

PART 2

Descriptive Statistics

CHAPTER 2

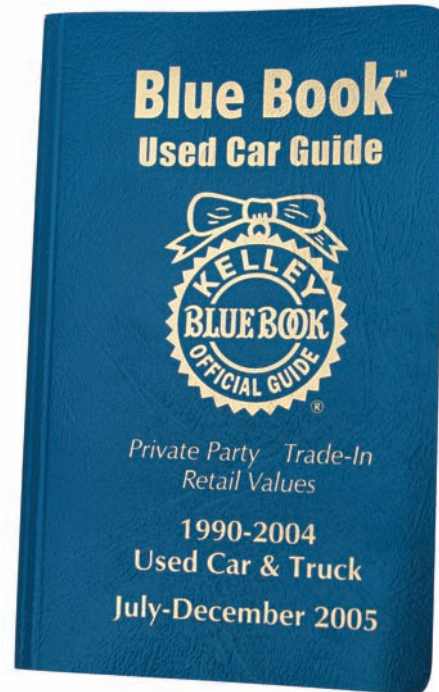
Organizing and Summarizing Data

CHAPTER 3

Numerically Summarizing Data

CHAPTER 4

Describing the Relation between Two Variables



Remember, statistics is a process. The first chapter (Part 1) dealt with the first two steps in the statistical process: (1) Identify the research objective and (2) collect the in-

formation needed to answer the questions in the research objective. The next three chapters deal (Part 2) with organizing, summarizing, and presenting the data collected. This step in the process is called descriptive statistics.

Organizing and Summarizing Data



Outline

- 2.1 Organizing Qualitative Data
- 2.2 Organizing Quantitative Data
- 2.3 Graphical Misrepresentations of Data
 - Chapter Review
 - Case Study: The Day the Sky Roared (On CD)

DECISIONS

Suppose that you work for the school newspaper. Your editor approaches you with a special reporting assignment. Your task is to write an article that describes the “typical” student at your school, complete with supporting information. How are you going to do this assignment? See the Decisions project on page 95.



●●● Putting It All Together

Chapter 1 discussed how to collect data. We learned that data can be obtained from four sources: (1) a census, (2) existing data sets, (3) surveys, or (4) designed experiments. When data are obtained, they are referred to as **raw data**. Raw data must be organized into a meaningful form.

Methods for organizing raw data include the creation of tables or graphs, which allow for a quick overview of the information collected. Describing data is the third step in the statistical process. The procedures used in this step depend on whether the data are qualitative, discrete, or continuous.

2.1 Organizing Qualitative Data

Preparing for This Section Before getting started, review the following:

- Qualitative data (Section 1.1, p. 6)

Objectives

- 1 Organize qualitative data in tables
- 2 Construct bar graphs
- 3 Construct pie charts

In this section we will concentrate on tabular and graphical summaries of qualitative data. In Section 2.2 we discuss methods for summarizing quantitative data.



In Other Words

Qualitative data is also known as categorical data.

1

Organize Qualitative Data in Tables

Recall that qualitative data provide measures that categorize or classify an individual. When qualitative data are collected, we are often interested in determining the number of individuals observed within each category.

Definition

A **frequency distribution** lists each category of data and the number of occurrences for each category of data.

EXAMPLE 1

Organizing Qualitative Data into a Frequency Distribution

Problem: A physical therapist wants to get a sense of the types of rehabilitation required by her patients. To do so, she obtains a simple random sample of 30 of her patients and records the body part requiring rehabilitation. See Table 1. Construct a frequency distribution of location of injury.



Table 1

Back	Back	Hand	Neck	Knee	Knee
Wrist	Back	Groin	Shoulder	Shoulder	Back
Elbow	Back	Back	Back	Back	Back
Back	Shoulder	Shoulder	Knee	Knee	Back
Hip	Knee	Hip	Hand	Back	Wrist

Source: Krystal Catton, student at Joliet Junior College

Approach: To construct a frequency distribution, we create a list of the body parts (categories) and tally each occurrence. Finally, we add up the number of tallies to determine the frequency.

Solution: See Table 2. From the table, we can see that the back is the most common body part requiring rehabilitation, with a total of 12.

Table 2

Body Part	Tally	Frequency
Back		12
Wrist		2
Elbow		1
Hip		2
Shoulder		4
Knee		5
Hand		2
Groin		1
Neck		1



CAUTION

The data in Table 2 are still qualitative. The frequency represents the count of each category.

With frequency distributions, it is a good idea to add up the frequency column to make sure that it sums to the number of observations. In the case of the data in Example 1, the frequency column adds up to 30, as it should.

Often, rather than being concerned with the frequency with which categories of data occur, we want to know the *relative frequency* of the categories.



In Other Words

A frequency distribution shows how many observations are in each category. A relative frequency distribution shows the proportion of observations that belong in each category.

Definition

The **relative frequency** is the proportion (or percent) of observations within a category and is found using the formula

$$\text{Relative frequency} = \frac{\text{frequency}}{\text{sum of all frequencies}} \quad (1)$$

A **relative frequency distribution** lists each category of data together with the relative frequency.

EXAMPLE 2

Constructing a Relative Frequency Distribution of Qualitative Data

Problem: Using the data in Table 2, construct a relative frequency distribution.

Approach: Add all the frequencies, and then use Formula (1) to compute the relative frequency of each category of data.

Solution: We add the values in the frequency column in Table 2:

$$\text{Sum of all frequencies} = 12 + 2 + 1 + 2 + 4 + 5 + 2 + 1 + 1 = 30$$

We now compute the relative frequency of each category. For example, the relative frequency of the category “Back” is

$$\frac{12}{30} = 0.4$$

After computing the relative frequency for the remaining categories, we obtain the relative frequency distribution shown in Table 3.

Body Part	Frequency	Relative Frequency
Back	12	$\frac{12}{30} = 0.4$
Wrist	2	$\frac{2}{30} \approx 0.0667$
Elbow	1	0.0333
Hip	2	0.0667
Shoulder	4	0.1333
Knee	5	0.1667
Hand	2	0.0667
Groin	1	0.0333
Neck	1	0.0333

From the table, we can see that the most common body part for rehabilitation is the back.



Using Technology

Some statistical spreadsheets such as MINITAB have a `Tally` command. This command will construct a frequency and relative frequency distribution of raw qualitative data.

It is a good idea to add up the entries in the relative frequency column to be sure they sum to 1. Sometimes the sum of the relative frequencies will differ slightly from 1 due to rounding. Nonetheless, the check is a good idea.

Now Work Problems 27(a) and (b).

2 Construct Bar Graphs

Once raw data are organized in a table, we can create graphs. Creating graphs allows us to see the data and get a sense of what the data are saying about the individuals in the study. In general, pictures of data result in a more powerful message than tables. Try the following exercise for yourself: Open a newspaper and look at a table and graph. Study each. Now put the paper away and close your eyes. What do you see in your mind's eye? Can you recall information obtained from the table or the graph? In general, people are more likely to recall information obtained from a graph than they are from a table.

One of the most common devices for graphically representing qualitative data is a bar graph.

Definition

A **bar graph** is constructed by labeling each category of data on a horizontal axis and the frequency or relative frequency of the category on the vertical axis. Rectangles of equal width are drawn for each category. The height of each rectangle is the category's frequency or relative frequency.

EXAMPLE 3

Constructing a Frequency and Relative Frequency Bar Graph

Problem: Use the data summarized in Table 3 to construct the following:

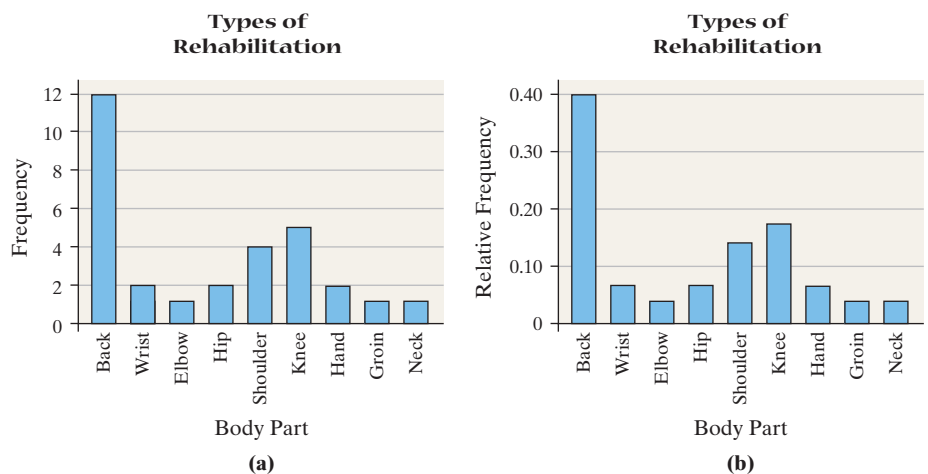
- (a) Frequency bar graph
- (b) Relative frequency bar graph

Approach: A horizontal axis is used to indicate the categories of the data (body parts, in this case), and a vertical axis is used to represent the frequency or relative frequency. Draw rectangles of equal width to the height that is the frequency or relative frequency for each category. The bars do not touch each other.

Solution

- (a) Figure 1(a) shows the frequency bar graph.
- (b) Figure 1(b) shows the relative frequency bar graph.

Figure 1



CAUTION

Watch out for graphs that start the scale at some value other than 0, have bars with unequal widths, or have bars with different colors, because they can misrepresent the data.

EXAMPLE 4

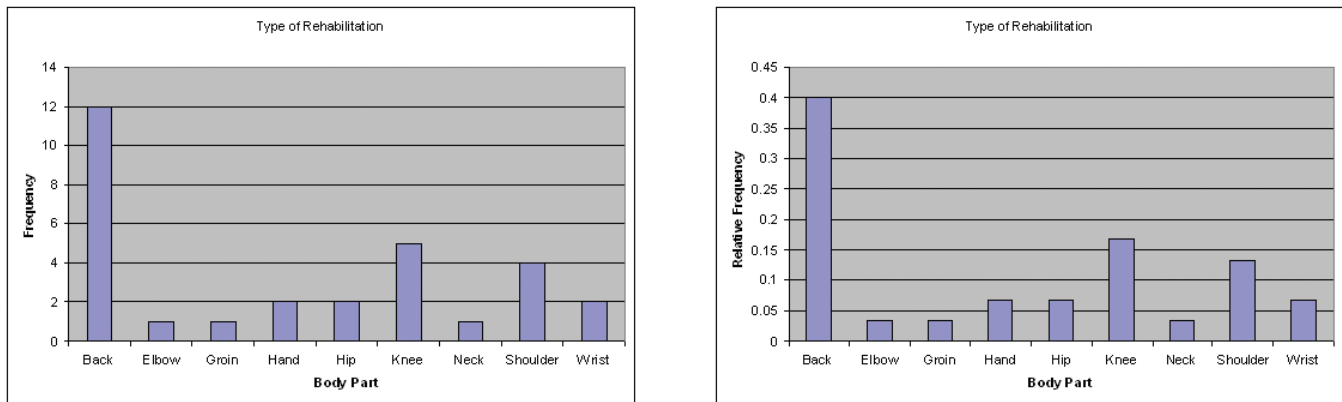
Constructing a Frequency or Relative Frequency Bar Graph Using Technology

Problem: Use a statistical spreadsheet to construct a frequency and relative frequency bar graph.

Approach: We will use Excel to construct the frequency and relative frequency bar graph. The steps for constructing the graphs using MINITAB or Excel are given in the Technology Step by Step on page 70. **Note:** The TI-83 and TI-84 Plus graphing calculators cannot draw frequency or relative frequency bar graphs.

Result: Figure 2(a) shows the frequency bar graph and Figure 2(b) shows the relative frequency bar graph obtained from Excel.

Figure 2



Now Work Problems 27(c) and (d).

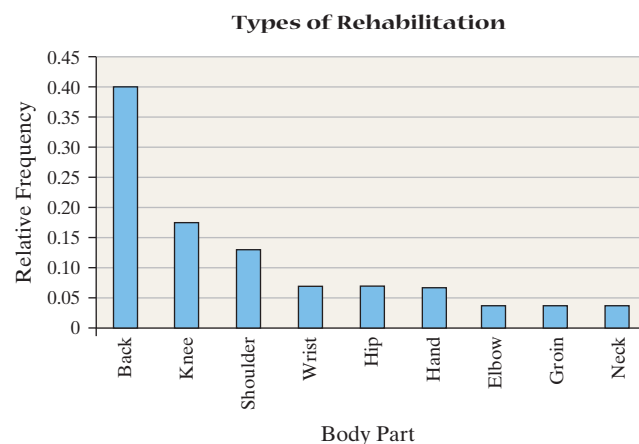
Some statisticians prefer to create bar graphs with the categories arranged in decreasing order of frequency.

Definition

A **Pareto chart** is a bar graph whose bars are drawn in decreasing order of frequency or relative frequency.

Figure 3 illustrates a relative frequency Pareto chart for the data in Table 3.

Figure 3



Using Technology

The graphs obtained from a different statistical package should not differ from those in Figure 2. Some packages use the word *count* in place of *frequency* or *percent* in place of *relative frequency*, however.

Side-by-Side Bar Graphs

Graphics provide insight when you are comparing two sets of data. For example, suppose we wanted to know if more people are finishing college today than in 1990. We could draw a **side-by-side bar graph** to compare the two data sets. Data sets should be compared by using relative frequencies, because different sample or population sizes make comparisons using frequencies difficult or misleading.

