FILTERS** area in the **PivotTable Fields**.*

We can also examine only a portion of the data in a PivotTable using the Filter option in Excel. To Filter data in a PivotTable, click on the **Filter Arrow** next to **Row Labels** or **Column Labels** and then uncheck the values that you want to remove from the PivotTable. For example, we could click on the arrow next to Row Labels and then uncheck the Good value to examine only Very Good and Excellent restaurants.

Recommended PivotTables in Excel

Excel also has the ability to recommend PivotTables for your data set. To illustrate Recommended PivotTables in Excel, we return to the restaurant data in Figure 3.8. To create a Recommended PivotTable, follow the steps below using the file *Restaurant*.

Hovering your pointer over the different options will display the full name of each option, as shown in Figure 3.14.

Step 1. Select any cell in table of data (for example, cell A1)
Step 2. Click the **Insert** tab on the Ribbon
Step 3. Click **Recommended PivotTables** in the **Tables** group
Step 4. When the **Recommended PivotTables** dialog box appears:
Select the **Count of Restaurant, Sum of Wait Time (min), Sum of Meal Price ($) by Quality Rating** option (see Figure 3.14)
Click **OK**

The steps above will create the PivotTable shown in Figure 3.15 on a new Worksheet. The Recommended PivotTables tool in Excel is useful for quickly creating commonly used PivotTables for a data set, but note that it may not give you the option to create the

FIGURE 3.13 PivotTable Report for the Restaurant Data with Average Wait Times Added

	A	B	C	D	E	F	G
1							
2							
3	**Average of Wait Time (min)**	**Column**					
4	**Row Labels**	**Labels 10–19**	**20–29**	**30–39**	**40–49**	**Grand Total**	
5	Good	2.6	2.5	0.5		2.5	
6	Very Good	12.6	12.6	12.0	10.0	12.3	
7	Excellent	25.5	29.1	34.0	32.3	32.1	
8	**Grand Total**	**7.6**	**11.1**	**19.8**	**27.5**	**13.9**	
9							
10							
11							
12							
13							
16							
15							
16							
17							

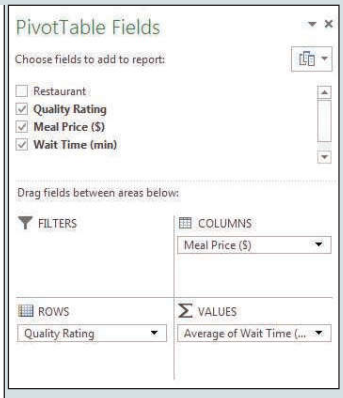


FIGURE 3.14 Recommended PivotTables Dialog Box in Excel

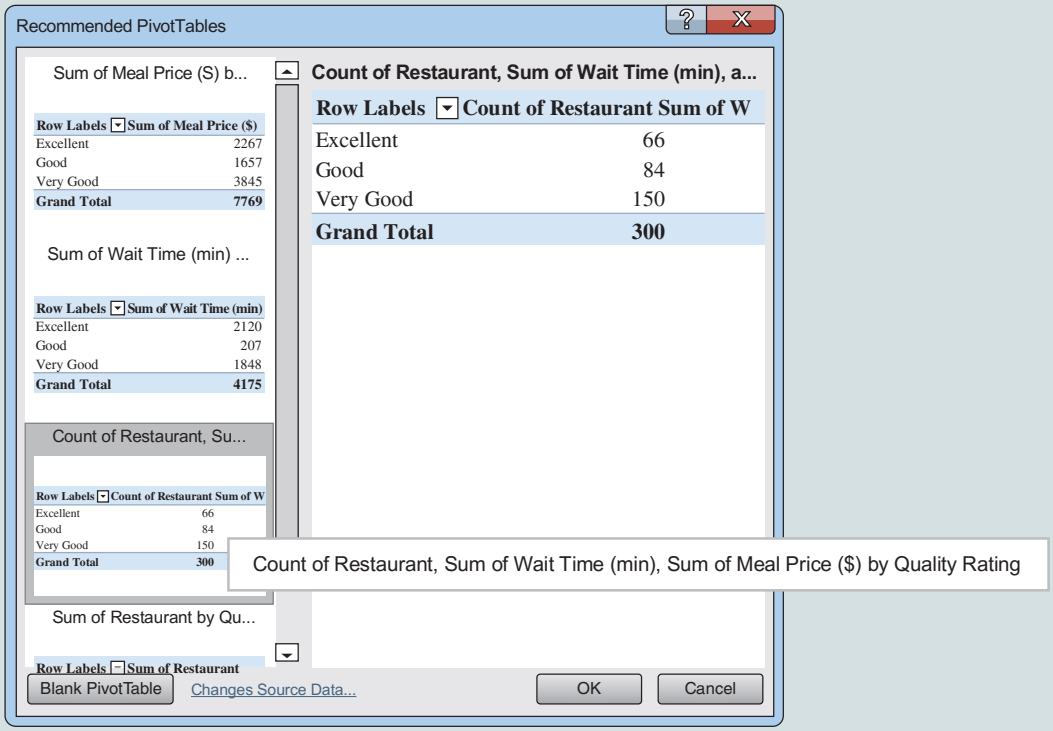
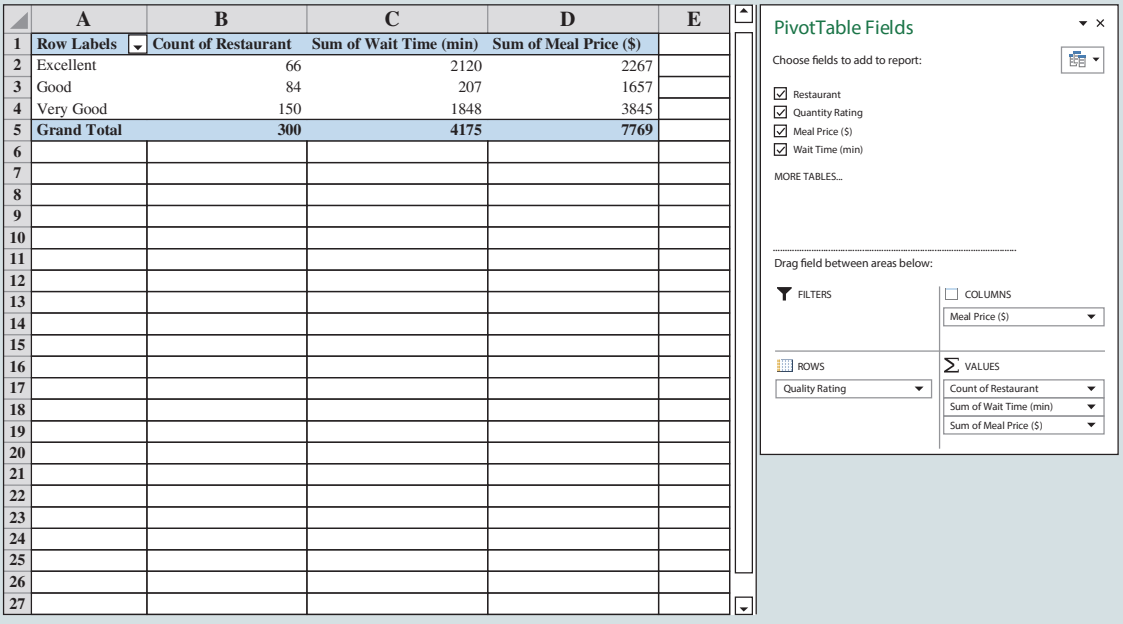
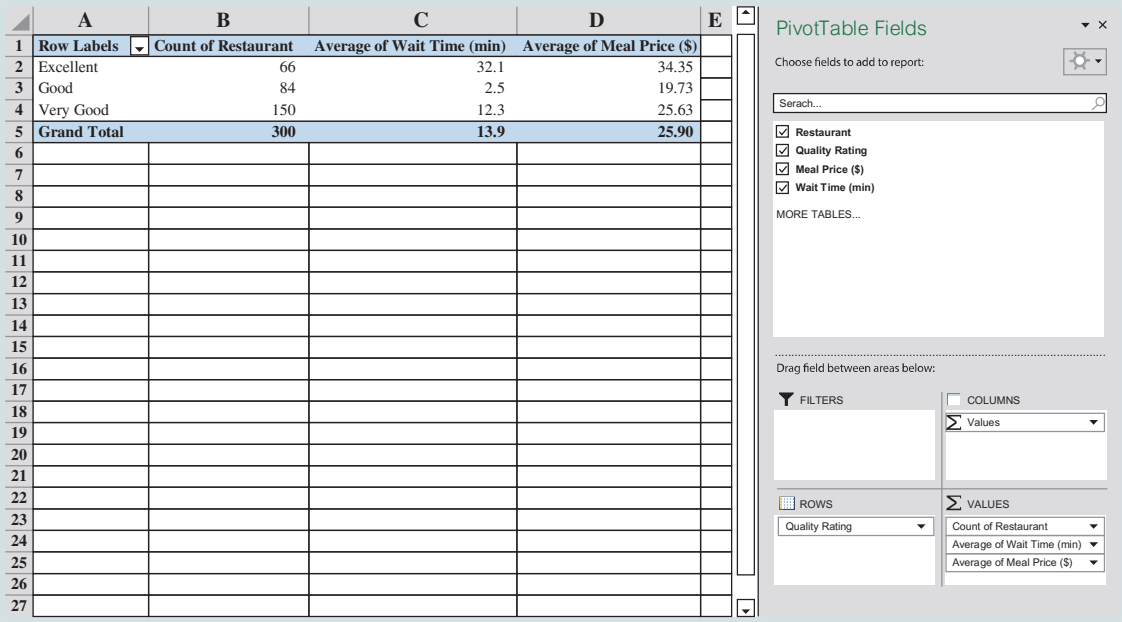


FIGURE 3.15 Default PivotTable Created for Restaurant Data Using Excel's Recommended PivotTables Tool



	A	B	C	D	E
1	Row Labels	Count of Restaurant	Sum of Wait Time (min)	Sum of Meal Price ($)	
2	Excellent	66	2120	2267	
3	Good	84	207	1657	
4	Very Good	150	1848	3845	
5	Grand Total	300	4175	7769	

FIGURE 3.16 Completed PivotTable for Restaurant Data Using Excel's Recommended PivotTables Tool



	A	B	C	D	E
1	Row Labels	Count of Restaurant	Average of Wait Time (min)	Average of Meal Price ($)	
2	Excellent	66	32.1	34.35	
3	Good	84	2.5	19.73	
4	Very Good	150	12.3	25.63	
5	Grand Total	300	13.9	25.90	

exact PivotTable that will be of the most use for your data analysis. Displaying the sum of wait times and the sum of meal prices within each quality-rating category, as shown in Figure 3.15, is not particularly useful here; the average wait times and average meal prices within each quality-rating category would be more useful to us. But we can easily modify the PivotTable in Figure 3.14 to show the average values by selecting any cell in the PivotTable to invoke the **PivotTable Fields** task pane, clicking on **Sum of Wait Time (min)** and then **Sum of Meal Price ($)**, and using the **Value Field Settings…** to change the **Summarize value field by** option to **Average**. The finished PivotTable is shown in Figure 3.16.

3.3 Charts

The appendix for this chapter available in the MindTap Reader demonstrates the use of the Excel Add-in Analytic Solver to create a scatter-chart matrix and a parallel-coordinates plot.

Charts (or graphs) are visual methods for displaying data. In this section, we introduce some of the most commonly used charts to display and analyze data including scatter charts, line charts, and bar charts. Excel is the most commonly used software package for creating simple charts. We explain how to use Excel to create scatter charts, line charts, sparklines, bar charts, bubble charts, and heat maps.

Scatter Charts

A **scatter chart** is a graphical presentation of the relationship between two quantitative variables. As an illustration, consider the advertising/sales relationship for an electronics store in San Francisco. On 10 occasions during the past three months, the store used weekend television commercials to promote sales at its stores. The managers want to investigate whether a relationship exists between the number of commercials shown and sales at the store the following week. Sample data for the 10 weeks, with sales in hundreds of dollars, are shown in Table 3.8.

TABLE 3.8 Sample Data for the San Francisco Electronics Store

Week	No. of Commercials x	Sales ($100s) y
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

DATA file Electronics

We will use the data from Table 3.8 to create a scatter chart using Excel's chart tools and the data in the file *Electronics*:

Hovering the pointer over the chart type buttons in Excel will display the names of the buttons and short descriptions of the types of chart.

Step 1. Select cells B2:C11
Step 2. Click the **Insert** tab in the Ribbon
Step 3. Click the **Insert Scatter (X,Y) or Bubble Chart** button in the **Charts** group
Step 4. When the list of scatter chart subtypes appears, click the **Scatter** button
Step 5. Click the **Design** tab under the **Chart Tools** Ribbon
Step 6. Click **Add Chart Element** in the **Chart Layouts** group
Select **Chart Title**, and click **Above Chart**
Click on the text box above the chart, and replace the text with *Scatter Chart for the San Francisco Electronics Store*
Step 7. Click **Add Chart Element** in the **Chart Layouts** group
Select **Axis Title**, and click **Primary Vertical**
Click on the text box under the horizontal axis, and replace "Axis Title" with *Number of Commercials*

Steps 9 and 10 are optional, but they improve the chart's readability. We would want to retain the gridlines only if they helped the reader to determine more precisely where data points are located relative to certain values on the horizontal and/or vertical axes.