EXAMPLE 5 Comparing Two Data Sets

Problem: The data in Table 4 represent the educational attainment in 1990 and 2003 of adults 25 years and older who are residents of the United States. The data are in thousands. So 16,502 represents 16,502,000.

- (a) Draw a side-by-side relative frequency bar graph of the data.
 (b) Is a greater proportion of Americans earning bachelor's degrees?

**Table 4**

Educational Attainment	1990	2003
Less than 9th grade	16,502	12,276
9th to 12th grade, no diploma	22,842	16,323
High school diploma	47,643	59,292
Some college, no degree	29,780	31,762
Associate's degree	9,792	15,147
Bachelor's degree	20,833	33,213
Graduate or professional degree	11,478	17,169
Totals	158,870	185,182

Source: U.S. Census Bureau

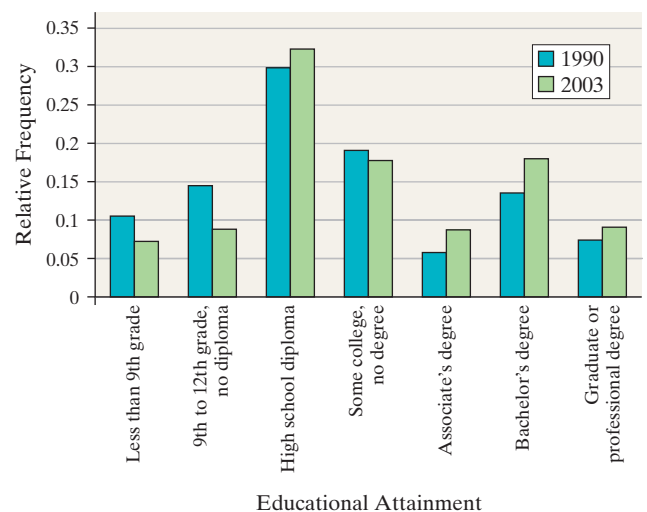
Approach: First, we determine the relative frequencies of each category for each year. To construct the side-by-side bar graphs, we draw two bars for each category of data. One of the bars will represent 1990 and the other will represent 2003.

Solution: Table 5 shows the relative frequency for each category.

- (a) The side-by-side bar graph is shown in Figure 4.

Table 5

Educational Attainment	1990	2003
Less than 9th grade	0.1039	0.0663
9th to 12th grade, no diploma	0.1438	0.0881
High school diploma	0.2999	0.3202
Some college, no degree	0.1874	0.1715
Associate's degree	0.0616	0.0818
Bachelor's degree	0.1311	0.1794
Graduate or professional degree	0.0722	0.0927

Figure 4 Educational Attainment in 1990 versus 2003

- (b) From the graph, we can see that the proportion of Americans 25 years and older who earned a bachelor's degree is higher in 2003. This information is not clear from the frequency table, because the sizes of the populations are different. Increases in the number of Americans who earned a bachelor's degree are due partly to the increases in the size of each population.

3 Construct Pie Charts

Pie charts are typically used to present the relative frequency of qualitative data.

Definition

A **pie chart** is a circle divided into sectors. Each sector represents a category of data. The area of each sector is proportional to the frequency of the category.

EXAMPLE 6 Constructing a Pie Chart

Problem: The data presented in Table 6 represent the educational attainment of residents of the United States 25 years or older in 2003, based on data obtained from the U.S. Census Bureau. The data are in thousands. Construct a pie chart of the data.



Table 6

Educational Attainment	2003
Less than 9th grade	12,276
9th to 12th grade, no diploma	16,323
High school diploma	59,292
Some college, no degree	31,762
Associate's degree	15,147
Bachelor's degree	33,213
Graduate or professional degree	17,169
Totals	185,182

Approach: The pie chart will have seven parts, or sectors, corresponding to the seven categories of data. The area of each sector is proportional to the frequency of each category. For example,

$$\frac{12,276}{185,182} = 0.0663$$

of all U.S. residents 25 years or older have less than a 9th-grade education. The category “less than 9th grade” will make up 6.63% of the area of the pie chart. Since a circle has 360 degrees, the degree measure of the sector for the category “less than 9th grade” will be $(0.0663)360^\circ \approx 24^\circ$. Use a protractor to measure each angle.

Solution: We follow the approach presented for the remaining categories of data to obtain Table 7.

Table 7

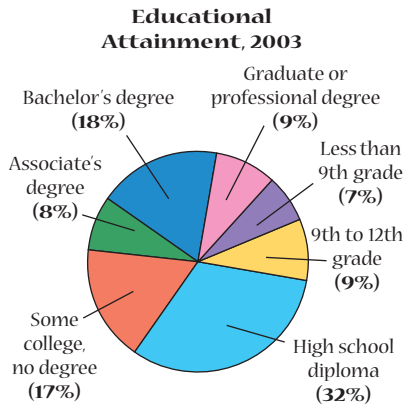
Education	Frequency	Relative Frequency	Degree Measure of Each Sector
Less than 9th grade	12,276	0.0663	24
9th to 12th grade, no diploma	16,323	0.0881	32
High school diploma	59,292	0.3202	115
Some college, no degree	31,762	0.1715	62
Associate's degree	15,147	0.0818	29
Bachelor's degree	33,213	0.1794	65
Graduate or professional degree	17,169	0.0927	33



USING TECHNOLOGY

Most statistical spreadsheets are capable of drawing pie charts. See the Technology Step by Step on page 70 for instructions on how to obtain a pie chart using MINITAB or Excel. The TI-83 and TI-84 Plus graphing calculators do not draw pie charts.

Figure 5



To construct a pie chart by hand, we use a protractor to approximate the angles for each sector. See Figure 5.

Pie charts can be created only if all the categories of the variable under consideration are represented. For example, from the data given in Example 1, we could create a bar graph that lists the proportion of patients requiring rehabilitation on their back, shoulder or knee only, but it would not make sense to construct a pie chart for this situation. Do you see why? Only 70% of the data would be represented.

When should a bar graph be used to display information? When should a pie chart be used? Pie charts are useful for showing the division of all possible values of a qualitative variable into its parts. However, because angles are often hard to judge in pie charts, they are not as useful in comparing two specific values of the qualitative variable. Instead the emphasis is on comparing the part to the whole. Bar graphs are useful when we want to compare the different parts, not the parts to the whole. For example, if we wanted to get the “big picture” regarding educational attainment in 2003, then a pie chart is a good visual summary. However, if we want to compare bachelor’s degrees to high school diplomas, then a bar graph is a good visual summary.

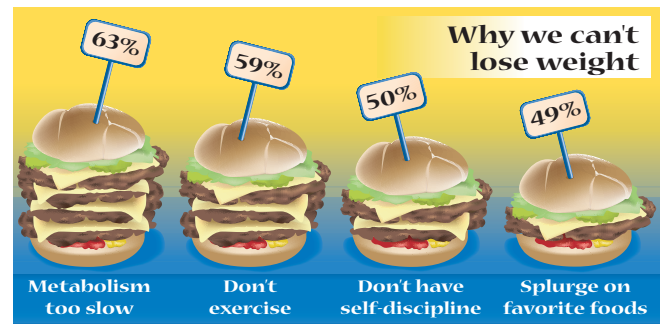
Now Work Problem 27(e).

2.1 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

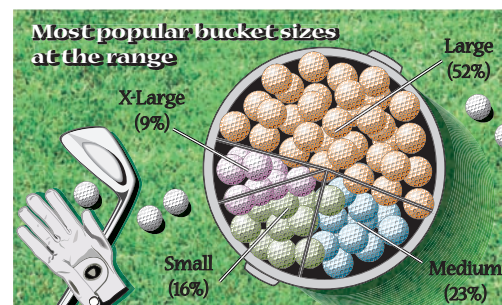
1. Define *raw data* in your own words.
2. A frequency distribution lists the _____ of occurrences of each category of data, while a relative frequency distribution lists the _____ of occurrences of each category of data.
3. When constructing a frequency distribution, why is it a good idea to add up the frequencies?
4. In a relative frequency distribution, what should the relative frequencies add up to?
5. What is a Pareto chart?
6. Why should relative frequencies be used when comparing two data sets?
7. Explain why Pareto charts might be preferred over bar graphs.

8. Consider the information in the “Why we can’t lose weight” chart shown below which is in the *USA Today* style of graph. Could the information provided be organized into a pie chart? Why or why not?

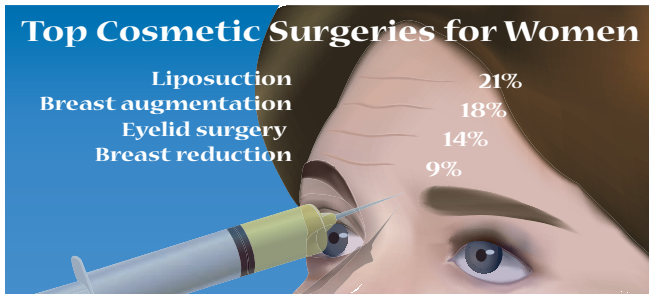


Skill Building

9. **Golf Driving Range** The pie chart at the right, the type we see in *USA Today*, depicts the bucket size golfers choose while at the driving range.
 - (a) What is the most popular size? What percentage of golfers choose this size?
 - (b) What is the least popular size? What percentage of golfers choose this size?
 - (c) What percentage of golfers choose a medium-sized bucket?



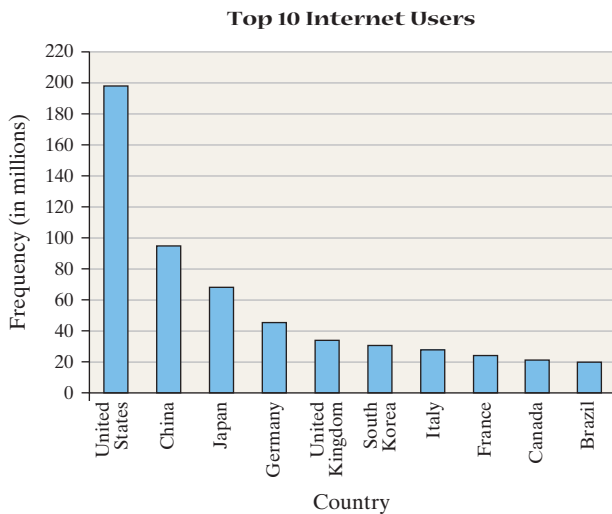
10. Cosmetic Surgery The *USA Today*-type chart shows the most frequent cosmetic surgeries for women in 2003.



- (a) If women had 1,548,000 cosmetic surgeries in 2003, how many of them were for liposuction?
- (b) How many were for eyelid surgery?
- (c) What percentage of surgeries are not accounted for in the graph?

11. Internet Users The following Pareto chart represents the top 10 countries in Internet users as of February 2005.

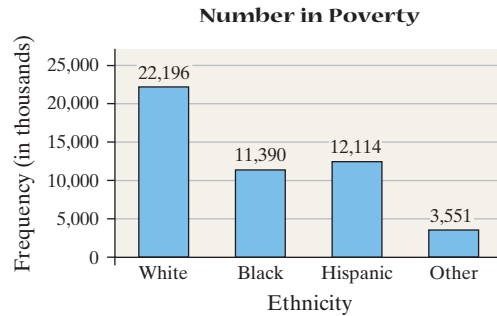
(Source: www.internetworldstats.com)



- (a) Which country has the most Internet users?
- (b) Approximately what is the Internet usage in Canada?
- (c) Approximately how many more users are in China than in Japan?

12. Poverty Every year the U.S. Census Bureau counts the number of people living in poverty. The bureau uses money income thresholds as its definition of poverty, so noncash benefits such as Medicaid and food stamps do not count toward poverty thresholds. For example, in 2002 the poverty threshold for a family of three was \$14,348. The

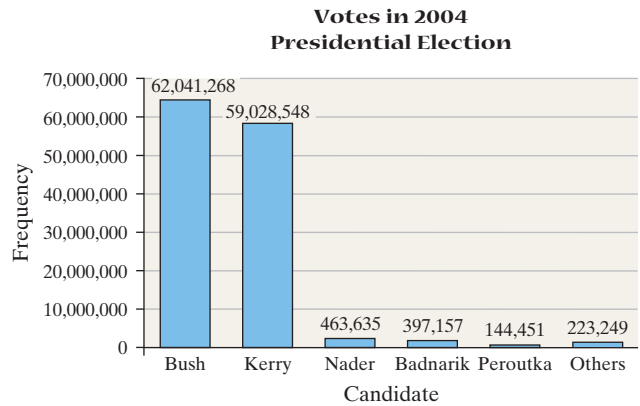
bar chart represents the number of people living in poverty in the United States in 2003, by ethnicity, based on March 2004 estimates.



Source: The Henry Kaiser Family Foundation (www.statehealthfacts.kff.org)

- (a) How many whites were living in poverty in 2003?
- (b) Of the impoverished, what percent were Hispanic?
- (c) How might this graph be misleading?

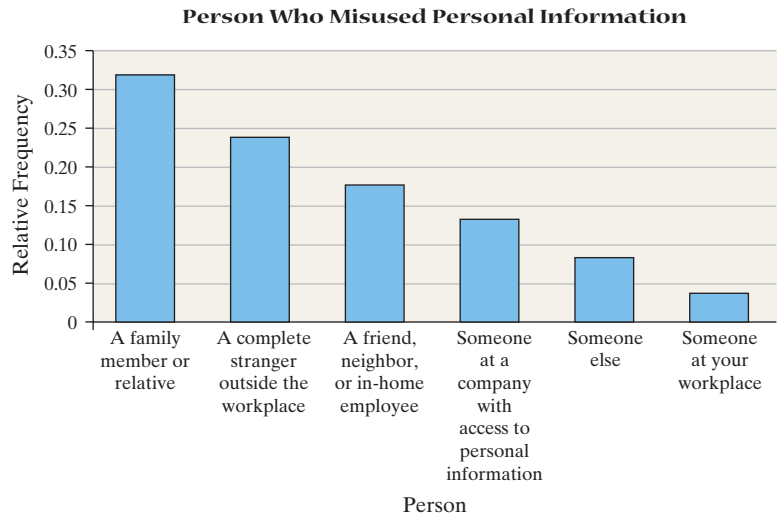
13. 2004 Presidential Election The following bar chart represents the number of voters who cast votes for the top five vote-getters and all others in the 2004 presidential election. The number of votes received for each candidate appears above the bar.



Source: David Leip's Atlas of U.S. Presidential Elections (www.uselectionatlas.org)

- (a) How many votes were cast for Michael Badnarik of the Libertarian Party?
- (b) What percent of the votes was cast for George W. Bush? For John F. Kerry?
- (c) To receive presidential election funds, a candidate must garner at least 5% of the vote. Will Ralph Nader receive presidential election funds?

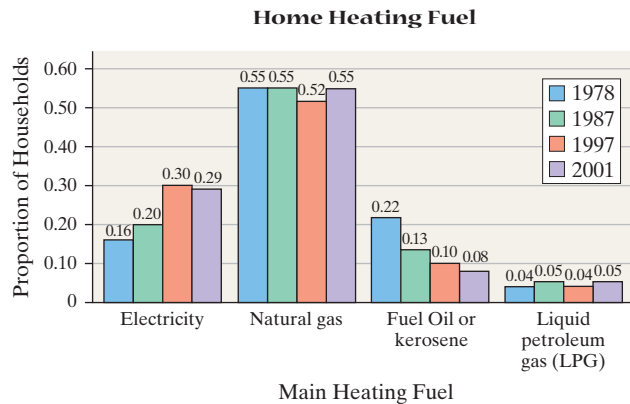
14. Identity Fraud In a study conducted by the Better Business Bureau and Javelin Strategy and Research, victims of identity fraud were asked, “Who was the person who misused your personal information?” The following Pareto chart represents the results for cases in the year 2004 for which the perpetrator’s identity was known.



Source: Javelin Strategy & Research, 2004.

- (a) Approximately what percentage of identity-fraud victims were victimized by a family member or relative?
- (b) If there were 9.3 million identity-fraud victims in 2004, how many were victimized by someone at a company with access to personal information?

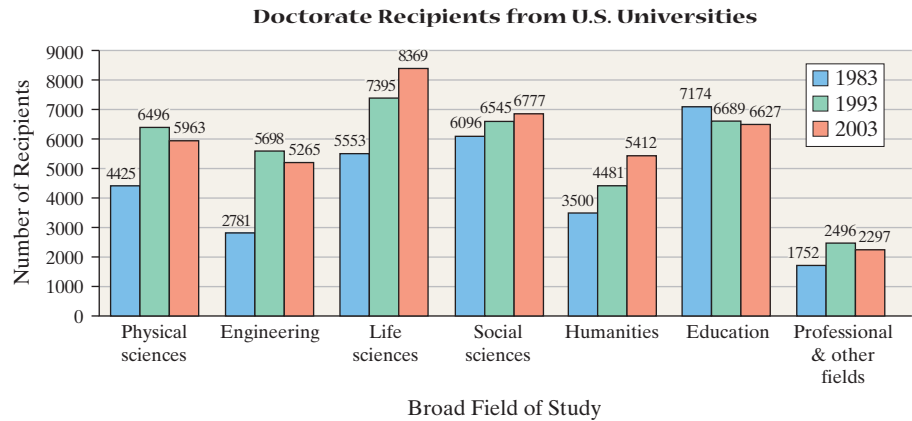
15. Home Heating Fuel The following side-by-side bar graph represents the proportion of households using four common sources of home heating fuel for the years 1978, 1987, 1997, and 2001.



Source: Energy Information Administration, Residential Energy Consumption Survey, 1978, 1987, 1997, 2001

- (a) What proportion of households used electricity as their main source of home heating fuel in 1978? in 1997?
- (b) What was the most popular source of home heating fuel in 1987?
- (c) If there were 107 million U.S. households in 2001, how many used LPG as their main source of home heating fuel?
- (d) What might account for the rise in homes that use electricity as the main source of home heating fuel?
- (e) Which source of home heating fuel has been decreasing rather steadily? What might account for this?
- (f) Which sources of home heating fuel have remained rather steady?

16. Doctorate Recipients The following side-by-side bar graph represents the number of doctorate recipients from U.S. universities within broad fields of study for the years 1983, 1993, and 2003.



Source: NSF/NIH/USED/NEH/USDA/NASA, 2003 Survey of Earned Doctorates

- How many more engineering doctorates were awarded in 1993 than in 2003?
- In 2003, what percentage of doctoral recipients received degrees in physical science? in education?
- What field of study consistently decreased in the number of doctoral degree recipients in the 3 years of the survey?
- Which field of study had the largest increase in the number of doctoral degree recipients in the 3 years of the survey?

Applying the Concepts

17. Government Income For fiscal year 2003 (October 1, 2002 to September 30, 2003), the federal government's income was \$1,782.3 billion. The various sources of income are broken down in the following table.

Source of Income	Amount (in billions of dollars)
Individual income taxes	793.7
Corporate income taxes	131.8
Social insurance taxes	713.0
Excise, estate and gift taxes, customs, and miscellaneous receipts	143.8

Source: Congressional Budget Office

- Construct a relative frequency distribution of the data shown.
- What percentage of total income is attributable to individual income taxes?
- Construct a frequency bar graph of the data.
- Construct a relative frequency bar graph of the data.
- Construct a pie chart of the data.

(f) In your opinion, which graph appears to place more emphasis on social insurance taxes as a source of income, the relative frequency bar graph or the pie chart? Why?

18. Government Expenditures For fiscal year 2003 (October 1, 2002 to September 30, 2003), the federal government spent \$2,158 billion. The breakdown of expenditures is given in the following table.

Category	Expenditure (in billions of dollars)
National defense and foreign affairs	474.8
Social programs	1251.6
Physical, human, and community development	215.8
Net interest on the debt	151.1
Law enforcement and general government	64.7

Source: Budget of the United States and Internal Revenue Service

- Construct a relative frequency distribution of the data shown.

- (b) What percentage of total expenditures is attributable to net interest on debt?
- (c) Construct a frequency bar graph of government expenditures.
- (d) Construct a relative frequency bar graph of government expenditures.
- (e) Construct a pie chart of government expenditures.
- (f) In your opinion, which graph appears to place more emphasis on net interest on the debt, the relative frequency bar graph or the pie chart? Why?

- 19. College Survey** In a national survey conducted by the Centers for Disease Control to determine health-risk behaviors among college students, college students were asked, “How often do you wear a seat belt when riding in a car driven by someone else?” The frequencies were as follows:



Response	Frequency
Never	125
Rarely	324
Sometimes	552
Most of the time	1257
Always	2518

- (a) Construct a relative frequency distribution.
- (b) What percentage of respondents answered “Always”?
- (c) What percentage of respondents answered “Never” or “Rarely”?
- (d) Construct a frequency bar graph.
- (e) Construct a relative frequency bar graph.
- (f) Construct a pie chart.
- (g) Suppose that a representative from the Centers for Disease Control says, “52.7% of all college students never wear a seat belt.” Is this a descriptive or inferential statement?

- 20. College Survey** In a national survey conducted by the Centers for Disease Control to determine health-risk behaviors among college students, college students were asked, “How often do you wear a seat belt when driving a car?” The frequencies were as follows:



Response	Frequency
I do not drive a car	249
Never	118
Rarely	249
Sometimes	345
Most of the time	716
Always	3093

- (a) Construct a relative frequency distribution.

- (b) What percentage of respondents answered “Always”?
- (c) What percentage of respondents answered “Never” or “Rarely”?
- (d) Construct a frequency bar graph.
- (e) Construct a relative frequency bar graph.
- (f) Construct a pie chart.
- (g) Compute the relative frequencies of “Never,” “Rarely,” “Sometimes,” “Most of the time,” and “Always,” excluding those that do not drive. Compare with those in Problem 19. What might you conclude?
- (h) Suppose that a representative from the Centers for Disease Control says, “2.5% of the college students in this survey responded that they never wear a seat belt.” Is this a descriptive or inferential statement?

- 21. Foreign-Born Population** The following data represent the region of birth of foreign-born residents of the United States in 2003.



Region	Number (thousands)
Caribbean	3,384
Central America	12,362
South America	2,111
Asia	8,375
Europe	4,590
Other Regions	2,680

Source: U.S. Census Bureau

- (a) Construct a relative frequency distribution.
- (b) What percentage of foreign-born residents was born in Asia?
- (c) Construct a frequency bar graph.
- (d) Construct a relative frequency bar graph.
- (e) Construct a pie chart.

- 22. Robbery** The following data represent the number of offenses for various robberies in 2003.



Type of Robbery	Number (thousands)
Street or highway	131
Commercial	61
Gas station	10
Convenience store	26
Residence	41
Bank	7

Source: U.S. Federal Bureau of Investigation

- (a) Construct a relative frequency distribution.
- (b) What percentage of robberies was of gas stations?
- (c) Construct a frequency bar graph.
- (d) Construct a relative frequency bar graph.
- (e) Construct a pie chart.

23. Educational Attainment On the basis of the 2003 Current Population Survey, there were 88.7 million males and 96.6 million females 25 years old or older in the United States. The educational attainment of the males and females was as follows:



Educational Attainment	Males (in millions)	Females (in millions)
Not a high school graduate	14.1	14.5
High school graduate	27.4	31.9
Some college, but no degree	15.2	16.6
Associate's degree	6.4	8.8
Bachelor's degree	16.4	16.9
Advanced degree	9.2	7.9

Source: U.S. Census Bureau

- (a) Construct a relative frequency distribution for males.
- (b) Construct a relative frequency distribution for females.
- (c) Construct a side-by-side relative frequency bar graph.
- (d) Compare each gender's educational attainment. Make a conjecture about the reasons for the differences.

24. Internet Access The following data represent the number of people who had Internet access in the years 2000 and 2003 by level of education. Data are in thousands of U.S. residents.



Educational Attainment	2000	2003
No college	24,662	65,862
Some college	31,462	50,931
Graduated college	34,379	49,106

Source: U.S. Statistical Abstract, 2004

- (a) Construct a relative frequency distribution for 2000.
- (b) Construct a relative frequency distribution for 2003.
- (c) Construct a side-by-side relative frequency bar graph.
- (d) Compare each year's Internet access. Make some conjectures about the reasons for any differences or similarities.

25. Murder Victims A criminologist wanted to know if there was any relation between age and gender of murder victims. The following data represent the number of male and female murder victims by age in 2002.



Age	Number of Males	Number of Females
Less than 17	650	444
17–24	3435	704
25–34	2990	704
35–54	2859	1085
55 or older	787	442

Source: U.S. Federal Bureau of Investigation

- (a) Construct a relative frequency distribution for males.
- (b) Construct a relative frequency distribution for females.
- (c) Construct a side-by-side relative frequency bar graph.
- (d) Compare each gender's age percentages. Make a conjecture about the reasons for the differences or similarities.

26. Car Color DuPont Automotive is a major supplier of paint to the automotive industry. It conducted a survey of 100 randomly selected autos in the luxury car segment and 100 randomly selected autos in the sports car segment that were recently purchased and obtained the following colors.



Color	Number of Luxury Cars	Number of Sports Cars
Silver	32	27
White	30	11
Blue	9	13
Black	9	16
Gray	7	7
Red	6	18
Gold	3	4
Green	2	0
Brown	2	4

- (a) Construct a relative frequency distribution for each car type.
- (b) Draw a side-by-side relative frequency bar graph.
- (c) Compare the colors for the two types of car. Make a conjecture about the reasons for the differences.

27. 2004 Presidential Election An exit poll was conducted in Los Alamos County, New Mexico, in which a random sample of 40 voters revealed whom they voted for in the presidential election. The results of the survey are as follows:



Kerry	Kerry	Bush	Bush
Bush	Kerry	Kerry	Bush
Kerry	Bush	Kerry	Bush
Bush	Bush	Kerry	Kerry
Kerry	Bush	Bush	Kerry
Badnarik	Bush	Kerry	Bush
Kerry	Bush	Kerry	Bush
Bush	Bush	Kerry	Kerry
Bush	Bush	Bush	Nader
Bush	Kerry	Bush	Kerry

- (a) Construct a frequency distribution.
- (b) Construct a relative frequency distribution.
- (c) Construct a frequency bar graph.
- (d) Construct a relative frequency bar graph.
- (e) Construct a pie chart.
- (f) On the basis of the data, make a conjecture about which candidate will win Los Alamos County. Would your conjecture be descriptive statistics or inferential statistics? If George W. Bush wins Los Alamos County, what conclusions might be drawn, assuming that the sample was conducted appropriately? Would you be confident in making this prediction with a sample of 40? If the sample consisted of 100 voters, would your confidence increase? Why?

28. Hospital Admissions The following data represent the diagnoses of a random sample of 20 patients admitted to a hospital.