Step 8. Click **Add Chart Element** in the **Chart Layouts** group
Select **Axis Title,** and click **Primary Horizontal**
Click on the text box next to the vertical axis, and replace "Axis Title" with *Sales ($100s)*
Step 9. Right-click on one of the horizontal grid lines in the body of the chart, and click **Delete**
Step 10. Right-click on one of the vertical grid lines in the body of the chart, and click **Delete**

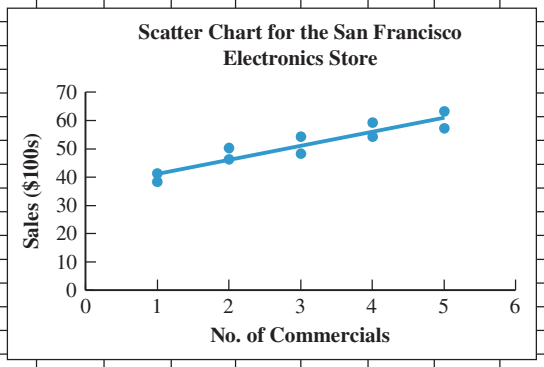
We can also use Excel to add a trendline to the scatter chart. A **trendline** is a line that provides an approximation of the relationship between the variables. To add a linear trendline using Excel, we use the following steps:

Step 1. Right-click on one of the data points in the scatter chart, and select **Add Trendline…**
Step 2. When the **Format Trendline** task pane appears, select **Linear** under **Trendline Options**

Figure 3.17 shows the scatter chart and linear trendline created with Excel for the data in Table 3.8. The number of commercials (x) is shown on the horizontal axis, and sales (y)

FIGURE 3.17 Scatter Chart for the San Francisco Electronics Store

	A	B	C
1	Week	No. of Commercials	Sales Volume
2	1	2	50
3	2	5	57
4	3	1	41
5	4	3	54
6	5	4	54
7	6	1	38
8	7	5	63
9	8	3	48
10	9	4	59
11	10	2	46



are shown on the vertical axis. For week 1, $x = 2$ and $y = 50$. A point is plotted on the scatter chart at those coordinates; similar points are plotted for the other nine weeks. Note that during two of the weeks, one commercial was shown, during two of the weeks, two commercials were shown, and so on.

Scatter charts are often referred to as scatter plots or scatter diagrams.

The completed scatter chart in Figure 3.17 indicates a positive linear relationship (or positive correlation) between the number of commercials and sales: Higher sales are associated with a higher number of commercials. The linear relationship is not perfect because not all of the points are on a straight line. However, the general pattern of the points and the trendline suggest that the overall relationship is positive. This implies that the covariance between sales and commercials is positive and that the correlation coefficient between these two variables is between 0 and +1.

Chapter 2 introduces scatter charts and relates them to the concepts of covariance and correlation.

The **Chart Buttons** in Excel allow users to quickly modify and format charts. Three buttons appear next to a chart whenever you click on a chart to make it active. Clicking on the **Chart Elements** button brings up a list of check boxes to quickly add and remove axes, axis titles, chart titles, data labels, trendlines, and more. Clicking on the **Chart Styles** button allows the user to quickly choose from many preformatted styles to change the look of the chart. Clicking on the **Chart Filter** button allows the user to select the data to be included in the chart. The Chart Filter button is very useful for performing additional data analysis.

Recommended Charts in Excel

Similar to the ability to recommend PivotTables, Excel has the ability to recommend charts for a given data set. The steps below demonstrate the Recommended Charts tool in Excel for the *Electronics* data.

DATA file
Electronics

Step 1. Select cells B2:C11
Step 2: Click the **Insert** tab in the Ribbon
Step 3: Click the **Recommended Charts** button in the **Charts** group
Step 4: When the **Insert Chart** dialog box appears, select the **Scatter** option (see Figure 3.18)
Click **OK**

These steps create the basic scatter chart that can then be formatted (using the Chart Buttons or Chart Tools Ribbon) to create the completed scatter chart shown in Figure 3.17. Note that the Recommended Charts tool gives several possible recommendations for the electronics data in Figure 3.18. These recommendations include scatter charts, line charts, and bar charts, which will be covered later in this chapter. Excel's Recommended Charts tool generally does a good job of interpreting your data and providing recommended charts, but take care to ensure that the selected chart is meaningful and follows good design practice.

Line Charts

A line chart for time series data is often called a time series plot.

Line charts are similar to scatter charts, but a line connects the points in the chart. Line charts are very useful for time series data collected over a period of time (minutes, hours, days, years, etc.). As an example, Kirkland Industries sells air compressors to manufacturing companies. Table 3.9 contains total sales amounts (in $100s) for air compressors during

FIGURE 3.18 Insert Chart Dialog Box from Recommended Charts Tool in Excel

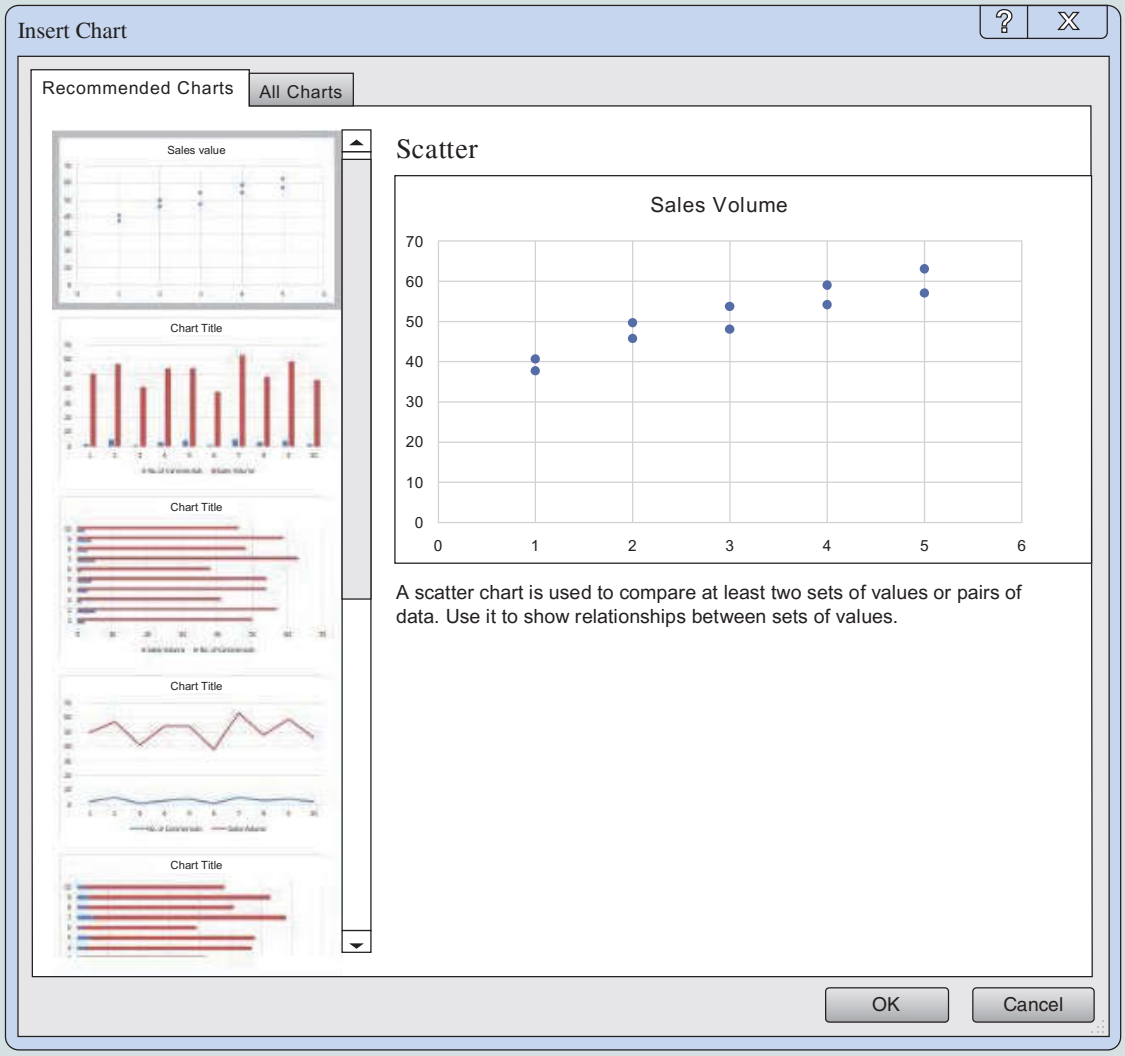


TABLE 3.9 Monthly Sales Data of Air Compressors at Kirkland Industries

Month	Sales ($100s)
Jan	150
Feb	145
Mar	185
Apr	195
May	170
Jun	125
Jul	210
Aug	175
Sep	160
Oct	120
Nov	115
Dec	120

DATA file
Kirkland

each month in the most recent calendar year. Figure 3.19 displays a scatter chart and a line chart created in Excel for these sales data. The line chart connects the points of the scatter chart. The addition of lines between the points suggests continuity, and it is easier for the reader to interpret changes over time.

To create the line chart in Figure 3.19 in Excel, we follow these steps:

Step 1. Select cells A2:B13
Step 2. Click the **Insert** tab on the Ribbon
Step 3. Click the **Insert Line Chart** button in the **Charts** group
Step 4. When the list of line chart subtypes appears, click the **Line with Markers** button under **2-D Line**

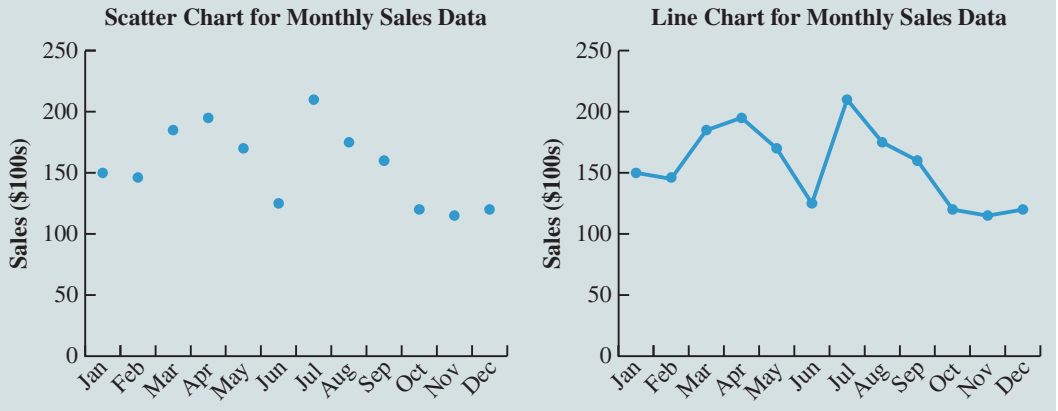
This creates a line chart for sales with a basic layout and minimum formatting

Step 5. Select the line chart that was just created to reveal the **Chart Buttons**

*Because the gridlines do not add any meaningful information here, we do not select the check box for **Gridlines** in **Chart Elements**, as it increases the data-ink ratio.*

In the line chart in Figure 3.19, we have kept the markers at each data point. This is a matter of personal taste, but removing the markers tends to suggest that the data are continuous when in fact we have only one data point per month.

FIGURE 3.19 Scatter Chart and Line Chart for Monthly Sales Data at Kirkland Industries



Step 6. Click the **Chart Elements** button
Select the check boxes for **Axes**, **Axis Titles**, and **Chart Title**. **Deselect** the check box for **Gridlines**.
Click on the text box next to the vertical axis, and replace "Axis Title" with *Sales ($100s)*
Click on the text box next to the horizontal axis and replace "Axis Title" with *Month*
Click on the text box above the chart, and replace "Sales ($100s)" with *Line Chart for Monthly Sales Data*

Figure 3.20 shows the line chart created in Excel along with the selected options for the Chart Elements button.

Line charts can also be used to graph multiple lines. Suppose we want to break out Kirkland's sales data by region (North and South), as shown in Table 3.10. We can create a line chart in Excel that shows sales in both regions, as in Figure 3.21 by following similar steps but selecting cells A2:C14 in the file *KirklandRegional* before creating the line chart. Figure 3.21 shows an interesting pattern. Sales in both the North and the South regions seemed to follow the same increasing/decreasing pattern until October. Starting in October, sales in the North continued to decrease while sales in the South increased. We would probably want to investigate any changes that occurred in the North region around October.