Cancer	Motor vehicle accident	Congestive heart failure
Gunshot wound	Fall	Gunshot wound
Gunshot wound	Motor vehicle accident	Gunshot wound
Assault	Motor vehicle accident	Gunshot wound
Motor vehicle accident	Motor vehicle accident	Gunshot wound
Motor vehicle accident	Gunshot wound	Motor vehicle accident
Fall	Gunshot wound	

Source: Tamela Ohm, student at Joliet Junior College

- (a) Construct a frequency distribution.
- (b) Construct a relative frequency distribution.
- (c) Which diagnosis had the most admissions?
- (d) What percentage of diagnoses was motor vehicle accidents?
- (e) Construct a frequency bar graph.
- (f) Construct a relative frequency bar graph.
- (g) Construct a pie chart.
- (h) Suppose that an admission specialist at the hospital stated that 40% of all admissions were gunshot wounds. Would this statement be descriptive or inferential? Why?

29. Which Position in Baseball Pays the Most? You are a prospective baseball agent and are in search of clients. You would like to recruit the highest-paid players as clients, so you perform a study in which you identify the 24 top-paid players for the 2004 season and their positions. The table shows the results of your study.



Player	Position
Manny Ramirez	Left field
Alex Rodriguez	Third base
Carlos Delgado	First base
Derek Jeter	Shortstop
Barry Bonds	Left field
Pedro Martinez	Pitcher
Mo Vaughn	First base
Shawn Green	Right field
Mike Piazza	Catcher
Jeff Bagwell	First base
Randy Johnson	Pitcher
Mike Mussina	Pitcher
Sammy Sosa	Right field
Kevin Brown	Pitcher
Chipper Jones	Third base
Mike Hampton	Pitcher
Magglio Ordonez	Right field
Chan Ho Park	Pitcher
Gary Sheffield	Right field
Larry Walker	Right field
Ken Griffey, Jr.	Center field
Richard Hidalgo	Right field
Andruw Jones	Center field
Matt Morris	Pitcher

Source: usatoday.com

- (a) Construct a frequency distribution of position.
- (b) Construct a relative frequency distribution of position.
- (c) Which position appears to be the most lucrative? For which position would you recruit?
- (d) Are there any positions that you would avoid recruiting? Why?
- (e) Draw a frequency bar graph.
- (f) Draw a relative frequency bar graph.
- (g) Draw a pie chart.

30. Blood Type A phlebotomist draws the blood of a random sample of 50 patients and determines their blood types as shown:



O	O	A	A	O
B	O	B	A	O
AB	B	A	B	AB
O	O	A	A	O
AB	O	A	B	A
O	A	A	O	A
O	A	O	AB	A
O	B	A	A	O
O	O	O	A	O
O	A	O	A	O

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- According to the data, which blood type is most common?
- According to the data, which blood type is least common?
- Use the results of the sample to conjecture the percentage of the population that has type O blood. Is this an example of descriptive or inferential statistics?

- Contact a local hospital and ask them the percentage of the population that is blood type O. Why might the results differ?
- Draw a frequency bar graph.
- Draw a relative frequency bar graph.
- Draw a pie chart.

31. Foreign Language According to the Modern Language Association, the number of college students studying foreign language is increasing. The following data represent the foreign language being studied based on a simple random sample of 30 students learning a foreign language.



Spanish	Chinese	Spanish	Spanish	Spanish
Chinese	German	Spanish	Spanish	French
Spanish	Spanish	Japanese	Latin	Spanish
German	German	Spanish	Italian	Spanish
Italian	Japanese	Chinese	Spanish	French
Spanish	Spanish	Russian	Latin	French

Source: Based on data obtained from the Modern Language Association

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- Construct a frequency bar graph.
- Construct a relative frequency bar graph.
- Construct a pie chart.

32. The following table lists the presidents of the United States and their state of birth.



Birthplace of U.S. President

President	State of Birth	President	State of Birth	President	State of Birth
Washington	Virginia	Lincoln	Kentucky	Coolidge	Vermont
J. Adams	Massachusetts	A. Johnson	North Carolina	Hoover	Iowa
Jefferson	Virginia	Grant	Ohio	F. D. Roosevelt	New York
Madison	Virginia	Hayes	Ohio	Truman	Missouri
Monroe	Virginia	Garfield	Ohio	Eisenhower	Texas
J. Q. Adams	Massachusetts	Arthur	Vermont	Kennedy	Massachusetts
Jackson	South Carolina	Cleveland	New Jersey	L. B. Johnson	Texas
Van Buren	New York	B. Harrison	Ohio	Nixon	California
W. H. Harrison	Virginia	Cleveland	New Jersey	Ford	Nebraska
Tyler	Virginia	McKinley	Ohio	Carter	Georgia
Polk	North Carolina	T. Roosevelt	New York	Reagan	Illinois
Taylor	Virginia	Taft	Ohio	George H. Bush	Massachusetts
Fillmore	New York	Wilson	Virginia	Clinton	Arkansas
Pierce	New Hampshire	Harding	Ohio	George W. Bush	Connecticut
Buchanan	Pennsylvania				

- Construct a frequency bar graph for state of birth.
- Which state has yielded the most presidents?
- Explain why the answer obtained in part (b) may be considered to be misleading.



Consumer Reports Rates Treadmills

A study that compared exercisers who worked out equally hard for the same time on several different types of machines found that they generally burned the most calories on treadmills. Our own research has shown that treadmills are less likely than other machines to sit unused. So it should come as no surprise that treadmills are the best-selling home exercise machine in the United States.

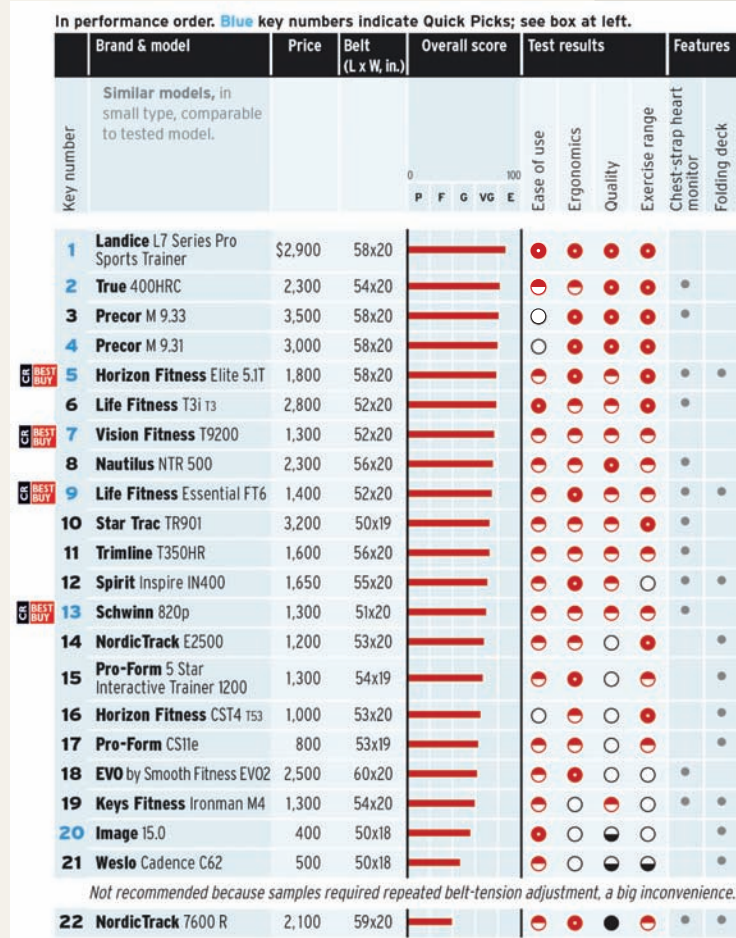
In a study by *Consumer Reports* (January 2005), we tested 22 treadmills from best-selling brands, ranging in price from \$400 to \$3,500. The treadmills were rated on ease of use, ergonomics, quality, and exercise range. Ease of use was based on how easy it was to adjust exercise difficulty, read the display, and operate the monitor functions. Ergonomics was based on how well the machine accommodates users of different sizes, including belt sizes, foot rail, and hand grip design. Quality was based on our judgment of construction quality, sample defects, and the results of our durability test. Exercise range was based on how well each machine pro-

vides an effective workout to users of various fitness levels, from beginners to very fit individuals.

In order to help compare the treadmills, these attribute scores were combined into an overall score. The figure below is a ratings chart for the 22 treadmills based on our test results. In addition to the performance ratings, other useful information, such as the models' price and belt size, is included.

- What type of graph is illustrated to display overall score in the figure?
- Which model has the highest construction score? Which models have the lowest ease of use score?
- For ease of use, how many treadmills rated excellent? very good? good? fair? poor?
- Draw a frequency bar graph for each rating category. In other words, draw a bar graph for ease of use, ergonomics, and so on.

Ratings Chart for Threadmills



Source: Used with the permission of *Consumer Reports*

TI-83/84 Plus The TI-83 or TI-84 Plus does not have the ability to draw bar graphs or pie charts.

MINITAB **Frequency or Relative Frequency Distributions from Raw Data**

Step 1: Enter the raw data in C1.

Step 2: Select **Stat** and highlight **Tables** and select **Tally Individual Variables . . .**

Step 3: Fill in the window with appropriate values. In the “Variables” box, enter C1. Check “counts” for a frequency distribution and/or “percents” for a relative frequency distribution. Click OK.

Bar Graphs from Summarized Data

Step 1: Enter the categories in C1 and the frequency or relative frequency in C2.

Step 2: Select **Graph** and highlight **Bar Chart**.

Step 3: In the “Bars represent” pull-down menu, select “Values from a table” and highlight “Simple.” Press OK.

Step 4: Fill in the window with the appropriate values. In the “Graph variables” box, enter C2. In the “Categorical variable” box, enter C1. By pressing Labels, you can add a title to the graph. Click OK to obtain the bar graph.

Bar Graphs from Raw Data

Step 1: Enter the raw data in C1.

Step 2: Select **Graph** and highlight **Bar Chart**.

Step 3: In the “Bars represent” pull-down menu, select “Counts of unique values” and highlight “Simple.” Press OK.

Step 4: Fill in the window with the appropriate values. In the “Categorical variable” box, enter C1. By pressing Labels, you can add a title to the graph. Click OK to obtain the bar graph.

Pie Chart from Raw or Summarized Data

Step 1: If the data are in a summarized table, enter the categories in C1 and the frequency or relative frequency in C2. If the data are raw, enter the data in C1.

Step 2: Select **Graph** and highlight **Pie Chart**.

Step 3: Fill in the window with the appropriate values. If the data are summarized, click the “Chart values from a table” radio button; If the data are raw, click the “Chart raw data” radio button. For summarized data, enter C1 in the “Categorical variable” box and C2 in the “Summary variable” box. If the data are raw, enter C1 in the “Categorical variable” box. By pressing Labels, you can add a title to the graph. Click OK to obtain the pie chart.

Excel **Bar Graphs from Summarized Data**

Step 1: Enter the categories in column A and the frequency or relative frequency in column B.

Step 2: Select the chart wizard icon. Click the “column” chart type. Select the chart type in the upper-left-hand corner and hit “Next.”

Step 3: Click inside the data range cell. Use the mouse to highlight the data to be graphed. Click “Next.”

Step 4: Click the “Titles” tab to enter *x*-axis, *y*-axis, and chart titles. Click “Finish.”

Pie Charts from Summarized Data

Step 1: Enter the categories in column A and the frequencies in column B. Select the chart wizard icon and click the “pie” chart type. Select the pie chart in the upper-left-hand corner.

Step 2: Click inside the data range cell. Use the mouse to highlight the data to be graphed. Click “Next.”

Step 3: Click the “Titles” tab to the chart title. Click the “Data Labels” tab and select “Show label and percent.” Click “Finish.”

2.2 Organizing Quantitative Data

Preparing for This Section Before getting started, review the following:

- Quantitative data (Section 1.1, p. 6)
- Discrete data (Section 1.1, p. 7)
- Continuous data (Section 1.1, p. 7)

Objectives

- 1 Organize discrete data in tables
- 2 Construct histograms of discrete data
- 3 Organize continuous data in tables
- 4 Construct histograms of continuous data
- 5 Draw stem-and-leaf plots
- 6 Draw dot plots
- 7 Identify the shape of a distribution
- 8 Draw time-series graphs

The first step in summarizing quantitative data is to determine whether the data are discrete or continuous. If the data are discrete and there are relatively few different values of the variable, the categories of data will be the observations (as in qualitative data). If the data are discrete, but there are many different values of the variables or, if the data are continuous, categories of data (called *classes*) must be created using intervals of numbers. We will first present the techniques required to organize discrete quantitative data when there are relatively few different values and then proceed to organizing continuous quantitative data.

1 Organize Discrete Data in Tables

The values of a discrete variable are used to create the categories of data.

EXAMPLE 1

Constructing Frequency and Relative Frequency Distributions from Discrete Data

Problem: The manager of a Wendy's fast-food restaurant is interested in studying the typical number of customers who arrive during the lunch hour. The data in Table 8 represent the number of customers who arrive at Wendy's for 40 randomly selected 15-minute intervals of time during lunch. For example, during one 15-minute interval, seven customers arrived. Construct a frequency and relative frequency distribution.



Table 8

Number of Arrivals at Wendy's				
7	6	6	6	4
5	6	6	11	4
2	7	1	2	4
6	5	5	3	7
2	2	9	7	5
6	2	6	5	7
6	8	2	6	5
4	6	9	8	5

Approach: The number of people arriving could be 0, 1, 2, 3, From Table 8, we see that there are 11 categories of data from this study: 1, 2, 3, . . . , 11. We tally the number of observations for each category, add up each tally, and create the frequency and relative frequency distributions.

Solution: The frequency and relative frequency distributions are shown in Table 9.