A special type of line chart is a **sparkline**, which is a minimalist type of line chart that can be placed directly into a cell in Excel. Sparklines contain no axes; they display only the line for the data. Sparklines take up very little space, and they can be effectively used to provide information on overall trends for time series data. Figure 3.22 illustrates the use of sparklines in Excel for the regional sales data. To create a sparkline in Excel:

DATA file
KirklandRegional

Step 1. Click the **Insert** tab on the Ribbon
Step 2. Click **Line** in the **Sparklines** group

FIGURE 3.20 Line Chart and Excel's Chart Elements Button Options for Monthly Sales Data at Kirkland Industries

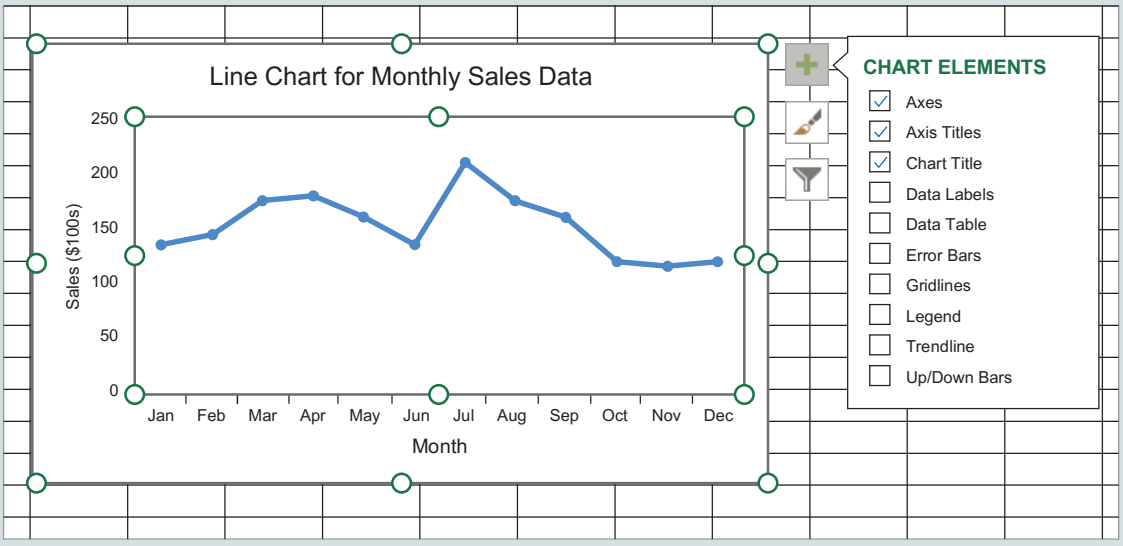


TABLE 3.10 Regional Sales Data by Month for Air Compressors at Kirkland Industries

Month	Sales ($100s) North	Sales ($100s) South
Jan	95	40
Feb	100	45
Mar	120	55
Apr	115	65
May	100	60
Jun	85	50
Jul	135	75
Aug	110	65
Sep	100	60
Oct	50	70
Nov	40	75
Dec	40	80

DATA file
KirklandRegional

Step 3. When the **Create Sparklines** dialog box opens,
Enter *B3:B14* in the **Data Range:** box
Enter *B15* in the **Location Range:** box
Click **OK**

Step 4. Copy cell B15 to cell C15

The sparklines in cells B15 and C15 do not indicate the magnitude of sales in the North and the South regions, but they do show the overall trend for these data. Sales in the North appear to be decreasing and sales in the South increasing overall. Because sparklines are input directly into the cell in Excel, we can also type text directly into the same cell that will then be overlaid on the sparkline, or we can add shading to the cell, which will appear as the background. In Figure 3.22, we have shaded cells B15 and C15 to highlight the sparklines. As can be seen, sparklines provide an efficient and simple way to display basic information about a time series.

In the line chart in Figure 3.21, we have replaced Excel's default legend with text boxes labeling the lines corresponding to sales in the North and the South. This can often make the chart look cleaner and easier to interpret.

FIGURE 3.21 Line Chart of Regional Sales Data at Kirkland Industries

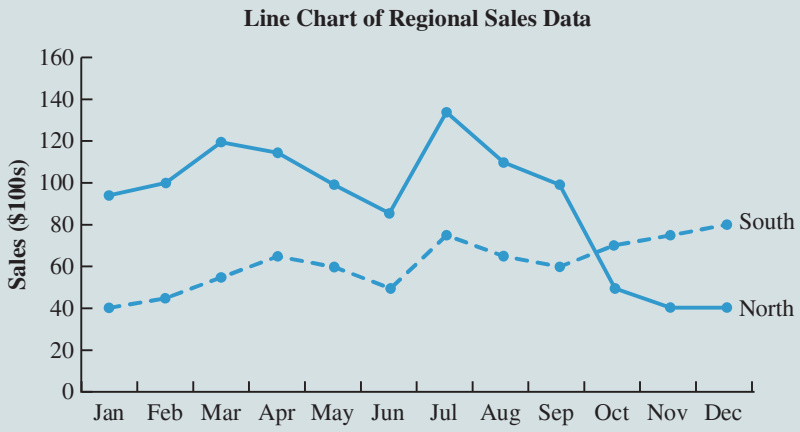
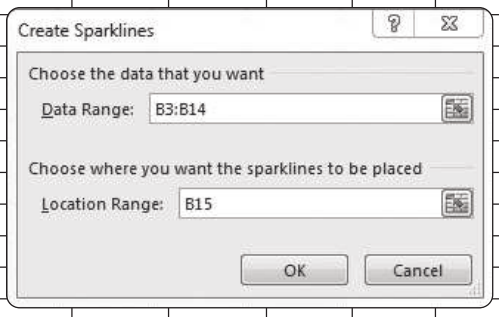


FIGURE 3.22 Sparklines for the Regional Sales Data at Kirkland Industries

	A	B	C	D	E	F	G	H	I
1		Sales ($100s)							
2	Month	North	South						
3	Jan	95	40						
4	Feb	100	45						
5	Mar	120	55						
6	Apr	115	65						
7	May	100	60						
8	Jun	85	50						
9	Jul	135	75						
10	Aug	110	65						
11	Sep	100	60						
12	Oct	50	70						
13	Nov	40	75						
14	Dec	40	80						
15									



In versions of Excel prior to Excel 2016, ***Insert Bar Chart*** *and* ***Insert Column Chart*** *each have separate buttons in the* ***Charts*** *group, but these are combined under the* ***Insert Column or Bar Chart*** *button in Excel 2016.*

Bar Charts and Column Charts

Bar charts and column charts provide a graphical summary of categorical data. **Bar charts** use horizontal bars to display the magnitude of the quantitative variable. **Column charts** use vertical bars to display the magnitude of the quantitative variable. Bar and column charts are very helpful in making comparisons between categorical variables. Consider a regional supervisor who wants to examine the number of accounts being handled by each manager. Figure 3.23 shows a bar chart created in Excel displaying these data. To create this bar chart in Excel:

Step 1. Select cells A2:B9
Step 2. Click the **Insert** tab on the Ribbon
Step 3. Click the **Insert Column or Bar Chart** button in the **Charts** group
Step 4. When the list of bar chart subtypes appears:
Click the **Clustered Bar** button in the **2-D Bar** section

DATA file
AccountsManaged

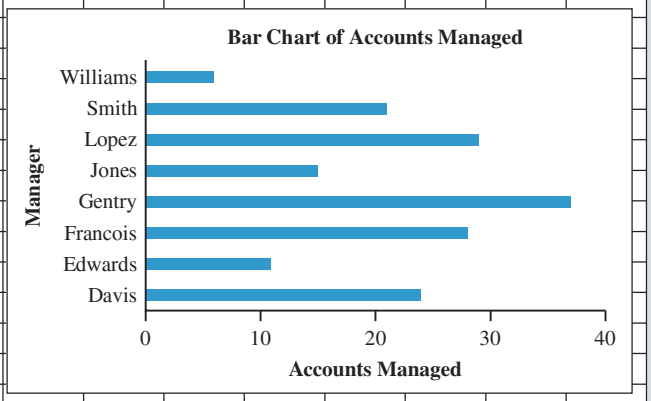
Step 5. Select the bar chart that was just created to reveal the **Chart Buttons**
Step 6. Click the **Chart Elements** button
Select the check boxes for **Axes**, **Axis Titles**, and **Chart Title**. Deselect the check box for **Gridlines**.
Click on the text box next to the vertical axis, and replace "Axis Title" with *Accounts Managed*
Click on the text box next to the vertical axis, and replace "Axis Title" with *Manager*
Click on the text box above the chart, and replace "Chart Title" with *Bar Chart of Accounts Managed*

From Figure 3.23 we can see that Gentry manages the greatest number of accounts and Williams the fewest. We can make this bar chart even easier to read by ordering the results by the number of accounts managed. We can do this with the following steps:

Step 1. Select cells A1:B9
Step 2. Right-click any of the cells A1:B9
Select **Sort**
Click **Custom Sort**

FIGURE 3.23 Bar Chart for Accounts Managed Data

	A	B
1	Manager	Accounts Managed
2	Davis	24
3	Edwards	11
4	Francois	28
5	Gentry	37
6	Jones	15
7	Lopez	29
8	Smith	21
9	Williams	6



Step 3. When the **Sort** dialog box appears:
Make sure that the check box for **My data has headers** is checked
Select **Accounts Managed** in the **Sort by** box under **Column**
Select **Smallest to Largest** under **Order**
Click **OK**

In the completed bar chart in Excel, shown in Figure 3.24, we can easily compare the relative number of accounts managed for all managers. However, note that it is difficult to interpret from the bar chart exactly how many accounts are assigned to each manager. If this information is necessary, these data are better presented as a table or by adding data labels to the bar chart, as in Figure 3.25, which is created in Excel using the following steps:

Alternatively, you can add Data Labels by right-clicking on a bar in the chart and selecting ***Add Data Labels****.*

Step 1. Select the chart to reveal the **Chart Buttons**
Step 2. Click the **Chart Elements** button
Select the check box for **Data Labels**

This adds labels of the number of accounts managed to the end of each bar so that the reader can easily look up exact values displayed in the bar chart.

A Note on Pie Charts and Three-Dimensional Charts

Pie charts are another common form of chart used to compare categorical data. However, many experts argue that pie charts are inferior to bar charts for comparing data. The pie chart in Figure 3.26 displays the data for the number of accounts managed in Figure 3.23. Visually, it is still relatively easy to see that Gentry has the greatest number of accounts and that Williams has the fewest. However, it is difficult to say whether Lopez or Francois has more accounts. Research has shown that people find it very difficult to perceive differences in area. Compare Figure 3.26 to Figure 3.24. Making visual comparisons is much easier in the bar chart than in the pie chart (particularly when using a limited number of colors for differentiation). Therefore, we recommend against using pie charts in most situations and suggest instead using bar charts for comparing categorical data.

FIGURE 3.24 Sorted Bar Chart for Accounts Managed Data

	A	B
1	Manager	Accounts Managed
2	Williams	6
3	Edwards	11
4	Jones	15
5	Smith	21
6	Davis	24
7	Francois	28
8	Lopez	29
9	Gentry	37

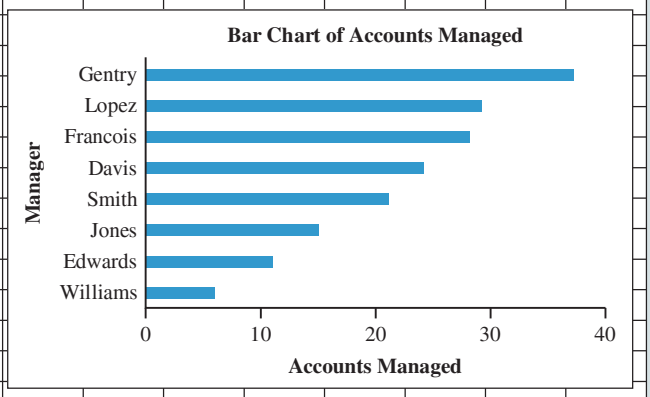


FIGURE 3.25 Bar Chart with Data Labels for Accounts Managed Data

	A	B
1	Manager	Accounts Managed
2	Williams	6
3	Edwards	11
4	Jones	15
5	Smith	21
6	Davis	24
7	Francois	28
8	Lopez	29
9	Gentry	37

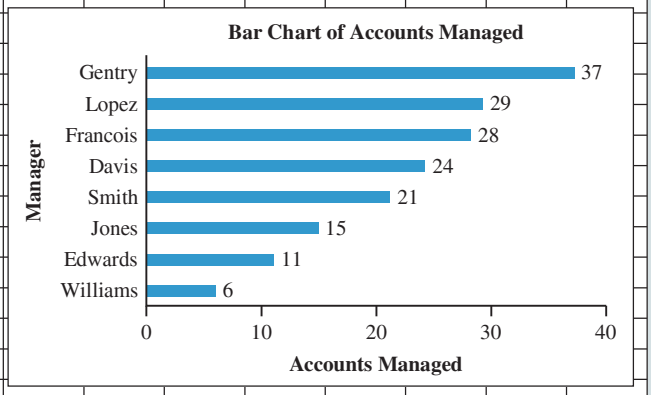
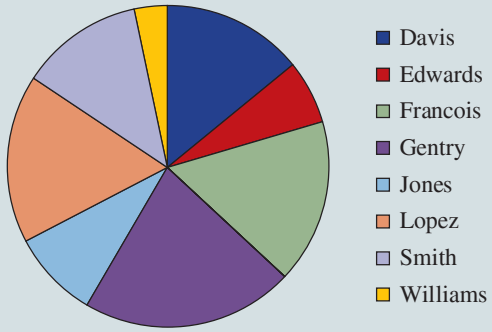


FIGURE 3.26 Pie Chart of Accounts Managed



Because of the difficulty in visually comparing area, many experts also recommend against the use of three-dimensional (3-D) charts in most settings. Excel makes it very easy to create 3-D bar, line, pie, and other types of charts. In most cases, however, the 3-D effect simply adds unnecessary detail that does not help explain the data. As an alternative, consider the use of multiple lines on a line chart (instead of adding a *z*-axis), employing multiple charts, or creating bubble charts in which the size of the bubble can represent the *z*-axis value. Never use a 3-D chart when a two-dimensional chart will suffice.