Number of Customers	Tally	Frequency	Relative Frequency
1		1	$\frac{1}{40} = 0.025$
2		6	0.15
3		1	0.025
4		4	0.1
5		7	0.175
6		11	0.275
7		5	0.125
8		2	0.05
9		2	0.05
10		0	0.0
11		1	0.025

On the basis of the relative frequencies, 27.5% of the 15-minute intervals had six customers arrive at Wendy's during the lunch hour. ■

Now Work Problems 27(a)–(d).



Construct Histograms of Discrete Data

As with qualitative data, quantitative data may also be represented graphically. We begin our discussion with a graph called the *histogram*, which is similar to the bar graph drawn for qualitative data.

Definition

A **histogram** is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. The width of each rectangle is the same and the rectangles touch each other.

EXAMPLE 2

Drawing a Histogram for Discrete Data

Problem: Construct a frequency histogram and a relative frequency histogram using the data summarized in Table 9.

Approach: On the horizontal axis, we place the value of each category of data (number of customers). The vertical axis will be the frequency or relative frequency of each category. Rectangles of equal width are drawn, with the center of each rectangle located at the value of each category. For example, the first rectangle is centered at 1. For the frequency histogram, the height of the rectangle will be the frequency of the category. For the relative frequency histogram, the height of the rectangle will be the relative frequency of the category. Remember, the rectangles touch for histograms.

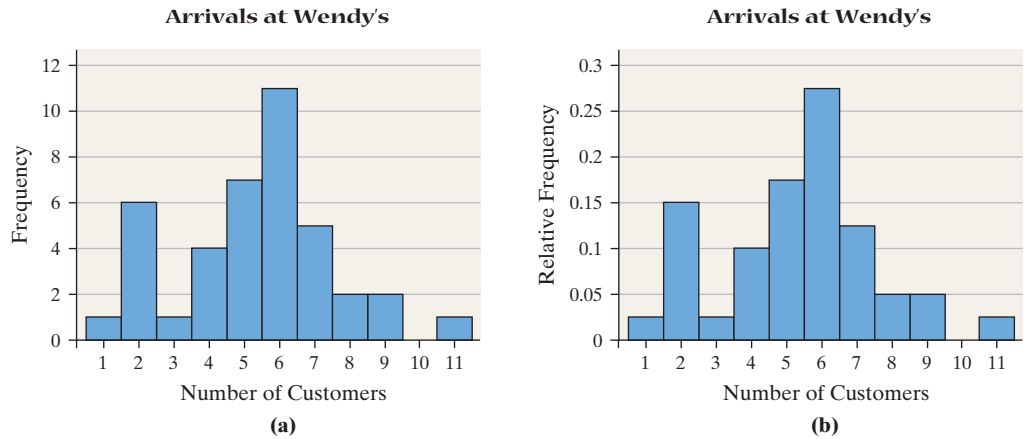
Solution: Figure 6(a) shows the frequency histogram. Figure 6(b) shows the relative frequency histogram.



CAUTION

The rectangles in histograms touch, while the rectangles in bar graphs do not touch.

Figure 6



Now Work Problems 27(e) and (f).

3 Organize Continuous Data in Tables

Raw continuous data do not have any predetermined categories that can be used to construct a frequency distribution. Therefore, the categories must be created. Categories of data are created by using intervals of numbers called **classes**.

Table 10 is a typical frequency distribution created from continuous data. The data represent the number of U.S. residents between the ages of 25 and 74 who have earned a bachelor’s degree. The data are based on the Current Population Survey conducted in 2003.

In the table, we notice that the data are categorized, or grouped, by intervals of numbers. Each interval represents a class. For example, the first class is 25- to 34-year-old residents of the United States who have a bachelor’s degree. We read this interval as follows: “The number of residents of the United States in 2003 who were between 25 and 34 years of age and have a bachelor’s degree was 8,849,000.” There are five classes in the table, each with a *lower bound* and an *upper bound*. The **lower class limit** of a class is the smallest value within the class, while the **upper class limit** of a class is the largest value within the class. The lower class limit for the first class in Table 10 is 25; the upper class limit is 34. The **class width** is the difference between consecutive lower class limits. The class width for the data in Table 10 is $35 - 25 = 10$.

The classes in Table 10 do not overlap, so there is no confusion as to which class a data value belongs. Notice that the class widths are equal for all classes. One exception to this requirement is in open-ended tables. A table is **open ended** if the first class has no lower class limit or the last class does not have an upper class limit. The data in Table 11 represent the number of persons under sentence of death as of December 31, 2003, in the United States. The last class in the table, “60 and older,” is open ended.



Table 10

Age	Number (in thousands)
25–34	8849
35–44	8915
45–54	7576
55–64	4254
65–74	2047



Table 11

Age	Number
20–29	533
30–39	1147
40–49	1090
50–59	493
60 and older	110

Source: U.S. Justice Department

EXAMPLE 3

Organizing Continuous Data into a Frequency and Relative Frequency Distribution

Problem: Suppose you are considering investing in a Roth IRA. You collect the data in Table 12, which represent the three-year rate of return (in percent) for a simple random sample of 40 small-capitalization growth mutual funds. Construct a frequency and relative frequency distribution of the data.

Approach: To construct a frequency distribution, we first create classes of equal width. There are 40 observations in Table 12, and they range from 0.05 to 14.48, so we decide to create the classes such that the lower class limit of the first



In Other Words

For qualitative and many discrete data, the categories of data are formed by using the data. For continuous data, the categories are formed by using an interval of numbers, such as 30–39.



Table 12

Three-Year Rate of Return of Mutual Funds

5.37	4.31	4.13	8.58	5.99	7.90	9.11	6.11
3.06	14.48	12.50	8.33	10.10	8.21	6.83	10.94
2.34	0.97	8.33	8.89	6.07	6.50	5.99	9.38
0.05	13.88	3.71	10.07	9.88	4.93	6.38	10.34
2.27	11.91	11.69	12.06	9.84	7.75	2.86	6.68

Source: Morningstar.com


CAUTION

Watch out for tables with class widths that overlap, such as a first class of 20–30 and a second class of 30–40.

class is 0 (a little smaller than the smallest data value) and the class width is 2. There is nothing magical about the choice of 2 as a class width. We could have selected a class width of 8 (or any other class width, as well). We choose a class width that we think will nicely summarize the data. If our choice doesn't accomplish this, we can always try another one. The lower class limit of the second class will be $0 + 2 = 2$. Because the classes must not overlap, the upper class limit of the first class is 1.99. Continuing in this fashion, we obtain the following classes:

0–1.99
2–3.99
⋮
14–15.99

This gives us eight classes. We tally the number of observations in each class, add up the tallies, and create the frequency distribution. The relative frequency distribution would be created by dividing each class's frequency by 40, the number of observations.

Solution: We tally the data as shown in the second column of Table 13. The third column in the table shows the frequency of each class. From the frequency distribution, we conclude that a three-year rate of return between 8.0% and 9.99% occurs with the most frequency. The fourth column in the table shows the relative frequency of each class. So 22.5% of the small-capitalization growth mutual funds had a three-year rate of return between 8% and 9.99%.

**Historical Note**

Florence Nightingale was born in Italy on May 12, 1820. She was named after the city of her birth. Nightingale was educated by her father, who attended Cambridge University. Between 1849 and 1851, she studied nursing throughout Europe. In 1854, she was asked to oversee the introduction of female nurses into the military hospitals in Turkey. While there, she greatly improved the mortality rate of wounded soldiers. She collected data and invented graphs (the polar area diagram), tables, and charts to show that improving sanitary conditions would lead to decreased mortality rates. In 1869, Nightingale founded the Nightingale School Home for Nurses. After a long and eventful life as a reformer of health care and contributor to graphics in statistics, Florence Nightingale died on August 13, 1910.

Table 13

Class (3-Year Rate of Return)	Tally	Frequency	Relative Frequency
0–1.99		2	$2/40 = 0.05$
2–3.99		5	$5/40 = 0.125$
4–5.99		6	$6/40 = 0.15$
6–7.99		8	0.2
8–9.99		9	0.225
10–11.99		6	0.15
12–13.99		3	0.075
14–15.99		1	0.025

Only one mutual fund had a three-year rate of return between 14% and 15.99%. We might consider this mutual fund worthy of our investment. This type of information would be more difficult to obtain from the raw data. ■

Notice that the choice of the lower class limit of the first class and the class width was rather arbitrary. While formulas and procedures do exist for creating frequency distributions from raw data, they do not necessarily provide better summaries. It is incorrect to say that a particular frequency distribution is the correct one. Constructing frequency distributions is somewhat of an art form in

which the distribution that seems to provide the best overall summary of the data should be used.

Consider the frequency distribution in Table 14, which also summarizes the three-year rate of return data discussed in Example 3. Here, the lower class limit of the first class is 0 and the class width is 4. Do you think Table 13 or Table 14 provides a better summary of the distribution of three-year rates of return? In forming your opinion, consider the following: Too few classes will cause a bunching effect. Too many classes will spread the data out, thereby not revealing any pattern.



In Other Words

Creating the classes for summarizing continuous data is an art form. There is no such thing as the correct frequency distribution. However, there can be less desirable frequency distributions. The larger the class width, the fewer classes a frequency distribution will have.

Class	Tally	Frequency
0–3.99		7
4–7.99		14
8–11.99		15
12–15.99		4

The goal in constructing a frequency distribution is to reveal interesting features of the data. With that said, when constructing frequency distributions, we typically want the number of classes to be between 5 and 20. When the data set is small, we want fewer classes. When the data set is large, we want more classes. Why do you think this is reasonable?

Now Work Problems 29(a) and (b).

4 Construct Histograms of Continuous Data

We are now ready to draw histograms of continuous data.

EXAMPLE 4

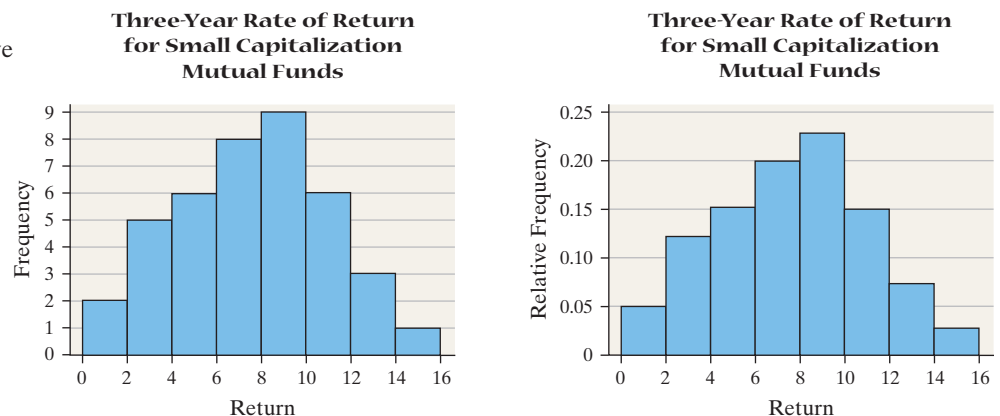
Drawing a Histogram of Continuous Data

Problem: Construct a frequency and relative frequency histogram of the three-year rate of return data discussed in Example 3.

Approach: To draw the frequency histogram, we will use the frequency distribution in Table 13. We label the lower class limits of each class on the horizontal axis. Then, for each class, we draw a rectangle whose width is the class width and whose height is the frequency. To construct the relative frequency histogram, we let the height of the rectangle be the relative frequency, instead of the frequency.

Solution: Figure 7(a) represents the frequency histogram, and Figure 7(b) represents the relative frequency histogram.

Figure 7
Frequency and Relative
Frequency Histograms



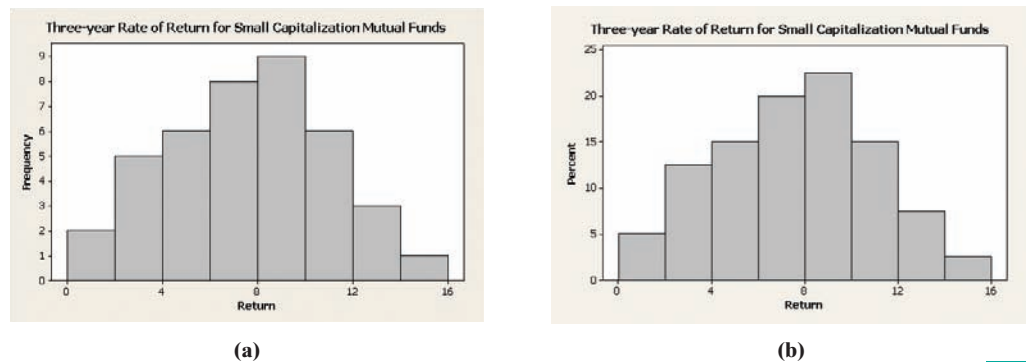
EXAMPLE 5**Drawing a Histogram for Continuous Data Using Technology**

Problem: Construct a frequency and relative frequency histogram of the three-year rate of return data discussed in Example 3.

Approach: We will use MINITAB to construct the frequency and relative frequency histograms. The steps for constructing the graphs using the TI-83/84 Plus graphing calculators, MINITAB, and Excel are given in the Technology Step by Step on page 92.

Result: Figure 8(a) shows the frequency histogram and Figure 8(b) shows the relative frequency histogram obtained from MINITAB. Note that MINITAB expresses relative frequencies using percent.

Figure 8



Now Work Problems 29(c)–(d).

5**Draw Stem-and-Leaf Plots**

A **stem-and-leaf plot** is another way to represent quantitative data graphically. Stem-and-leaf plots have some advantages over histograms. Use the following steps to construct a stem-and-leaf plot.

Construction of a Stem-and-Leaf Plot

Step 1: The **stem** of the graph will consist of the digits to the left of the rightmost digit. The **leaf** of the graph will be the rightmost digit. Sometimes it is necessary to modify the method of choosing the stem if a different class width is desired.

Step 2: Write the stems in a vertical column in increasing order. Draw a vertical line to the right of the stems.

Step 3: Write each leaf corresponding to the stems to the right of the vertical line.

Step 4: Write the leaves in ascending order.

EXAMPLE 6**Constructing a Stem-and-Leaf Plot**

Problem: The data in Table 15 represent the percentage of persons in poverty, by state, in 2002. Draw a stem-and-leaf plot of the data.

Approach:

Step 1: We will treat the integer portion of the number as the stem and the decimal portion as the leaf. For example, the stem of Alabama will be 14 and the leaf will be 6. The stem of 14 will include all data from 14.0 to 14.9.

Step 2: Write the stems vertically in ascending order, and then draw a vertical line to the right of the stems.



USING TECHNOLOGY

In MINITAB, there is a column of numbers left of the stem. The (6) indicates that there are 6 observations in the class containing the middle value (called the *median*). The values above the (6) represent the number of observations less than or equal to the upper class limit of the class. For example, 12 states have percentage in poverty less than 8.9. The values in the left column below the (6) indicate the number of observations greater than or equal to the lower class limit of the class. For example, 7 states have percentages in poverty greater than 15.0.

Approach: We will use MINITAB to construct the stem-and-leaf plot. The steps for constructing the graphs using MINITAB are given in the Technology Step by Step on page 92. **Note:** The TI graphing calculators and Excel are not capable of drawing stem-and-leaf plots.

Result: Figure 10 shows the stem-and-leaf plot obtained from MINITAB.

Figure 10

1	5	6
2	6	5
5	7	388
12	8	1333677
21	9	234455669
(6)	10	123368
24	11	22389
19	12	118
16	13	11357
11	14	0267
7	15	3
6	16	08
4	17	068
1	18	0

Now Work Problem 35.

If you look at the stem-and-leaf plot carefully, you'll notice that it looks much like a histogram turned on its side. The stem serves as the class. For example, the stem 10 contains all data from 10.0 to 10.9. The leaf serves as the frequency (height of the rectangle). Therefore, it is important to space the leaves equally when drawing a stem-and-leaf plot.

One advantage of the stem-and-leaf plot over frequency distributions and histograms is that the raw data can be retrieved from the stem-and-leaf plot.

Once a frequency distribution or histogram of continuous data is created, the raw data are lost. However, the raw data can be retrieved from the stem-and-leaf plot.

The steps listed for creating stem-and-leaf plots sometimes must be modified to meet the needs of the data. Consider the next example.



In Other Words

The choice of the stem in the construction of a stem-and-leaf diagram is also an art form. It acts just like the class width. For example, the stem of 7 in Figure 10 represents the class 7.0–7.9. The stem of 8 represents the class 8.0–8.9. Notice that the class width is 1.0. The number of leaves is the frequency of each category.

EXAMPLE 8

Constructing a Stem-and-Leaf Plot after Modifying the Data

Problem: Construct a stem-and-leaf plot of the three-year rate of return data listed in Table 12 on page 74.

Approach

Step 1: If we follow the approach from Example 6 and use the integer portion as the stem and the decimals as the leaves, the stems will be 0, 1, 2, ..., 14. This is fine. However, the leaves will be two digits (such as 37, 06, and so on). This is not acceptable. To address this problem, we will round the data to the nearest tenth. Then the stem can be the whole numbers 0, 1, 2, ..., 14, and the leaves will be the decimal portion.

Step 2: Create a vertical column of the whole-number stems in increasing order.

Step 3: Write the leaves corresponding to each stem.

Step 4: Write the leaves in ascending order.

Solution:

Step 1: We round the data to the nearest tenth as shown in Table 16.



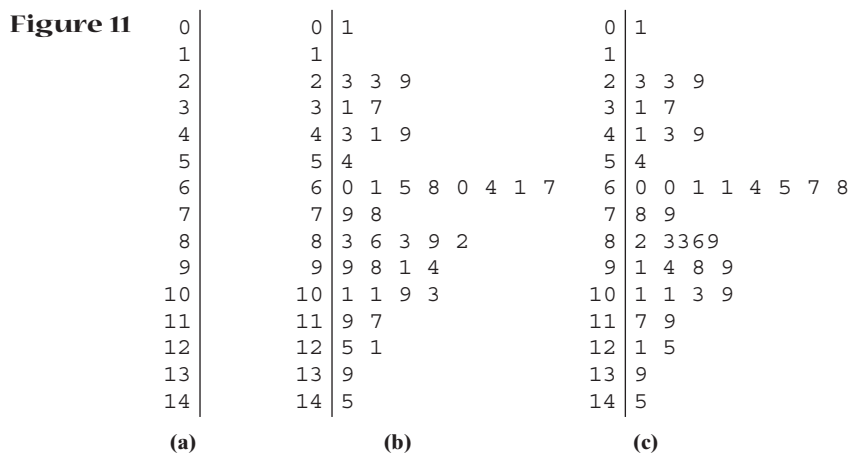
Table 16

5.4	4.3	4.1	8.6	6.0	7.9	9.1	6.1
3.1	14.5	12.5	8.3	10.1	8.2	6.8	10.9
2.3	1.0	8.3	8.9	6.1	6.5	6.0	9.4
0.1	13.9	3.7	10.1	9.9	4.9	6.4	10.3
2.3	11.9	11.7	12.1	9.8	7.8	2.9	6.7

Step 2: Write the stems vertically in ascending order as shown in Figure 11(a).

Step 3: Write the leaves corresponding to each stem as shown in Figure 11(b).

Step 4: Write the leaves in ascending order as shown in Figure 11(c).



In Other Words

Using split stems is like adding more classes to a frequency distribution.

Split Stems

Consider the data shown in Table 17. The data range from 11 to 48. If we drew a stem-and-leaf plot using the tens digit as the stem and the ones digit as the leaf, we would obtain the results shown in Figure 12. The data appear rather “bunched.” To resolve this problem, we can use **split stems**. For example, rather than using one stem for the class of data 10–19, we could use two stems, one for the 10–14 interval and the second for the 15–19 interval. We do this in Figure 13.

Table 17

27	17	11	24	36
13	29	22	18	17
23	30	12	46	17
32	48	11	18	23
18	32	26	24	38
24	15	13	31	22
18	21	27	20	16
15	37	19	19	29

Figure 12

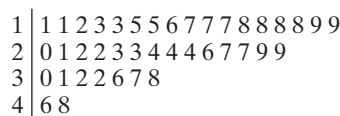
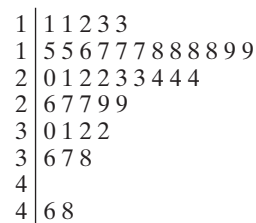


Figure 13



The stem-and-leaf plot shown in Figure 13 reveals the distribution of the data better. As with the construction of class intervals in the creation of frequency histograms, judgment plays a major role. There is no such thing as a correct stem-and-leaf plot. However, a quick comparison of Figures 12 and 13 shows that some are better than others.

One final note: **Stem-and-leaf plots are best used when the data set is small.**

6 Draw Dot Plots

One more graph! A **dot plot** is drawn by placing each observation horizontally in increasing order and placing a dot above the observation each time it is observed.

EXAMPLE 9

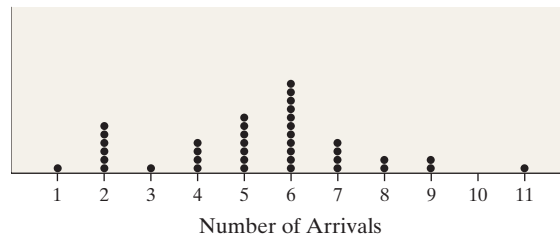
Drawing a Dot Plot

Problem: Draw a dot plot for the number of arrivals at Wendy's data from Example 1 on page 71.

Approach: The smallest observation in the data set is 1 and the largest is 11. We write the numbers 1 through 11 horizontally. For each observation, we place a dot above the value of the observation.

Solution: Figure 14 shows the dot plot.

Figure 14



Now Work Problem 45.

7 Identify the Shape of a Distribution

One way that a variable is described is through the shape of its distribution. Distribution shapes are typically classified as symmetric, skewed left, or skewed right. Figure 15 displays various histograms and the shape of the distribution.

Figures 15(a) and (b) display symmetric distributions. These distributions are symmetric because, if we split the histogram down the middle, the right and left sides of the histograms are mirror images. Figure 15(a) is a **uniform distribution**, because the frequency of each value of the variable is evenly spread out across the values of the variable. Figure 15(b) displays a **bell-shaped distribution**, because the highest frequency occurs in the middle and frequencies tail off to the left and right of the middle so the graph looks like the profile of a bell. Figure 15(c) illustrates a distribution that is **skewed right**. Notice that the tail to the right of the peak is longer than the tail to the left of the peak. Finally, Figure 15(d) illustrates a distribution that is **skewed left**, because the tail to the left of the peak is longer than the tail to the right of the peak.

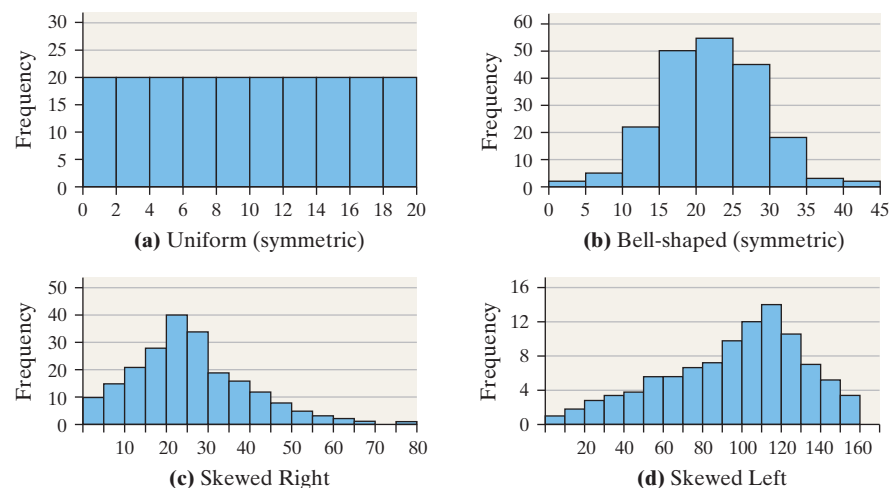
CAUTION

We do not describe qualitative data as skewed left, skewed right, or uniform.

CAUTION

It is important to recognize that data will not always exhibit behavior that perfectly matches any of the shapes given in Figure 15. To identify the shape of a distribution, some flexibility is required. In addition, people may disagree on the shape, since identifying shape is subjective.

Figure 15

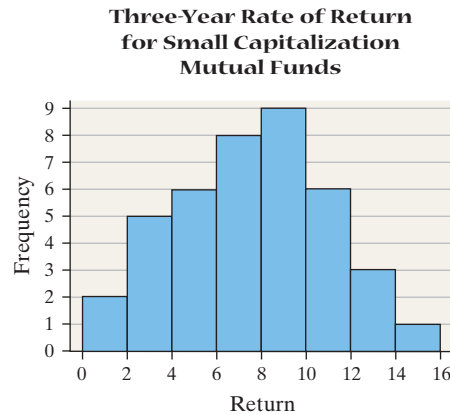


EXAMPLE 10**Identifying the Shape of a Distribution**

Problem: Figure 16 displays the histogram obtained for the three-year rates of return for small-capitalization stocks. Describe the shape of the distribution.

Approach: We compare the shape of the distribution displayed in Figure 16 with those in Figure 15.

Solution: Since the histogram looks most like Figure 15(b), the distribution is bell shaped.

Figure 16

Now Work Problem 29(e).

In-Class Activity: Random-Number Generators, Pulse Rate, and Household Size

1. We saw in Chapter 1 how to use a graphing calculator or statistical spreadsheet to generate random numbers. Using either of these, randomly generate 200 integers between 0 and 20, inclusive. That is, the integer can be any value greater than or equal to 0 or less than or equal to 20. Store these data in your calculator or spreadsheet.
2. Everyone in the class should determine his or her resting pulse rate. Collect these data for the class and store them in your calculator or spreadsheet.
3. Everyone in class should share how many people live in their household. Collect the data for the class, and input them into your calculator or spreadsheet.
 - (a) What shape do you expect the distribution of random integers to have? Why?
 - (b) What shape do you expect the distribution of pulse rates to have? Why?
 - (c) What shape do you expect the distribution of household size to have? Why?
 - (d) Draw a histogram of each data set. For the random integer data, use a class width of 2.
 - (e) What shape did each have? Are you surprised?

8

Draw Time-Series Graphs

If the value of a variable is measured at different points in time, the data are referred to as **time-series data**. The closing price of Cisco Systems stock each month for the past 12 years is an example of time-series data.

Definition

A **time-series plot** is obtained by plotting the time in which a variable is measured on the horizontal axis and the corresponding value of the variable on the vertical axis. Lines are then drawn connecting the points.

Time-series plots are very useful in identifying trends in the data.