Bubble Charts

A **bubble chart** is a graphical means of visualizing three variables in a two-dimensional graph and is therefore sometimes a preferred alternative to a 3-D graph. Suppose that we want to compare the number of billionaires in various countries. Table 3.11 provides a sample of six countries, showing, for each country, the number of billionaires per 10 million residents, the per capita income, and the total number of billionaires. We can create a bubble chart using Excel to further examine these data:

DATA file
Billionaires

Step 1. Select cells B2:D7
Step 2. Click the **Insert** tab on the Ribbon
Step 3. In the **Charts** group, click **Insert Scatter (X,Y) or Bubble Chart**
In the **Bubble** subgroup, click **Bubble**
Step 4. Select the chart that was just created to reveal the **Chart Buttons**

TABLE 3.11 Sample Data on Billionaires per Country

Country	Billionaires per 10M Residents	Per Capita Income	No. of Billionaires
United States	54.7	$54,600	1,764
China	1.5	$12,880	213
Germany	12.5	$45,888	103
India	0.7	$ 5,855	90
Russia	6.2	$24,850	88
Mexico	1.2	$17,881	15

Step 5. Click the **Chart Elements** button
Select the check boxes for **Axes**, **Axis Titles**, **Chart Title** and **Data Labels**. **Deselect** the check box for **Gridlines**.
Click on the text box under the horizontal axis, and replace "Axis Title" with *Billionaires per 10 Million Residents*
Click on the text box next to the vertical axis, and replace "Axis Title" with *Per Capita Income*
Click on the text box above the chart, and replace "Chart Title" with *Billionaires by Country*

Step 6. Double-click on one of the Data Labels in the chart (e.g., the "$54,600" next to the largest bubble in the chart) to reveal the **Format Data Labels** task pane

Step 7. In the **Format Data Labels** task pane, click the **Label Options** icon and open the **Label Options** area
Under **Label Contains**, select **Value from Cells** and click the **Select Range…** button
When the **Data Label Range** dialog box opens, select cells A2:A8 in the Worksheet
Click **OK**

Step 8. In the **Format Data Labels** task pane, deselect **Y Value** under **Label Contains**, and select **Right** under **Label Position**

The completed bubble chart appears in Figure 3.27. This size of each bubble in Figure 3.27 is proportionate to the number of billionaires in that country. The per capita income and billionaires per 10 million residents is displayed on the vertical and horizontal axes. This chart shows us that the United States has the most billionaires and the highest number of billionaires per 10 million residents. We can also see that China has quite a few billionaires but with much lower per capita income and much lower billionaires per 10 million residents (because of China's much larger population). Germany, Russia, and India all appear to have similar numbers of billionaires, but the per capita income and billionaires per 10 million residents are very different for each country. Bubble charts can be very effective for comparing categorical variables on two different quantitative values.

Heat Maps

A **heat map** is a two-dimensional graphical representation of data that uses different shades of color to indicate magnitude. Figure 3.28 shows a heat map indicating the magnitude of changes for a metric called same-store sales, which are commonly used in the retail industry to measure trends in sales. The cells shaded red in Figure 3.28 indicate declining same-store sales for the month, and cells shaded blue indicate increasing same-store sales for the month. Column N in Figure 3.28 also contains sparklines for the same-store sales data.

Figure 3.28 can be created in Excel by following these steps:

DATA file
SameStoreSales

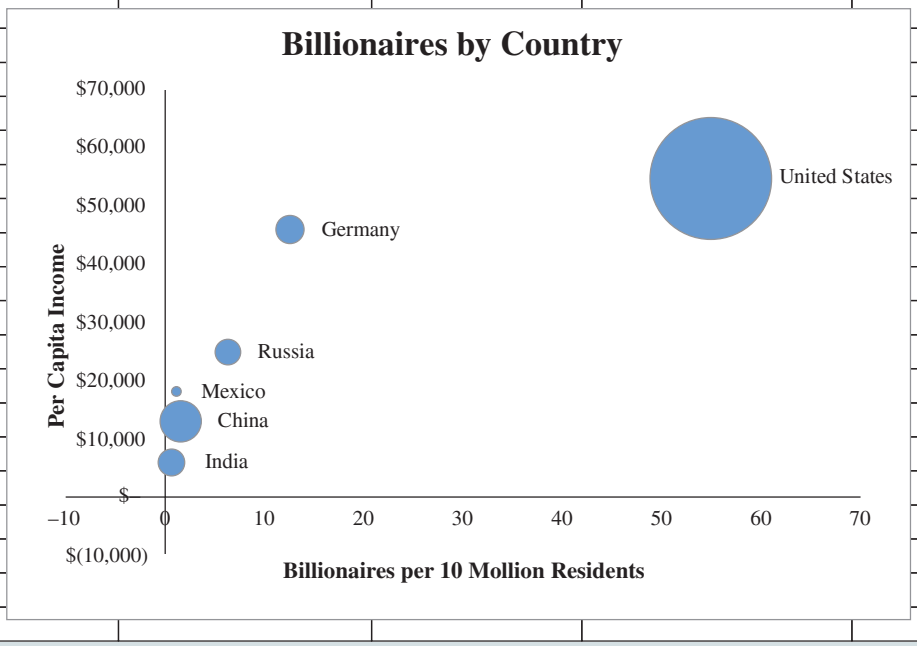
Step 1. Select cells B2:M17
Step 2. Click the **Home** tab on the Ribbon
Step 3. Click **Conditional Formatting** in the **Styles** group
Select **Color Scales** and click on **Blue–White–Red Color Scale**

To add the sparklines in column N, we use the following steps:

Step 4. Select cell N2
Step 5. Click the **Insert** tab on the Ribbon
Step 6. Click **Line** in the **Sparklines** group
Step 7. When the **Create Sparklines** dialog box opens:
Enter *B2:M2* in the **Data Range:** box
Enter *N2* in the **Location Range:** box and click **OK**
Step 8. Copy cell N2 to N3:N17

FIGURE 3.27 Bubble Chart Comparing Billionaires by Country

	A	B	C	D	E
1	**Country**	**Billionaires per 10M Residents**	**Per Capita Income**	**No. of Billionaires**	
2	United States	54.7	$ 54,600	1764	
3	China	1.5	$ 12,880	213	
4	Germany	12.5	$ 45,888	103	
5	India	0.7	$ 5,855	90	
6	Russia	6.2	$ 24,850	88	
7	Mexico	1.2	$ 17,881	15	



Both the heat map and the sparklines described here can also be created using the ***Quick Analysis*** *button*  . *To display this button, select cells B2:M17. The* ***Quick Analysis*** *button will appear at the bottom right of the selected cells. Click the button to display options for heat maps, sparklines, and other data-analysis tools.*

The heat map in Figure 3.28 helps the reader to easily identify trends and patterns. We can see that Austin has had positive increases throughout the year, while Pittsburgh has had consistently negative same-store sales results. Same-store sales at Cincinnati started the year negative but then became increasingly positive after May. In addition, we can differentiate between strong positive increases in Austin and less substantial positive increases in Chicago by means of color shadings. A sales manager could use the heat map in Figure 3.28 to identify stores that may require intervention and stores that may be used as models. Heat maps can be used effectively to convey data over different areas, across time, or both, as seen here.

Because heat maps depend strongly on the use of color to convey information, one must be careful to make sure that the colors can be easily differentiated and that they do not become overwhelming. To avoid problems with interpreting differences in color, we can add sparklines as shown in column N of Figure 3.28. The sparklines clearly show the overall trend (increasing or decreasing) for each location. However, we cannot gauge

FIGURE 3.28 Heat Map and Sparklines for Same-Store Sales Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	SPARKLINES
2	St. Louis	–2%	–1%	–1%	0%	2%	4%	3%	5%	6%	7%	8%	8%	
3	Phoenix	5%	4%	4%	2%	2%	–2%	–5%	–8%	–6%	–5%	–7%	–8%	
4	Albany	–5%	–6%	–4%	–5%	–2%	–5%	–5%	–3%	–1%	–2%	–1%	–2%	
5	Austin	16%	15%	15%	16%	18%	17%	14%	15%	16%	19%	18%	16%	
6	Cincinnati	–9%	–6%	–7%	–3%	3%	6%	8%	11%	10%	11%	13%	11%	
7	San Francisco	2%	4%	5%	8%	4%	2%	4%	3%	1%	–1%	1%	2%	
8	Seattle	7%	7%	8%	7%	5%	4%	2%	0%	–2%	–4%	–6%	–5%	
9	Chicago	5%	3%	2%	6%	8%	7%	8%	5%	8%	10%	9%	8%	
10	Atlanta	12%	14%	13%	17%	12%	11%	8%	7%	7%	8%	5%	3%	
11	Miami	2%	3%	0%	1%	–1%	–4%	–6%	–8%	–11%	–13%	–11%	–10%	
12	Minneapolis	–6%	–6%	–8%	–5%	–6%	–5%	–5%	–7%	–5%	–2%	–1%	–2%	
13	Denver	5%	4%	1%	1%	2%	3%	1%	–1%	0%	1%	2%	3%	
14	Salt Lake City	7%	7%	7%	13%	12%	8%	5%	9%	10%	9%	7%	6%	
15	Raleigh	4%	2%	0%	5%	4%	3%	5%	5%	9%	11%	8%	6%	
16	Boston	–5%	–5%	–3%	4%	–5%	–4%	–3%	–1%	1%	2%	3%	5%	
17	Pittsburgh	–6%	–6%	–4%	–5%	–3%	–3%	–1%	–2%	–2%	–1%	–2%	–1%	

differences in the magnitudes of increases and decreases among locations using sparklines. The combination of a heat map and sparklines here is a particularly effective way to show both trend and magnitude.

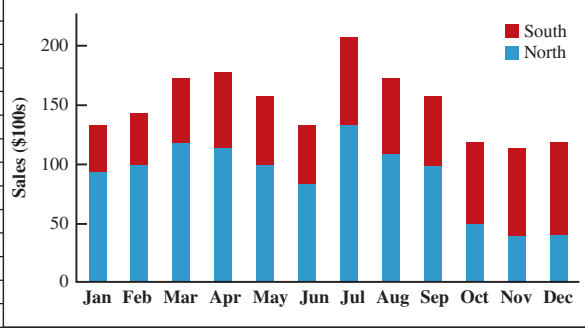
DATA file
KirklandRegional

Additional Charts for Multiple Variables

Figure 3.29 provides an alternative display for the regional sales data of air compressors for Kirkland Industries. The figure uses a **stacked-column chart** to display the North and the South regional sales data previously shown in a line chart in Figure 3.21. We could also

FIGURE 3.29 Stacked-Column Chart for Regional Sales Data for Kirkland Industries

	A	B	C
1		Sales ($100s)	
2	Month	North	South
3	Jan	95	40
4	Feb	100	45
5	Mar	120	55
6	Apr	115	65
7	May	100	60
8	Jun	85	50
9	Jul	135	75
10	Aug	110	65
11	Sep	100	60
12	Oct	50	70
13	Nov	40	75
14	Dec	40	80



use a stacked-bar chart to display the same data by using horizontal bars instead of vertical. To create the stacked-column chart shown in Figure 3.29, we use the following steps:

*Note that here we have not included the additional steps for formatting the chart in Excel using the **Chart Elements** button, but the steps are similar to those used to create the previous charts.*

Step 1. Select cells A2:C14
Step 2. Click the **Insert** tab on the Ribbon
Step 3. In the **Charts** group, click the **Insert Column or Bar Chart** button
Select **Stacked Column** under **2-D Column**

Stacked-column and stacked-bar charts allow the reader to compare the relative values of quantitative variables for the same category in a bar chart. However, these charts suffer from the same difficulties as pie charts because the human eye has difficulty perceiving small differences in areas. As a result, experts often recommend against the use of stacked-column and stacked-bar charts for more than a couple of quantitative variables in each category. An alternative chart for these same data is called a **clustered-column (or clustered-bar) chart**. It is created in Excel following the same steps but selecting **Clustered Column** under the **2-D Column** in Step 3. Clustered-column and clustered-bar charts are often superior to stacked-column and stacked-bar charts for comparing quantitative variables, but they can become cluttered for more than a few quantitative variables per category.

Clustered-column (bar) charts are also referred to as side-by-side-column (bar) charts.

An alternative that is often preferred to both stacked and clustered charts, particularly when many quantitative variables need to be displayed, is to use multiple charts. For the regional sales data, we would include two column charts: one for sales in the North and one for sales in the South. For additional regions, we would simply add additional column charts. To facilitate comparisons between the data displayed in each chart, it is important to maintain consistent axes from one chart to another. The categorical variables should be listed in the same order in each chart, and the axis for the quantitative variable should have the same range. For instance, the vertical axis for both North and South sales starts at 0 and ends at 140. This makes it easy to see that, in most months, the North region has greater sales. Figure 3.30 compares the approaches using stacked-, clustered-, and multiple-bar charts for the regional sales data.

Figure 3.30 shows that the multiple-column charts require considerably more space than the stacked- and clustered-column charts. However, when comparing many quantitative variables, using multiple charts can often be superior even if each chart must be made smaller. Stacked-column and stacked-bar charts should be used only when comparing a few quantitative variables and when there are large differences in the relative values of the quantitative variables within the category.

An especially useful chart for displaying multiple variables is the **scatter-chart matrix**. Table 3.12 contains a partial listing of the data for each of New York City's 55 subboroughs (a designation of a community within New York City) on monthly median rent, percentage of college graduates, poverty rate, and mean travel time to work. Suppose we want to examine the relationship between these different categorical variables. Figure 3.31 displays a scatter-chart matrix (scatter-plot matrix) for data related to rentals in New York City.

A scatter-chart matrix allows the reader to easily see the relationships among multiple variables. Each scatter chart in the matrix is created in the same manner as for creating a single scatter chart. Each column and row in the scatter-chart matrix corresponds to one categorical variable. For instance, row 1 and column 1 in Figure 3.31 correspond to the median monthly rent variable. Row 2 and column 2 correspond to the percentage of college graduates variable. Therefore, the scatter chart shown in row 1, column 2 shows the relationship between median monthly rent (on the y-axis) and the percentage of college graduates (on the x-axis) in New York City subboroughs. The scatter chart shown in row 2, column 3 shows the relationship between the percentage of college graduates (on the y-axis) and poverty rate (on the x-axis).

Figure 3.31 allows us to infer several interesting findings. Because the points in the scatter chart in row 1, column 2 generally get higher moving from left to right, this tells us that subboroughs with higher percentages of college graduates appear to have higher median monthly rents. The scatter chart in row 1, column 3 indicates that subboroughs with higher

FIGURE 3.30 Comparing Stacked-, Clustered-, and Multiple-Column Charts for the Regional Sales Data for Kirkland Industries

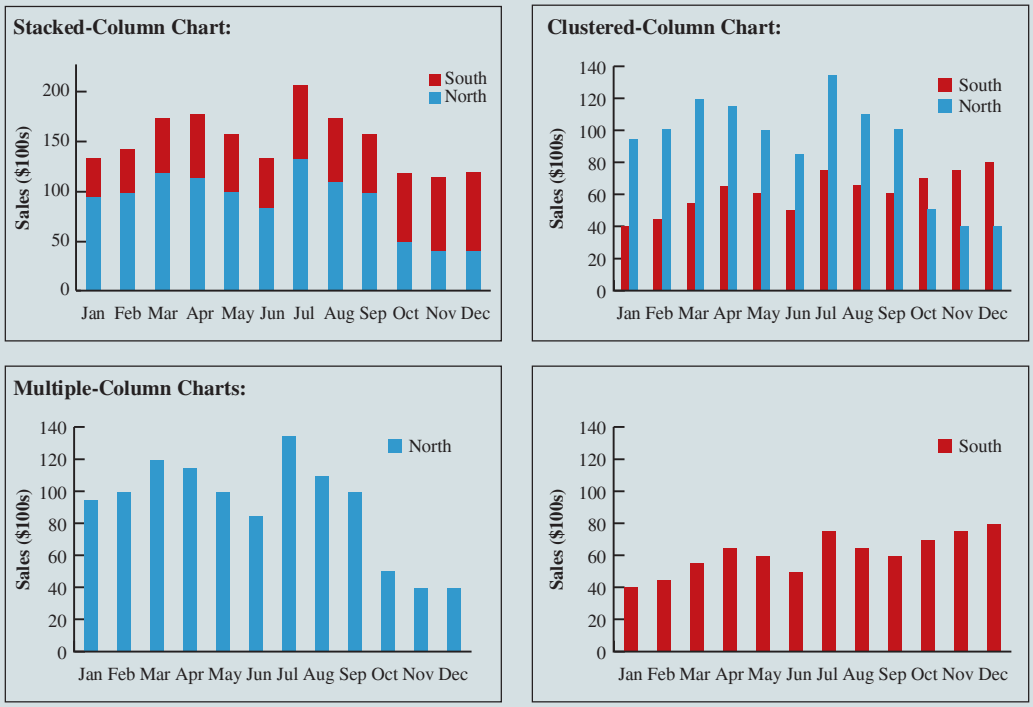


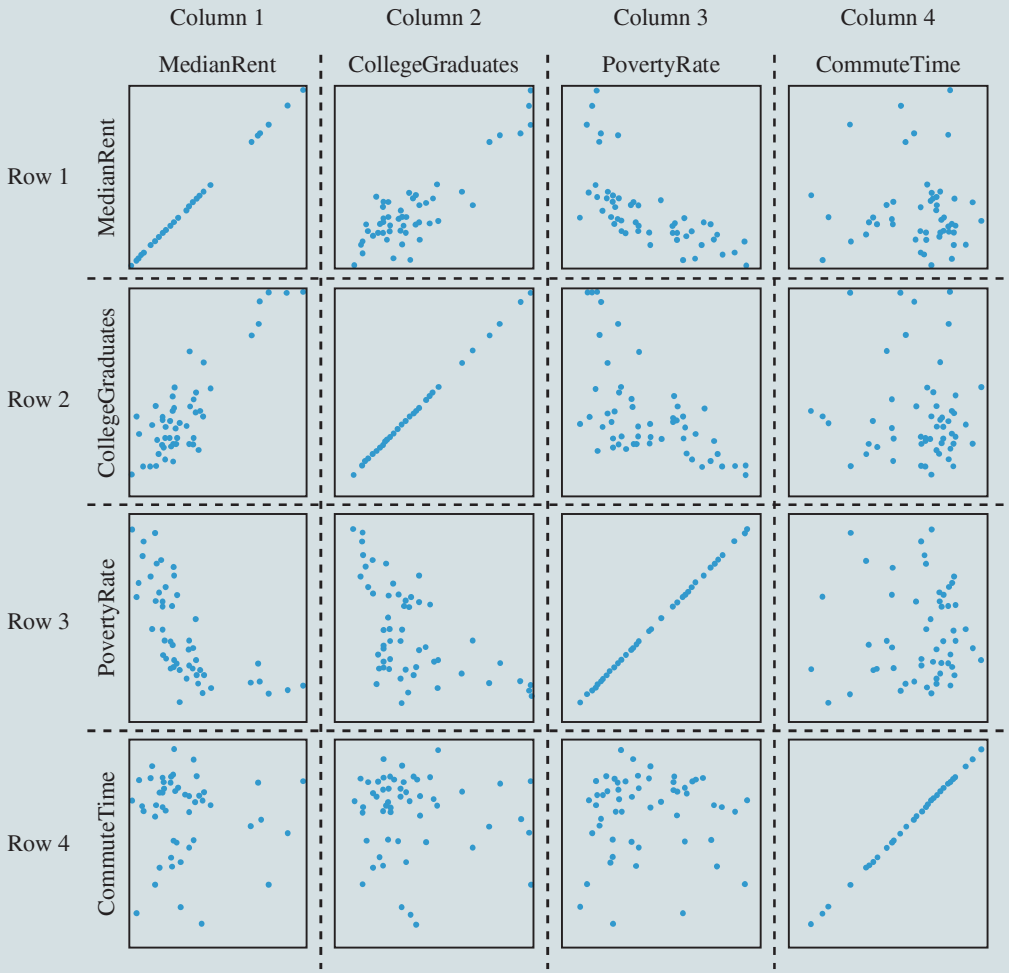
TABLE 3.12 Data for New York City Subboroughs

DATA file
NYCityData

Area	Median Monthly Rent ($)	Percentage College Graduates (%)	Poverty Rate (%)	Travel Time (min)
Astoria	1,106	36.8	15.9	35.4
Bay Ridge	1,082	34.3	15.6	41.9
Bayside/Little Neck	1,243	41.3	7.6	40.6
Bedford Stuyvesant	822	21.0	34.2	40.5
Bensonhurst	876	17.7	14.4	44.0
Borough Park	980	26.0	27.6	35.3
Brooklyn Heights/ Fort Greene	1,086	55.3	17.4	34.5
Brownsville/ Ocean Hill	714	11.6	36.0	40.3
Bushwick	945	13.3	33.5	35.5
Central Harlem	665	30.6	27.1	25.0
Chelsea/Clinton/ Midtown	1,624	66.1	12.7	43.7
Coney Island	786	27.2	20.0	46.3
⋮	⋮	⋮	⋮	⋮

The scatter charts along the diagonal in a scatter-chart matrix (e.g., in row 1, column 1 and in row 2, column 2) display the relationship between a variable and itself. Therefore, the points in these scatter charts will always fall along a straight line at a 45-degree angle, as shown in Figure 3.31.

FIGURE 3.31 Scatter-Chart Matrix for New York City Rent Data



In the appendix available within the MindTap Reader, we demonstrate how to create a scatter-chart matrix similar to that shown in Figure 3.31 using the Excel Add-in Analytic Solver. Statistical software packages such as R, NCSS, JMP, and SAS can also be used to create these matrixes.

poverty rates appear to have lower median monthly rents. The data in row 2, column 3 show that subboroughs with higher poverty rates tend to have lower percentages of college graduates. The scatter charts in column 4 show that the relationships between the mean travel time and the other categorical variables are not as clear as relationships in other columns.

The scatter-chart matrix is very useful in analyzing relationships among variables. Unfortunately, it is not possible to generate a scatter-chart matrix using standard Excel functions.

PivotCharts in Excel

DATA file
Restaurant

To summarize and analyze data with both a crosstabulation and charting, Excel pairs **PivotCharts** with PivotTables. Using the restaurant data introduced in Table 3.7 and Figure 3.7, we can create a PivotChart by taking the following steps:

Step 1. Click the **Insert** tab on the Ribbon
Step 2. In the **Charts** group, select **PivotChart**
Step 3. When the **Create PivotChart** dialog box appears:
Choose **Select a Table or Range**

Enter *A1:D301* in the **Table/Range:** box
Select **New Worksheet** as the location for the PivotTable Report
Click **OK**

Step 4. In the **PivotChart Fields** area, under **Choose fields to add to report:**
Drag the **Quality Rating** field to the **AXIS (CATEGORIES)** area
Drag the **Meal Price ($)** field to the **LEGEND (SERIES)** area
Drag the **Wait Time (min)** field to the **VALUES** area

Step 5. Click on **Sum of Wait Time (min)** in the **Values** area

Step 6. Select **Value Field Settings…** from the list of options that appear

Step 7. When the **Value Field Settings** dialog box appears:
Under **Summarize value field by**, select **Average**
Click **Number Format**
In the **Category:** box, select **Number**
Enter *1* for **Decimal places:**
Click **OK**
When the **Value Field Settings** dialog box reappears, click **OK**

Step 8. Right-click in cell B2 or any cell containing a meal price column label

Step 9. Select **Group** from the list of options that appears

Step 10. When the **Grouping** dialog box appears:
Enter *10* in the **Starting at:** box
Enter *49* in the **Ending at:** box
Enter *10* in the **By:** box
Click **OK**

Step 11. Right-click on "Excellent" in cell A5

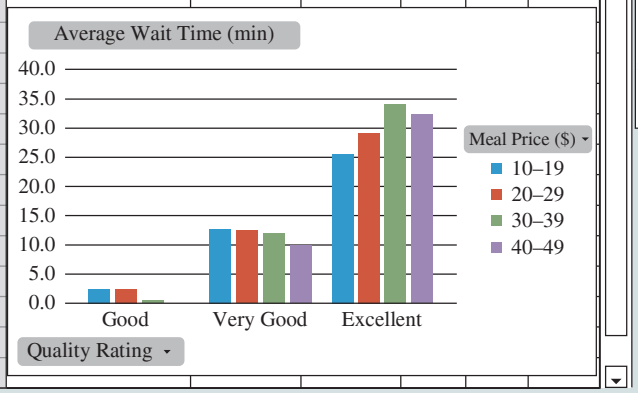
Step 12. Select **Move** and click **Move "Excellent" to End**

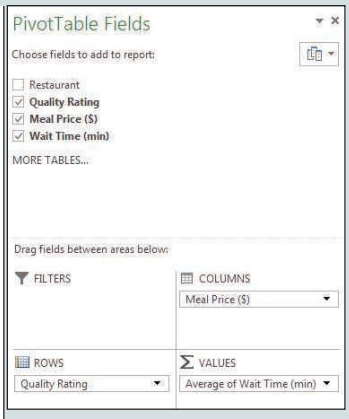
The completed PivotTable and PivotChart appear in Figure 3.32. The PivotChart is a clustered-column chart whose column heights correspond to the average wait times and are clustered into the categorical groupings of Good, Very Good, and Excellent. The columns

Like PivotTables, PivotCharts are interactive. You can use the arrows on the axes and legend labels to change the categorical data being displayed. For example, you can click on the ***Quality Rating*** *horizontal axis label (see Figure 3.32) and choose to look at only Very Good and Excellent restaurants, or you can click on the* ***Meal Price ($)*** *legend label and choose to view only certain meal price categories.*

FIGURE 3.32 PivotTable and PivotChart for the Restaurant Data

	A	B	C	D	E	F
1	Average of Wait Time(min)	Columns Lables				Grand
2	Row Labels	10–19	20–29	30–39	40–49	Total
3	Good	2.6	2.5	0.5		2.5
4	Very Good	12.6	12.6	12.0	10.0	12.3
5	Excellent	25.5	29.1	34.0	32.3	32.1
6	Grand Total	7.6	11.1	19.8	27.5	13.9





are different colors to differentiate the wait times at restaurants in the various meal price ranges. Figure 3.32 shows that Excellent restaurants have longer wait times than Good and Very Good restaurants. We also see that Excellent restaurants in the price range of $30–$39 have the longest wait times. The PivotChart displays the same information as that of the PivotTable in Figure 3.13, but the column chart used here makes it easier to compare the restaurants based on quality rating and meal price.