Table 18

Date	Closing Price	Date	Closing Price
3/03	12.98	3/04	23.57
4/03	15.00	4/04	20.91
5/03	16.41	5/04	22.37
6/03	16.79	6/04	23.70
7/03	19.49	7/04	20.92
8/03	19.14	8/04	18.76
9/03	19.59	9/04	18.10
10/03	20.93	10/04	19.21
11/03	22.70	11/04	18.75
12/03	24.23	12/04	19.32
1/04	25.71	1/05	18.04
2/04	23.16	2/05	17.42

Source: NASDAQ

EXAMPLE 11

Drawing a Time-Series Plot

Problem: The data in Table 18 represent the closing price of Cisco Systems stock at the end of each month from March 2003 through February 2005. Construct a time-series plot of the data.

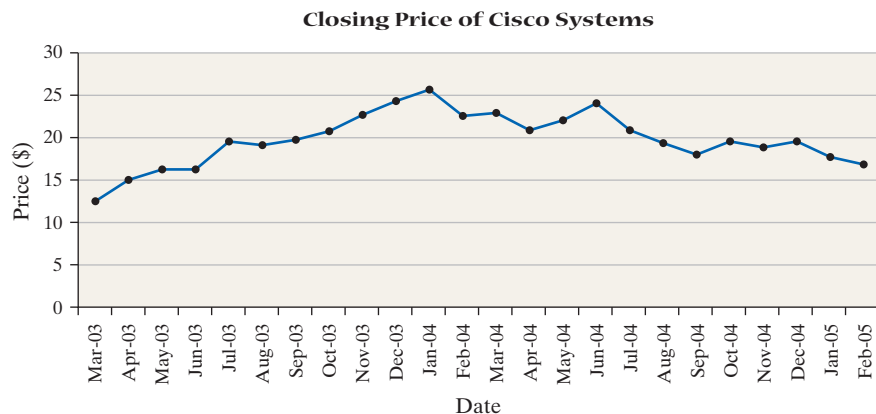
Approach

Step 1: Plot points for each month, with the date on the horizontal axis and the closing price on the vertical axis.

Step 2: Connect the points with straight lines.

Solution: Figure 17 shows the time-series plot. The trend since June, 2004 does not bode well for investors of Cisco Systems stock.

Figure 17



USING TECHNOLOGY:

Statistical spreadsheets, such as Excel or MINITAB, and certain graphing calculators, such as the TI-83 or TI-84 Plus, have the ability to create time-series graphs.

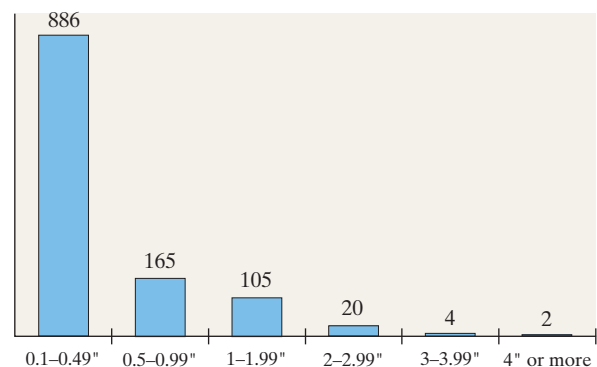
Now Work Problem 47.

2.2 ASSESS YOUR UNDERSTANDING

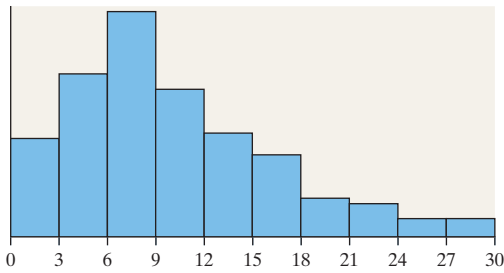
Concepts and Vocabulary

1. Discuss circumstances under which it is preferable to use relative frequency distributions, instead of frequency distributions.
2. Why shouldn't classes overlap when one summarizes continuous data?
3. The histogram to the right represents the total rainfall for each time it rained in Chicago during the month of August since 1871. The histogram was taken from the *Chicago Tribune* on August 14, 2001. What is wrong with the histogram?
4. State the advantages and disadvantages of histograms versus stem-and-leaf plots.
5. Contrast the differences between histograms and bar graphs.
6. *True or False:* There is not one particular frequency distribution that is correct, but there are frequency distributions that are less desirable than others.

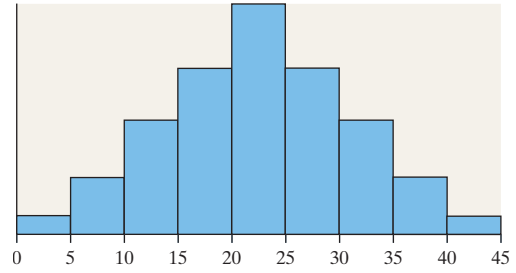
Total August Rain Events Since 1871 in Chicago



7. *True or False:* The shape of the distribution shown is best classified as skewed left.

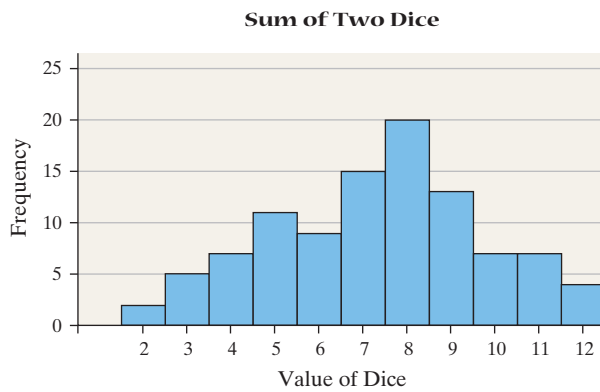


8. *True or False:* The shape of the distribution shown is best classified as uniform.



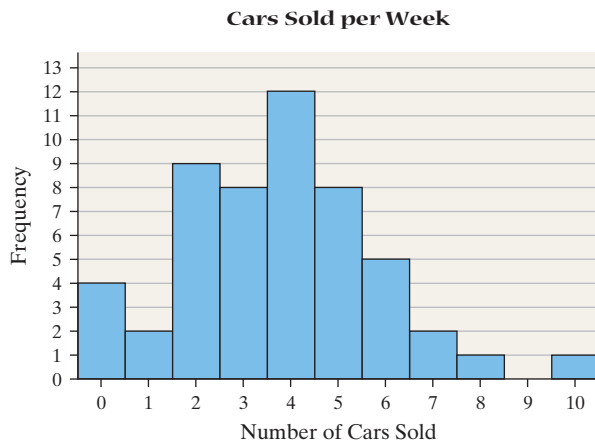
Skill Building

9. **Rolling the Dice** An experiment was conducted in which two fair dice were thrown 100 times. The sum of the pips showing on the dice was then recorded. The following frequency histogram gives the results.



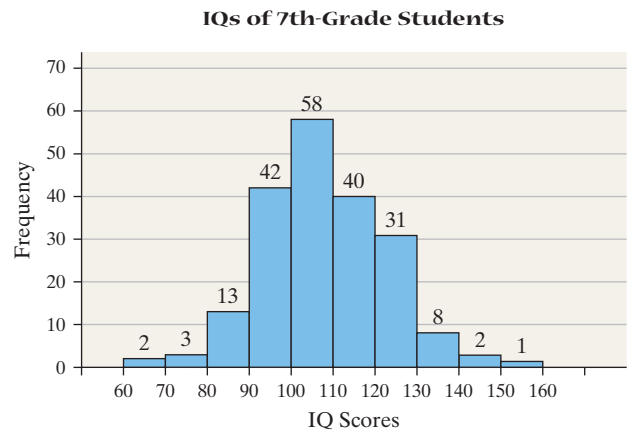
- What was the most frequent outcome of the experiment?
- What was the least frequent?
- How many times did we observe a 7?
- Determine the percentage of time a 7 was observed.
- Describe the shape of the distribution.

10. **Car Sales** A car salesman records the number of cars he sold each week for the past year. The following frequency histogram shows the results.



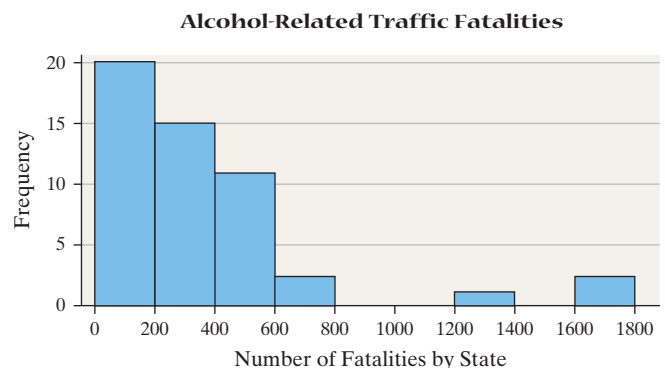
- What is the most frequent number of cars sold in a week?
- For how many weeks were two cars sold?
- Determine the percentage of time two cars were sold.
- Describe the shape of the distribution.

11. **IQ Scores** The following frequency histogram represents the IQ scores of a random sample of seventh-grade students. IQs are measured to the nearest whole number. The frequency of each class is labeled above each rectangle.



- How many students were sampled?
- Determine the class width.
- Identify the classes and their frequencies.
- Which class has the highest frequency?
- Which class has the lowest frequency?

12. **Alcohol-Related Traffic Fatalities** The following frequency histogram represents the number of alcohol-related traffic fatalities by state in 2003 according to data obtained from the National Highway Traffic Safety Administration.



- Determine the class width.
- Identify the classes.
- Which class has the highest frequency?
- Describe the shape of the distribution.

- A reporter writes the following statement: "According to the data, Texas had 1709 alcohol-related deaths, while Vermont had only 29. So the roads in Vermont are much safer." Explain what is wrong with this statement and how a fair comparison can be made between alcohol-related traffic fatalities in Texas versus Vermont.

Applying the Concepts

- 13. Predicting School Enrollment** To predict future enrollment, a local school district wants to know the number of children under the age of 5. Fifty households within the district were sampled, and the head of household was asked to disclose the number of children under the age of 5 living in the household. The results of the survey are presented in the following table.



Number of Children under 5	Number of Households
0	16
1	18
2	12
3	3
4	1

- Construct a relative frequency distribution of the data.
- What percentage of households has two children under the age of 5?
- What percentage of households has one or two children under the age of 5?

- 14. Free Throws** A basketball player habitually makes 70% of her free throws. In an experiment, a researcher asks this basketball player to record the number of free throws she shoots until she misses. The experiment is repeated 50 times. The following table lists the distribution of the number of free throws attempted until a miss is recorded.



Number of Free Throws until a Miss	Frequency
1	16
2	11
3	9
4	7
5	2
6	3
7	0
8	1
9	0
10	1

- Construct a relative frequency distribution of the data.
- What percentage of the time did she miss on her fourth free throw?
- What percentage of the time did she make nine in a row and then miss the tenth free throw?
- What percentage of the time did she make at least five in a row?

In Problems 15 and 16, determine the original set of data. The stem represents the tens digit and the leaf represents the ones digit.

15.

1	0 1 4
2	1 4 4 7 9
3	3 5 5 5 7 7 8
4	0 0 1 2 6 6 8 9 9
5	3 3 5 8
6	1 2

16.

4	0 4 7
5	2 2 3 9 9
6	3 4 5 8 8 9
7	0 1 1 3 6 6
8	2 3 8

In Problems 17 and 18, determine the original set of data. The stem represents the ones digits and the leaf represents the tenths digit.

17.


1	2 4 6
2	1 4 7 7 9
3	3 3 3 5 7 7 8
4	0 1 1 3 6 6 8 8 9
5	3 4 5 8
6	2 4

18.

12	3 7 9 9
13	0 4 5 7 8 9 9
14	2 4 4 7 7 8 9
15	1 2 2 5 6
16	0 3

In Problems 19–22, find (a) the number of classes, (b) the class limits, and (c) the class width.


- 19. Health Insurance** The following data represent the number of people aged 25 to 64 covered by health insurance in 2003.



Age	Number (millions)
25–34	28.9
35–44	35.7
45–54	35.1
55–64	24.7

Source: U.S. Census Bureau, Current Population Survey, 2004 Annual Social and Economic Supplement


- 21. Hospital Admissions** The following data represent community hospital admissions for hospitals with 100 to 499 beds during the year 2002.



Number of Beds	Admissions (in thousands)
100–199	6,826
200–299	6,800
300–399	5,607
400–499	3,593

Source: Health, United States, 2004, from the National Center for Health Statistics


- 20. Earthquakes** The following data represent the number of earthquakes worldwide whose magnitude was less than 8.0 in 2004.



Magnitude	Number
0–0.9	3,179
1.0–1.9	1,340
2.0–2.9	6,297
3.0–3.9	7,826
4.0–4.9	10,975
5.0–5.9	1,430
6.0–6.9	139
7.0–7.9	13

Source: U.S. Geological Survey, Earthquake Hazards Program

- 22. Multiple Births** The following data represent the number of live multiple births (three or more babies) in 2002 for women 15 to 44 years old.



Age	Multiple Births
15–19	93
20–24	511
25–29	1,628
30–34	2,832
35–39	1,843
40–44	377

Source: National Vital Statistics Reports, Vol. 52, No. 10, December 17, 2003

In Problems 23–26, construct (a) a relative frequency distribution, (b) a frequency histogram, and (c) a relative frequency histogram for the given data. Then answer the questions that follow.