NOTES + COMMENTS

1. The steps for modifying and formatting charts were changed in Excel 2013. In versions of Excel prior to 2013, most chart-formatting options can be found in the **Layout** tab in the **Chart Tools** Ribbon. This is where you will find options for adding a **Chart Title, Axis Titles, Data Labels**, and so on in older versions of Excel.
2. Excel assumes that line charts will be used to graph only time series data. The Line Chart tool in Excel is the most intuitive for creating charts that include text entries for the horizontal axis (e.g., the month labels of Jan, Feb, Mar, etc. for the monthly sales data in Figure 3.19). When the horizontal axis represents numerical values (1, 2, 3, etc.), then it is easiest to go to the **Charts** group under the **Insert** tab in the Ribbon, click the **Insert Scatter (X,Y) or Bubble Chart** button, and then select the **Scatter with Straight Lines and Markers** button.
3. Color is frequently used to differentiate elements in a chart. However, be wary of the use of color to differentiate for several reasons: (1) Many people are color-blind and may not be able to differentiate colors. (2) Many charts are printed in black and white as handouts, which reduces or eliminates the impact of color. (3) The use of too many colors in a chart can make the chart appear too busy and distract or even confuse the reader. In many cases, it is preferable to differentiate chart elements with dashed lines, patterns, or labels.
4. Histograms and box plots (discussed in Chapter 2 in relation to analyzing distributions) are other effective data-visualization tools for summarizing the distribution of data.

3.4 Advanced Data Visualization

In this chapter, we have presented only some of the most basic ideas for using data visualization effectively both to analyze data and to communicate data analysis to others. The charts discussed so far are those most commonly used and will suffice for most data-visualization needs. However, many additional concepts, charts, and tools can be used to improve your data-visualization techniques. In this section we briefly mention some of them.

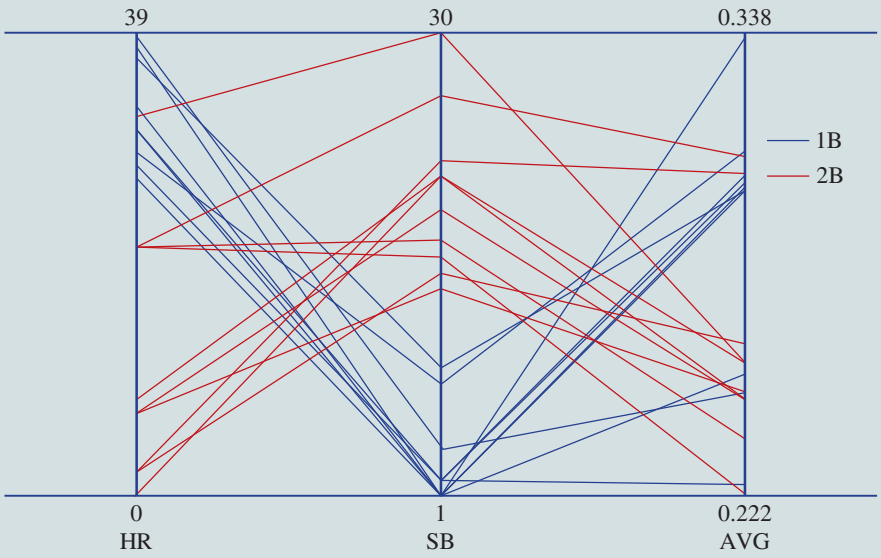
Advanced Charts

Although line charts, bar charts, scatter charts, and bubble charts suffice for most data-visualization applications, other charts can be very helpful in certain situations. One type of helpful chart for examining data with more than two variables is the **parallel-coordinates plot**, which includes a different vertical axis for each variable. Each observation in the data set is represented by drawing a line on the parallel-coordinates plot connecting each vertical axis. The height of the line on each vertical axis represents the value taken by that observation for the variable corresponding to the vertical axis. For instance, Figure 3.33 displays a parallel coordinates plot for a sample of Major League Baseball players. The figure contains data for 10 players who play first base (1B) and 10 players who play second base (2B). For each player, the leftmost vertical axis plots his total number of home runs (HR). The center vertical axis plots the player's total number of stolen bases (SB), and the rightmost vertical axis plots the player's batting average. Various colors differentiate 1B players from 2B players (1B players are in blue and 2B players are in red).

We can make several interesting statements upon examining Figure 3.33. The sample of 1B players tend to hit lots of HR but have very few SB. Conversely, the sample of 2B players steal more bases but generally have fewer HR, although some 2B players have many HR and many SB. Finally, 1B players tend to have higher batting averages (AVG) than 2B players. We may infer from Figure 3.33 that the traits of 1B players may be different from

The appendix for this chapter available in the MindTap Reader describes how to create a parallel coordinates plot similar to Figure 3.33 using the Analytic Solver Excel Add-in.

FIGURE 3.33 Parallel Coordinates Plot for Baseball Data



those of 2B players. In general, this statement is true. Players at 1B tend to be offensive stars who hit for power and average, whereas players at 2B are often faster and more agile in order to handle the defensive responsibilities of the position (traits that are not common in strong HR hitters). Parallel-coordinates plots, in which you can differentiate categorical variable values using color as in Figure 3.33, can be very helpful in identifying common traits across multiple dimensions.

A **treemap** is useful for visualizing hierarchical data along multiple dimensions. SmartMoney's Map of the Market, shown in Figure 3.34, is a treemap for analyzing stock market performance. In the Map of the Market, each rectangle represents a particular company (Apple, Inc. is highlighted in Figure 3.34). The color of the rectangle represents the overall performance of the company's stock over the previous 52 weeks. The Map of the Market is also divided into market sectors (Health Care, Financials, Oil & Gas, etc.). The size of each company's rectangle provides information on the company's market capitalization size relative to the market sector and the entire market. Figure 3.34 shows that Apple has a very large market capitalization relative to other firms in the Technology sector and that it has performed exceptionally well over the previous 52 weeks. An investor can use the treemap in Figure 3.34 to quickly get an idea of the performance of individual companies relative to other companies in their market sector as well as the performance of entire market sectors relative to other sectors.

Excel allows the user to create treemap charts. The step-by-step directions below explain how to create a treemap in Excel for the top-100 global companies based on 2014 market value using data in the file *Global100* . In this file we provide the continent where the company is headquartered in column A, the country headquarters in column B, the name of the company in column C, and the market value in column D. For the treemap to display properly in Excel, the data should be sorted by column A, "Continent," which is the highest level of the hierarchy.

Note that the treemap chart is not available in versions of Excel prior to Excel 2016.

Step 1. Select cells A1: D101

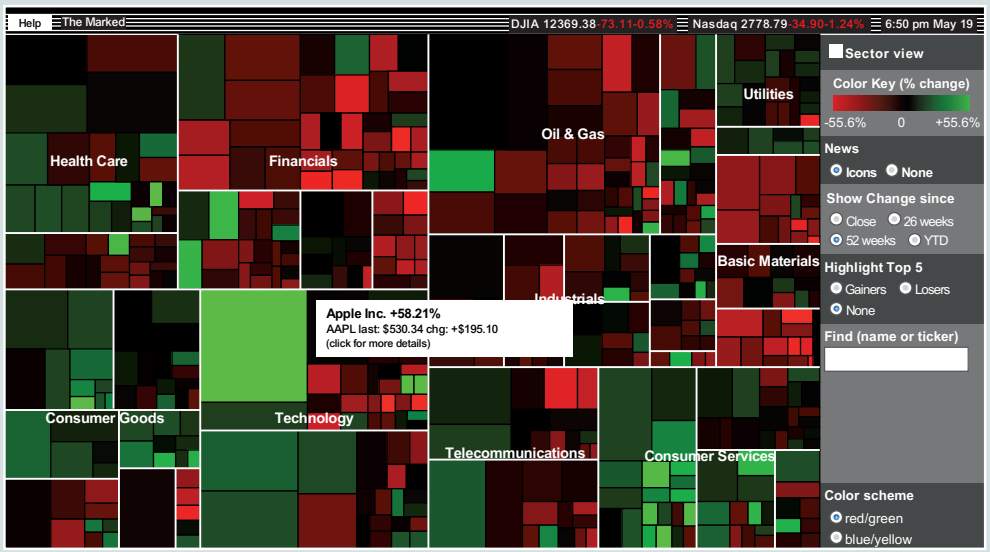
Step 2. Click **Insert** on the Ribbon

Click on the **Insert Hierarchy Chart** button  in the **Charts** group

Select **Treemap**  from the drop-down menu

The Map of the Market is based on work done by Professor Ben Shneiderman and students at the University of Maryland Human–Computer Interaction Lab.

FIGURE 3.34 SmartMoney's Map of the Market as an Example of a Treemap



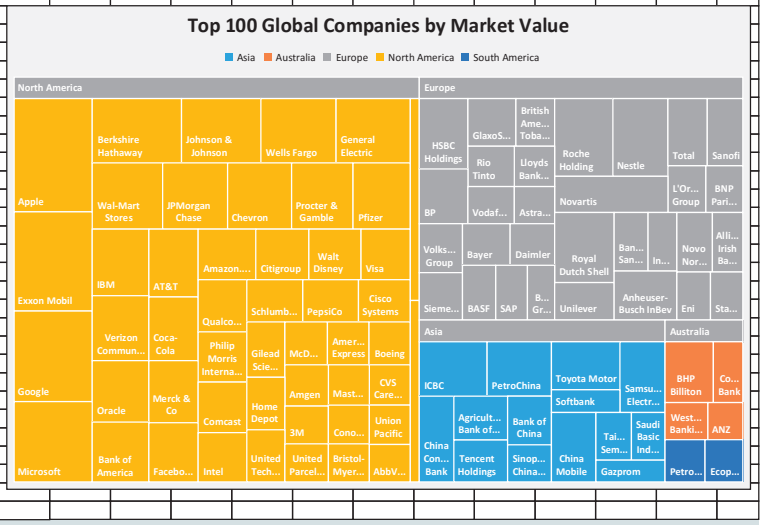
DATA file
Global100

Step 3. When the treemap chart appears, right-click on the treemap portion of the chart
Select **Format Data Series…** in the pop-up menu
When the **Format Data Series** task pane opens, select **Banner**

Figure 3.35 shows the completed treemap created with Excel. Selecting **Banner** in Step 3 places the name of each continent as a banner title within the treemap. Each continent is also

FIGURE 3.35 Treemap Created in Excel for Top 100 Global Companies Data

	A	B	C	D
1	Continent	Country	Company	Market Value (Billions US $)
2	Asia	China	Agricultural Bank of China	141.1
3	Asia	China	Bank of China	124.2
4	Asia	China	China Construction Bank	174.4
5	Asia	China	ICBC	215.6
6	Asia	China	PetroChina	202
7	Asia	China	Sinopec-China Petroleum	94.7
8	Asia	China	Tencent Holdings	135.4
9	Asia	Hong Kong	China Mobile	184.6
10	Asia	Japan	Softbank	91.2
11	Asia	Japan	Toyota Motor	193.5
12	Asia	Russia	Gazprom	88.8
13	Asia	Saudi Arabia	Saudi Basic Industries	94.4
14	Asia	South Korea	Samsung Electronics	186.5
15	Asia	Taiwan	Taiwan Semiconductor	102
16	Australia	Australia	ANZ	83.9
17	Australia	Australia	BHP Billiton	182.3
18	Australia	Australia	Commonwealth Bank	114.5
19	Australia	Australia	Westpac Banking Group	99
20	Europe	Belgium	Anheuser-Busch InBev	171.2
21	Europe	Denmark	Novo Nordisk	119.6
22	Europe	France	BNP Paribas	98.6
23	Europe	France	L'OrÃ©al Group	98.7
24	Europe	France	Sanofi	137.1
25	Europe	France	Total	149.8
26	Europe	Germany	BASF	102.3
27	Europe	Germany	Bayer	111.3
28	Europe	Germany	BMW Group	83.4
29	Europe	Germany	Daimler	102.9



assigned a different color within the treemap. From this figure we can see that North America has more top-100 companies than any other continent, followed by Europe and then Asia. The size of the rectangles for each company in the treemap represents their relative market value. We can see that Apple, ExxonMobile, Google, and Microsoft have the four highest market values. Australia has only four companies in the top 100 and South America has two. Africa and Antarctica have no companies in the top 100. Hovering your pointer over one of the companies in the treemap will display the market value for that company.

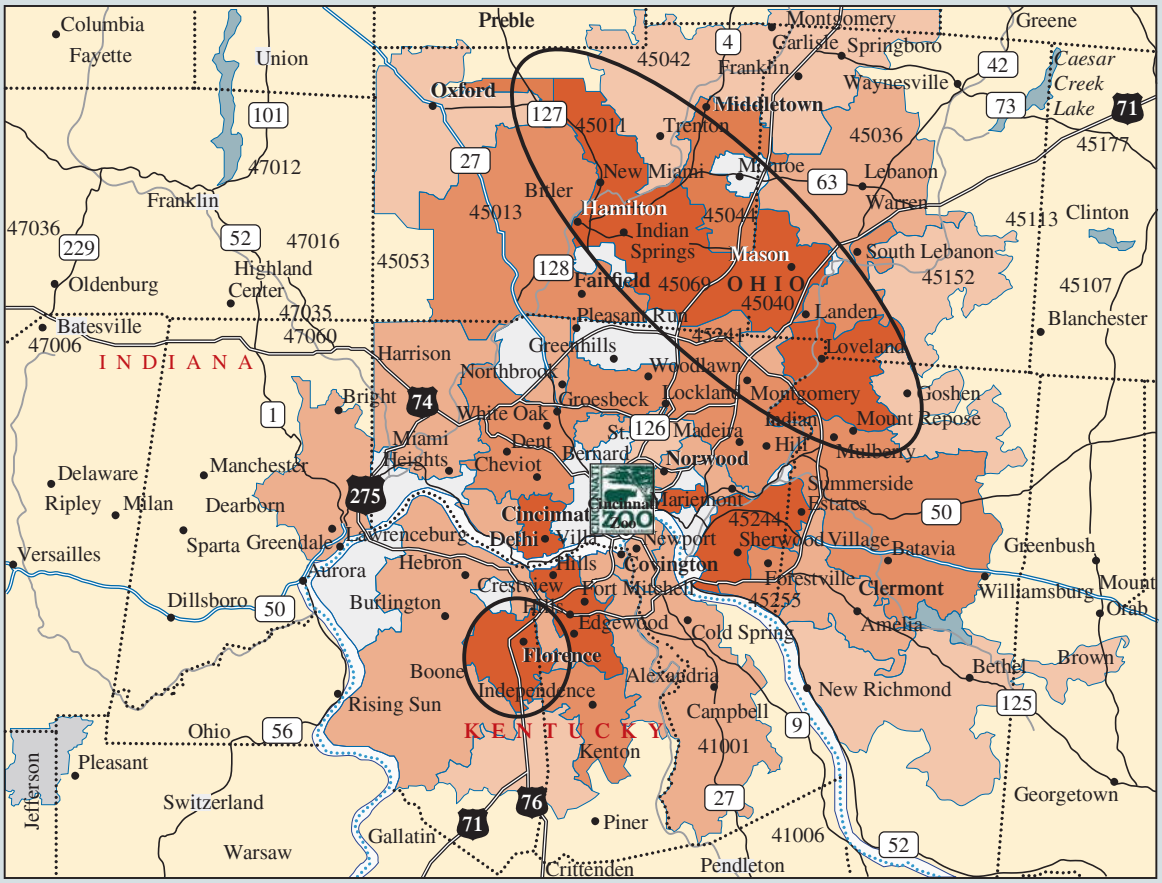
Geographic Information Systems Charts

A GIS chart such as that shown in Figure 3.36 is an example of geoanalytics, the use of data by geographical area or some other form of spatial referencing to generate insights.

Consider the case of the Cincinnati Zoo & Botanical Garden, which derives much of its revenue from selling annual memberships to customers. The Cincinnati Zoo would like to better understand where its current members are located. Figure 3.36 displays a map of the Cincinnati, Ohio, metropolitan area showing the relative concentrations of Cincinnati Zoo members. The more darkly shaded areas represent areas with a greater number of members. Figure 3.36 is an example of the output from a **geographic information system (GIS)**, which merges maps and statistics to present data collected over different geographic areas. Displaying geographic data on a map can often help in interpreting data and observing patterns.

The GIS chart in Figure 3.36 combines a heat map and a geographical map to help the reader analyze this data set. From the figure we can see that a high concentration of zoo members in a band to the northeast of the zoo that includes the cities of Mason and

FIGURE 3.36 GIS Chart for Cincinnati Zoo Member Data



Hamilton (circled). Also, a high concentration of zoo members lies to the southwest of the zoo around the city of Florence. These observations could prompt the zoo manager to identify the shared characteristics of the populations of Mason, Hamilton, and Florence to learn what is leading them to be zoo members. If these characteristics can be identified, the manager can then try to identify other nearby populations that share these characteristics as potential markets for increasing the number of zoo members.

3D Maps is called Power Map in Excel 2013, but it is not as fully integrated as it is in Excel 2016. This feature is not available in Excel versions prior to Excel 2013.

Excel has a feature called 3D Maps that allows the user to create interactive GIS-type charts. This tool is quite powerful, and the full capabilities are beyond the scope of this text. The step-by-step directions below show an example using data from the World Bank on gross domestic product (GDP) for countries around the world.

DATA file
WorldGDP

Step 1. Select cells A4:C267

Step 2. Click the **Insert** tab on the Ribbon

Click the **3D Map** button in the **Tours** group

Select **Open 3D Maps.** This will open a new Excel window that displays a world map (see Figure 3.37)

Step 3. Drag **GDP 2014 (Billions US $)** from the **Field List** to the **Height** box in the **Data** area of the **Layer 1** task pane.

Click the **Change the visualization to Region** button in the **Data** area of the **Layer 1** task pane.

Step 4. Click **Layer Options** in the **Layer 1** task pane.

Change the **Color** to a dark red color to give the countries more differentiation on the world map.

FIGURE 3.37 Initial Window Opened by Clicking on 3D Map Button in Excel for World GDP Data

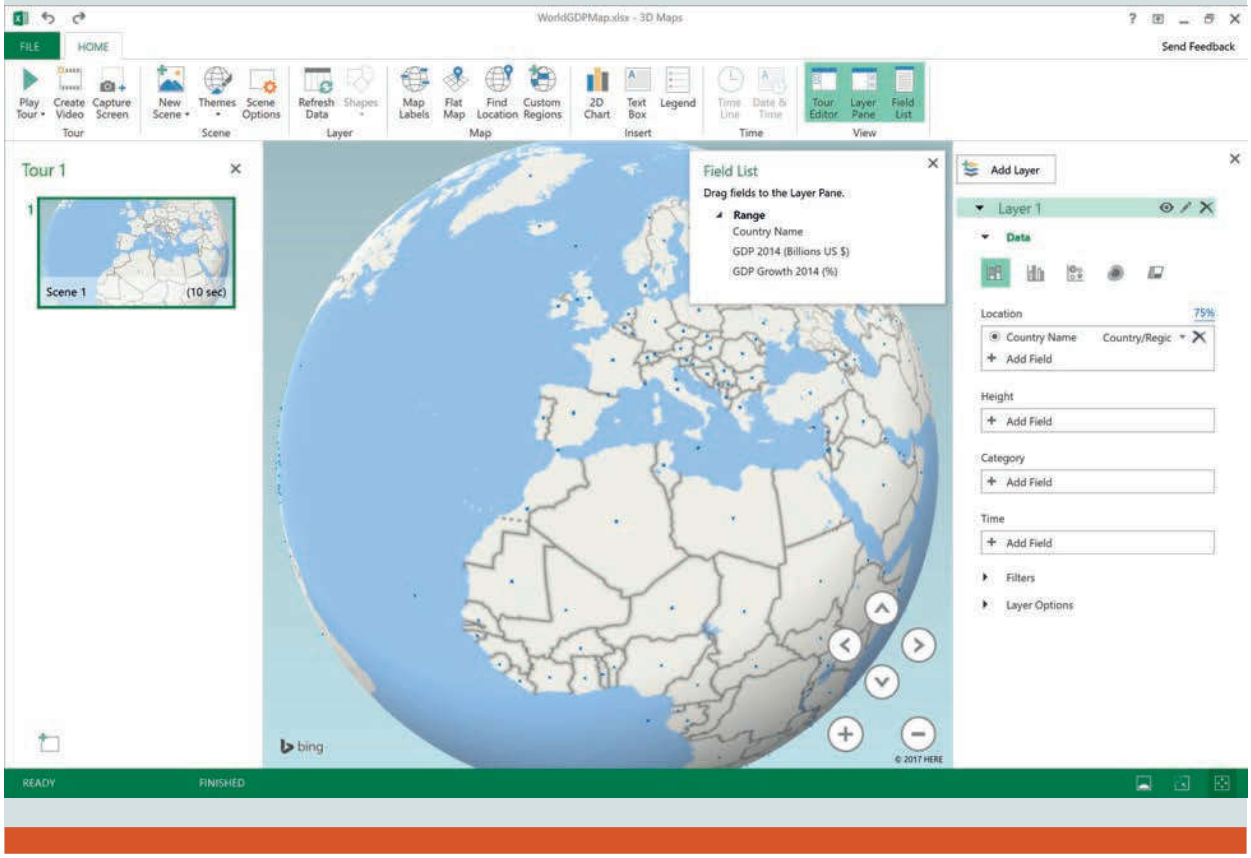
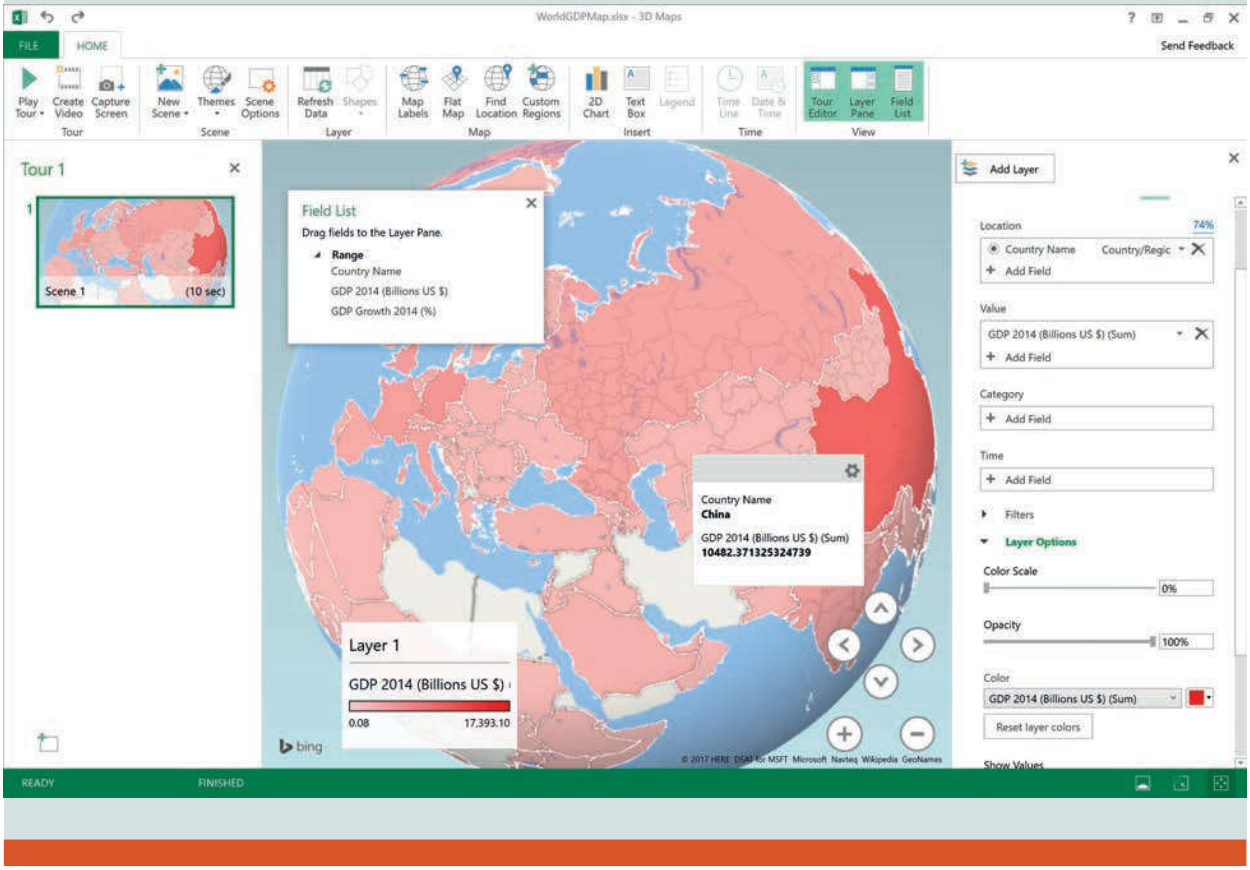


FIGURE 3.38 Completed 3D Map Created in Excel for World GDP Data



The completed GIS chart is shown in Figure 3.38. You can now click and drag the world map to view different parts of the world. Figure 3.38 shows much of Europe and Asia. The countries with the darker shading have higher GDPs. We can see that China has a very dark shading indicating very high GDP relative to other countries. Russia and Germany have slightly darker shadings than other countries shown indicating higher GDPs that most countries, but lower than China. If you hover over a country, it will display the Country Name and GDP 2014 (Billions US $) in a pop-up window. In Figure 3.38 we have hovered over China to display its GDP.

NOTES + COMMENTS

Spotfire, Tableau, QlikView, SAS Visual Analytics, R, and JMP are examples of software that include advanced data-visualization capabilities.

3.5 Data Dashboards

A **data dashboard** is a data-visualization tool that illustrates multiple metrics and automatically updates these metrics as new data become available. It is like an automobile's dashboard instrumentation that provides information on the vehicle's current speed, fuel level, and engine temperature so that a driver can assess current operating conditions and take effective action. Similarly, a data dashboard provides the important metrics that managers need to quickly assess the performance of their organization and react accordingly. In this section we provide guidelines for creating effective data dashboards and an example application.

Principles of Effective Data Dashboards

In an automobile dashboard, values such as current speed, fuel level, and oil pressure are displayed to give the driver a quick overview of current operating characteristics. In a business, the equivalent values are often indicative of the business's current operating characteristics, such as its financial position, the inventory on hand, customer service metrics, and the like. These values are typically known as **key performance indicators (KPIs)**. A data dashboard should provide timely summary information on KPIs that are important to the user, and it should do so in a manner that informs rather than overwhelms its user.

Key performance indicators are sometimes referred to as key performance metrics (KPMs).

Ideally, a data dashboard should present all KPIs as a single screen that a user can quickly scan to understand the business's current state of operations. Rather than requiring the user to scroll vertically and horizontally to see the entire dashboard, it is better to create multiple dashboards so that each dashboard can be viewed on a single screen.

The KPIs displayed in the data dashboard should convey meaning to its user and be related to the decisions the user makes. For example, the data dashboard for a marketing manager may have KPIs related to current sales measures and sales by region, while the data dashboard for a Chief Financial Officer should provide information on the current financial standing of the company, including cash on hand, current debt obligations, and so on.

A data dashboard should call attention to unusual measures that may require attention, but not in an overwhelming way. Color should be used to call attention to specific values to differentiate categorical variables, but the use of color should be restrained. Too many different or too bright colors make the presentation distracting and difficult to read.

Applications of Data Dashboards

To illustrate the use of a data dashboard in decision making, we discuss an application involving the Grogan Oil Company which has offices located in three cities in Texas: Austin (its headquarters), Houston, and Dallas. Grogan's Information Technology (IT) call center, located in Austin, handles calls from employees regarding computer-related problems involving software, Internet, and e-mail issues. For example, if a Grogan employee in Dallas has a computer software problem, the employee can call the IT call center for assistance.

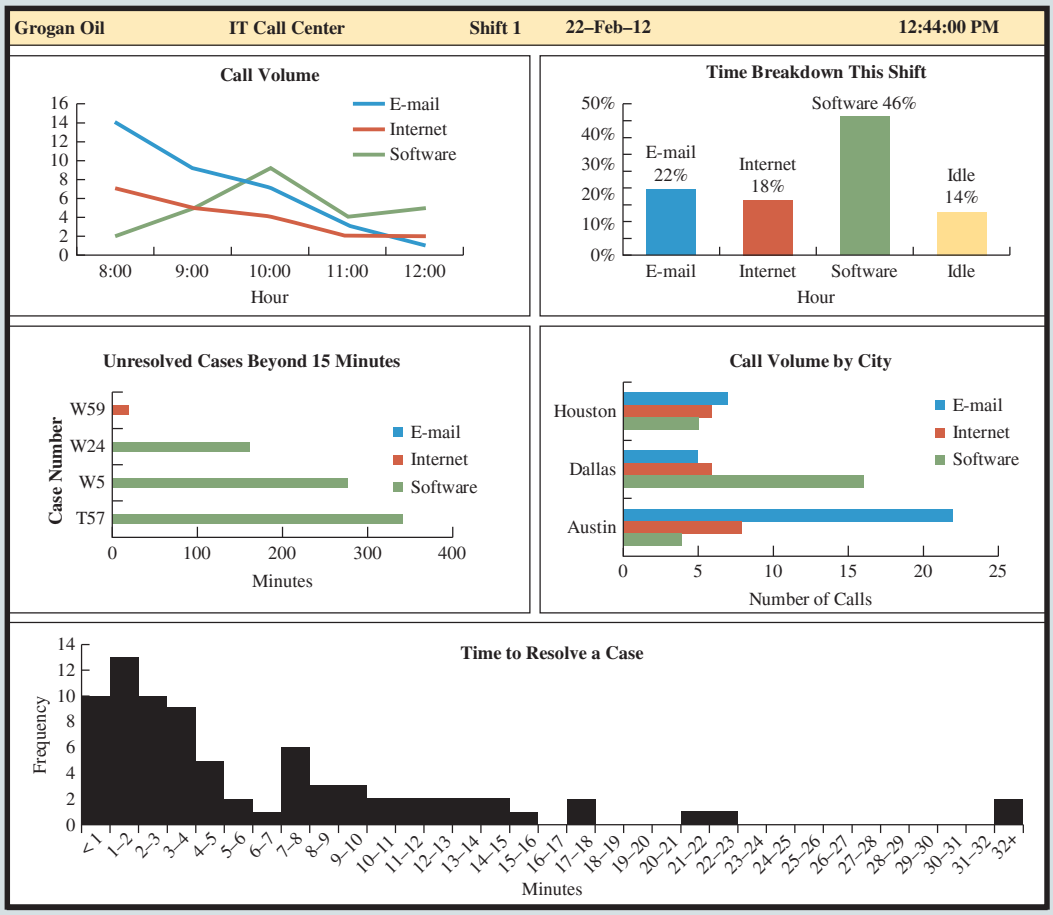
The data dashboard shown in Figure 3.39, developed to monitor the performance of the call center, combines several displays to track the call center's KPIs. The data presented are for the current shift, which started at 8:00 a.m. The line chart in the upper left-hand corner shows the call volume for each type of problem (Software, Internet, or E-mail) over time. This chart shows that call volume is heavier during the first few hours of the shift, that calls concerning e-mail issues appear to decrease over time, and that the volume of calls regarding software issues are highest at midmorning. A line chart is effective here because these are time series data and the line chart helps identify trends over time.

The column chart in the upper right-hand corner of the dashboard shows the percentage of time that call center employees spent on each type of problem or were idle (not working on a call). Both the line chart and the column chart are important displays in determining optimal staffing levels. For instance, knowing the call mix and how stressed the system is, as measured by percentage of idle time, can help the IT manager make sure that enough call center employees are available with the right level of expertise.

The clustered-bar chart in the middle right of the dashboard shows the call volume by type of problem for each of Grogan's offices. This allows the IT manager to quickly identify whether there is a particular type of problem by location. For example, the office in Austin seems to be reporting a relatively high number of issues with e-mail. If the source of the problem can be identified quickly, then the problem might be resolved quickly for many users all at once. Also, note that a relatively high number of software problems are coming from the Dallas office. In this case, the Dallas office is installing new software, resulting in more calls to the IT call center. Having been alerted to this by the Dallas office last week, the IT manager knew that calls coming from the Dallas office would spike, so the manager proactively increased staffing levels to handle the expected increase in calls.

For each unresolved case that was received more than 15 minutes ago, the bar chart shown in the middle left of the data dashboard displays the length of time for which each

FIGURE 3.39 Data Dashboard for the Grogan Oil Information Technology Call Center



case has been unresolved. This chart enables Grogan to quickly monitor the key problem cases and decide whether additional resources may be needed to resolve them. The worst case, T57, has been unresolved for over 300 minutes and is actually left over from the previous shift. Finally, the chart in the bottom panel shows the length of time required for resolved cases during the current shift. This chart is an example of a frequency distribution for quantitative data.

Chapter 2 discusses the construction of frequency distributions for quantitative and categorical data.

Throughout the dashboard, a consistent color coding scheme is used for problem type (E-mail, Software, and Internet). Because the Time to Resolve a Case chart is not broken down by problem type, dark shading is used so as not to confuse these values with a particular problem type. Other dashboard designs are certainly possible, and improvements could certainly be made to the design shown in Figure 3.39. However, what is important is that information is clearly communicated so that managers can improve their decision making.

The Grogan Oil data dashboard presents data at the operational level, is updated in real time, and is used for operational decisions such as staffing levels. Data dashboards may also be used at the tactical and strategic levels of management. For example, a sales manager could monitor sales by salesperson, by region, by product, and by customer. This would alert the sales manager to changes in sales patterns. At the highest level, a more strategic dashboard would allow upper management to quickly assess the financial health of the company by monitoring more aggregate financial, service-level, and capacity-utilization information.

NOTES + COMMENTS

The creation of data dashboards in Excel generally requires the use of macros written using Visual Basic for Applications (VBA). The use of VBA is beyond the scope of this textbook, but VBA is a powerful programming tool that can greatly increase the capabilities of Excel for analytics, including data visualization.

SUMMARY

In this chapter we covered techniques and tools related to data visualization. We discussed several important techniques for enhancing visual presentation, such as improving the clarity of tables and charts by removing unnecessary lines and presenting numerical values only to the precision necessary for analysis. We explained that tables can be preferable to charts for data visualization when the user needs to know exact numerical values. We introduced crosstabulation as a form of a table for two variables and explained how to use Excel to create a PivotTable.

We presented many charts in detail for data visualization, including scatter charts, line charts, bar and column charts, bubble charts, and heat maps. We explained that pie charts and three-dimensional charts are almost never preferred tools for data visualization and that bar (or column) charts are usually much more effective than pie charts. We also discussed several advanced data-visualization charts, such as parallel-coordinates plots, treemaps, and GIS charts. We introduced data dashboards as a data-visualization tool that provides a summary of a firm's operations in visual form to allow managers to quickly assess the current operating conditions and to aid decision making.

Many other types of charts can be used for specific forms of data visualization, but we have covered many of the most-popular and most-useful ones. Data visualization is very important for helping someone analyze data and identify important relations and patterns. The effective design of tables and charts is also necessary to communicate data analysis to others. Tables and charts should be only as complicated as necessary to help the user understand the patterns and relationships in the data.

GLOSSARY

Bar chart A graphical presentation that uses horizontal bars to display the magnitude of quantitative data. Each bar typically represents a class of a categorical variable.
Bubble chart A graphical presentation used to visualize three variables in a two-dimensional graph. The two axes represent two variables, and the magnitude of the third variable is given by the size of the bubble.
Chart A visual method for displaying data; also called a graph or a figure.
Clustered-column (or clustered-bar) chart A special type of column (bar) chart in which multiple bars are clustered in the same class to compare multiple variables; also known as a side-by-side-column (bar) chart.
Column chart A graphical presentation that uses vertical bars to display the magnitude of quantitative data. Each bar typically represents a class of a categorical variable.
Crosstabulation A tabular summary of data for two variables. The classes of one variable are represented by the rows; the classes for the other variable are represented by the columns.
Data dashboard A data-visualization tool that updates in real time and gives multiple outputs.
Data-ink ratio The ratio of the amount of ink used in a table or chart that is necessary to convey information to the total amount of ink used in the table and chart. Ink used that is not necessary to convey information reduces the data-ink ratio.
Geographic information system (GIS) A system that merges maps and statistics to present data collected over different geographies.

Heat map A two-dimensional graphical presentation of data in which color shadings indicate magnitudes.
Key performance indicator (KPI) A metric that is crucial for understanding the current performance of an organization; also known as a key performance metric (KPM).
Line chart A graphical presentation of time series data in which the data points are connected by a line.
Parallel-coordinates plot A graphical presentation used to examine more than two variables in which each variable is represented by a different vertical axis. Each observation in a data set is plotted in a parallel-coordinates plot by drawing a line between the values of each variable for the observation.
Pie chart A graphical presentation used to compare categorical data. Because of difficulties in comparing relative areas on a pie chart, these charts are not recommended. Bar or column charts are generally superior to pie charts for comparing categorical data.
PivotChart A graphical presentation created in Excel that functions similarly to a PivotTable.
PivotTable An interactive crosstabulation created in Excel.
Scatter chart A graphical presentation of the relationship between two quantitative variables. One variable is shown on the horizontal axis and the other on the vertical axis.
Scatter-chart matrix A graphical presentation that uses multiple scatter charts arranged as a matrix to illustrate the relationships among multiple variables.
Sparkline A special type of line chart that indicates the trend of data but not magnitude. A sparkline does not include axes or labels.
Stacked-column chart A special type of column (bar) chart in which multiple variables appear on the same bar.
Treemap A graphical presentation that is useful for visualizing hierarchical data along multiple dimensions. A treemap groups data according to the classes of a categorical variable and uses rectangles whose size relates to the magnitude of a quantitative variable.
Trendline A line that provides an approximation of the relationship between variables in a chart.

PROBLEMS

1. A sales manager is trying to determine appropriate sales performance bonuses for her team this year. The following table contains the data relevant to determining the bonuses, but it is not easy to read and interpret. Reformat the table to improve readability and to help the sales manager make her decisions about bonuses.

DATA file
SalesBonuses

Salesperson	Total Sales ($)	Average Performance Bonus Previous Years ($)	Customer Accounts	Years with Company
Smith, Michael	325,000.78	12,499.3452	124	14
Yu, Joe	13,678.21	239.9434	9	7
Reeves, Bill	452,359.19	21,987.2462	175	21
Hamilton, Joshua	87,423.91	7,642.9011	28	3
Harper, Derek	87,654.21	1,250.1393	21	4
Quinn, Dorothy	234,091.39	14,567.9833	48	9
Graves, Lorrie	379,401.94	27,981.4432	121	12
Sun, Yi	31,733.59	672.9111	7	1
Thompson, Nicole	127,845.22	13,322.9713	17	3