- 23.** Using the data in Problem 19, of the people covered by health insurance, what percentage is 25 to 34 years old? Of the people covered by health insurance, what percentage is 44 years or younger?
- 24.** Using the data in Problem 20, what percentage of earthquakes registered 4.0 to 4.9? What percentage of earthquakes registered 4.9 or less?
- 25.** Using the data in Problem 21, what percentage of the admissions was in hospitals with 300 to 399 beds? What percentage of the admissions was in hospitals with 300 or more beds?
- 26.** Using the data in Problem 22, what percentage of multiple births was to women 40 to 44 years old? What percentage of multiple births was to women 24 years or younger?

- 27. Waiting** The following data represent the number of customers waiting for a table at 6:00 P.M. for 40 consecutive Saturdays at Bobak’s Restaurant:



11	5	11	3
4	5	13	9
13	10	9	6
10	8	7	3
7	9	10	4
6	8	6	7
6	4	14	11
8	10	9	5
8	8	7	8
8	6	11	8

- (a) Construct a frequency distribution of the data.
- (b) Construct a relative frequency distribution of the data.
- (c) What percentage of the Saturdays had 10 or more customers waiting for a table at 6:00 P.M.?
- (d) What percentage of the Saturdays had five or fewer customers waiting for a table at 6:00 P.M.?
- (e) Construct a frequency histogram of the data.
- (f) Construct a relative frequency histogram of the data.
- (g) Describe the shape of the distribution.

28. Highway Repair The following data represent the number of potholes on 50 randomly selected 1-mile stretches of highway in the city of Chicago.



2	7	4	7	2	7
2	2	2	3	4	3
1	2	3	2	1	4
2	2	5	2	3	4
4	1	7	10	3	5
4	3	3	2	2	
1	6	5	7	9	
2	2	2	1	5	
3	5	1	3	5	

- (a) Construct a frequency distribution of the data.
- (b) Construct a relative frequency distribution of the data.
- (c) What percentage of the 1-mile stretches of highway had seven or more potholes?
- (d) What percentage of the 1-mile stretches of highway had two or fewer potholes?
- (e) Construct a frequency histogram of the data.
- (f) Construct a relative frequency histogram of the data.
- (g) Describe the shape of the distribution.

29. Average Income The following data represent the per capita (average) disposable income (income after taxes) for the 50 states and the District of Columbia in 2003.



24,028	30,641	24,293	22,123	29,798	30,507
36,726	28,960	42,345	27,610	26,356	27,837
23,584	30,063	25,929	26,409	27,033	23,567
23,889	25,900	32,637	34,570	27,275	30,397
21,677	26,317	23,528	27,865	28,188	31,251
35,411	23,301	31,527	25,307	26,902	26,684
24,169	26,102	28,557	28,365	23,753	27,149
26,314	26,922	22,581	27,750	29,683	30,288
22,252	27,508	29,600			

Source: U.S. Bureau of Economic Analysis, May 2004

With the first class having a lower class limit of 20,000 and a class width of 2500,

- (a) Construct a frequency distribution.
- (b) Construct a relative frequency distribution.
- (c) Construct a frequency histogram of the data.
- (d) Construct a relative frequency histogram of the data.
- (e) Describe the shape of the distribution.
- (f) Repeat parts (a)–(e) using a class width of 4000. Which frequency distribution seems to provide a better summary of the data?
- (g) The highest per capita disposable income exists in the District of Columbia, yet the District of Columbia has one of the highest unemployment rates (7% unemployed). Is this surprising to you? Why?

30. Poverty Every year the federal government adjusts the income level that must be earned to be above the poverty level. The dollar amount depends on the number of people living in the household. In 2002, a family of four needed to earn \$18,392 to be above the poverty line. The following data represent the percentage of each state's population (plus the District of Columbia) that is living in poverty.



14.6	8.3	13.3	18.0	12.8	9.4
7.8	8.1	16.8	12.1	12.1	10.6
11.8	11.2	8.7	8.3	9.4	13.1
17.0	11.3	7.3	9.6	10.3	6.5
17.6	9.6	13.7	9.5	8.3	5.6
7.8	17.8	14.0	13.1	11.9	10.1
14.7	11.2	9.2	10.3	13.5	10.2
14.2	15.3	9.3	9.9	8.7	10.8
16.0	8.6	9.5			

Source: U.S. Census Bureau

With the first class having a lower class limit of 5 and a class width of 1,

- (a) Construct a frequency distribution.
- (b) Construct a relative frequency distribution.
- (c) Construct a frequency histogram of the data.
- (d) Construct a relative frequency histogram of the data.
- (e) Describe the shape of the distribution.
- (f) Repeat parts (a)–(e) using a class width of 2. Which frequency distribution seems to provide a better summary of the data?
- (g) From Problem 29, we learned that the highest per capita disposable income exists in the District of Columbia. The poverty rate of the District of Columbia is 16.8. Is this surprising to you? Why?

31. Serum HDL Dr. Paul Oswiecinski randomly selects 40 of his 20- to 29-year-old patients and obtains the following data regarding their serum HDL cholesterol:

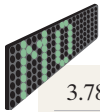


70	56	48	48	53	52	66	48
36	49	28	35	58	62	45	60
38	73	45	51	56	51	46	39
56	32	44	60	51	44	63	50
46	69	53	70	33	54	55	52

With the first class having a lower class limit of 20 and a class width of 10,

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- Construct a frequency histogram of the data.
- Construct a relative frequency histogram of the data.
- Describe the shape of the distribution.
- Repeat parts (a)–(e) using a class width of 5.
- Which frequency distribution seems to provide a better summary of the data?

32. Volume of Altria Group Stock The volume of a stock is the number of shares traded on a given day. The following data, in millions, so that 3.78 represents 3,780,000 shares traded, represent the volume of Altria Group stock traded for a random sample of 35 trading days in 2004.



3.78	8.74	4.35	5.02	8.40
6.06	5.75	5.34	6.92	6.23
5.32	3.25	6.57	7.57	6.07
3.04	5.64	5.00	7.16	4.88
10.32	3.38	7.25	6.52	4.43
3.38	5.53	4.74	9.70	3.56
10.96	4.50	7.97	3.01	5.58

Source: Yahoo.finance.com

With the first class having a lower class limit of 3 and a class width of 2,

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- Construct a frequency histogram of the data.
- Construct a relative frequency histogram of the data.
- Describe the shape of the distribution.
- Repeat parts (a)–(e) using a class width of 1.
- Which frequency distribution seems to provide a better summary of the data?

33. Dividend Yield A dividend is a payment from a publicly traded company to its shareholders. The dividend yield of a stock is determined by dividing the annual dividend of a stock by its price. The following data represent the dividend yields (in percent) of a random sample of 28 publicly traded stocks of companies with a value of at least \$5 billion.



1.7	0	1.15	0.62	1.06	2.45	2.38
2.83	2.16	1.05	1.22	1.68	0.89	0
2.59	0	1.7	0.64	0.67	2.07	0.94
2.04	0	0	1.35	0	0	0.41

Source: Yahoo! Finance

With the first class having a lower class limit of 0 and a class width of 0.40,

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- Construct a frequency histogram of the data.
- Construct a relative frequency histogram of the data.
- Describe the shape of the distribution.
- Repeat parts (a)–(e) using a class width of 0.8.
- Which frequency distribution seems to provide a better summary of the data?

34. Violent Crimes Violent crimes include murder, forcible rape, robbery, and aggravated assault. The following data represent the violent crime rate (crimes per 100,000 population) by state plus the District of Columbia in 2002.



444	563	553	424	593	352
311	599	1,633	770	459	262
255	621	357	286	377	279
662	108	770	484	540	268
343	539	352	314	638	161
375	740	496	470	78	351
503	292	402	285	822	177
717	579	237	107	291	345
234	225	274			

Source: U.S. Federal Bureau of Investigation

With the first class having a lower class limit of 0 and a class width of 150,

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- Construct a frequency histogram of the data.
- Construct a relative frequency histogram of the data.
- Describe the shape of the distribution.
- Repeat parts (a)–(e) using a class width of 300. Which frequency distribution seems to provide a better summary of the data?
- Do you believe that the violent crime rate is a good measure of how safe a state is? Why or why not?

In Problems 35–38, construct stem-and-leaf plots.

35. Age at Inauguration The following data represent the **NW** ages of the presidents of the United States on inauguration day.



57	61	57	57	58
54	68	51	49	64
65	52	56	46	54
47	55	55	54	42
55	51	54	51	60
55	56	61	52	69
57	50	49	51	62
61	48	50	56	43
64	46	54		

36. Divorce Rate The following data represent the divorce rates (per 1000 population) for most states in the United States in the year 2001. (**Note:** The list includes the District of Columbia but excludes California, Colorado, Indiana, and Louisiana because of failure to report.)



5.3	4.1	4.2	6.6	2.9	4.0	2.3
5.4	3.8	3.8	5.6	3.2	3.2	3.2
5.5	3.9	3.0	2.4	3.9	3.3	5.4
4.3	2.6	3.7	6.8	5.0	3.5	5.1
3.0	4.5	2.7	4.0	3.4	4.9	3.2
3.3	3.5	3.4	5.2	4.1	4.4	4.0
4.3	4.5	5.2	3.2	6.1		

Source: U.S. Census, *Statistical Abstract of the United States*, 2003

37. Grams of Fat in a McDonald's Breakfast The following data represent the number of grams of fat in breakfast meals offered at McDonald's.



12	23	28	2	28	33
31	11	23	40	35	1
23	33	23	16	11	8
8	17	16	15		

Source: McDonald's Corporation, *A Full Serving of Nutrition Facts*, April 2003

38. Miles per Gallon The following data represent the number of miles per gallon achieved on the highway for compact cars for the model year 2005.



30	29	30	21	18	29	27	30	29
34	34	30	28	30	20	32	28	32
34	35	26	26	31	25	35	32	25
19	26	19	24	22	24	19	31	26
34	32	34	25	34	34	32	29	25
31	29	30	30	34	32	29	38	39
46	31	31	30	27	29	26	29	24

Source: U.S. Department of Energy

39. Energy from Coal The following data represent the percentage of energy derived from coal for the 50 states plus the District of Columbia.



54.2	0	75.0	64.3	15.6	8.0	94.2
8.5	32.4	90.4	34.6	87.7	55.7	0
40.6	62.3	23.2	83.1	16.6	0	50.8
48.5	13.3	2.7	60.2	60.4	38.3	8.4
1.3	0.9	59.5	63.1	94.6	42.4	98.1
77.6	46.1	27.4	51.1	90.4	62.1	68.1
10.3	93.7	56.6	23.3	60.8	36.8	95.8
57.7	83.2					

Source: U.S. Energy Information Administration

- Round each observation to the nearest percent and draw a stem-and-leaf diagram.
- Describe the shape of the distribution.
- West Virginia is one of the largest coal mining states. Which number in the data set do you think is West Virginia?

40. Housing Prices The data at the right represent the percentage change in the price of housing from 1998 to 2003 for a random sample of 40 cities.



23.3	20.8	32.5	15.8	47.1	18.9	66.0	22.6
23.4	24.1	16.2	9.1	17.1	22.7	17.0	21.6
29.9	15.6	24.8	52.4	28.3	53.5	17.8	20.6
20.6	37.6	49.4	62.4	11.8	19.2	19.8	59.1
48.1	19.1	35.9	14.7	24.9	25.0	26.1	47.7

Source: Global Insight

- Round each observation to the nearest percent and draw a stem-and-leaf diagram.
- Describe the shape of the distribution.

41. Dependability Survey J.D. Power and Associates regularly surveys car owners and asks them about the reliability of their cars. The following data represent the number of problems per 100 vehicles for the 2000 model year for all makes.



Make	Problems per 100 Vehicles	Make	Problems per 100 Vehicles	Make	Problems per 100 Vehicles
Lexus	163	Subaru	266	Audi	318
Infiniti	174	Nissan	267	Mercedes Benz	318
Buick	179	GMC	269	Jeep	321
Porsche	193	Chevrolet	272	Volvo	330
Acura	196	Saturn	273	Mitsubishi	339
Toyota	201	Oldsmobile	283	Hyundai	342
Cadillac	209	Mazda	288	Isuzu	368
Lincoln	212	Pontiac	293	Volkswagen	391
Honda	218	Chrysler	295	Suzuki	403
Mercury	240	Ford	295	Daewoo	421
Jaguar	247	Plymouth	302	Land Rover	441
Saab	255	Dodge	312	Kia	509
BMW	262				

Source: Associated Press

- Round the data to the nearest tens (for example, round 163 as 160).
- Draw a stem-and-leaf diagram, treating the hundreds position as the stem and the tens position as the leaf. For example, for the observation 160, 1 is the stem and 6 is the leaf. Do you think this is a good summary of the data? Why?
- Redraw the stem-and-leaf diagram using split stems. For example, data between 200 and 240 is one stem and data between 250 and 290 is a second stem. Does this stem-and-leaf diagram better summarize the data? Why?

42. Crime Rates The data on the following page represent the violent crime rates per 100,000 population for the 50 states and the District of Columbia in 2002.



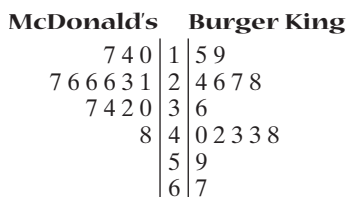
- Round the data to the nearest tens (for example, round 563 to 560).
- Draw a stem-and-leaf diagram, treating the hundreds position as the stem and the tens position as the leaf. For the observation 560, 5 is the stem and 6 is the leaf. Do you think this is a good summary of the data? Why?
- Redraw the stem-and-leaf diagram using split stems. For example, data between 200 and 240 is one stem and data between 250 and 290 is a second stem. Does this stem-and-leaf diagram better summarize the data? Why?



State	Violent Crime Rate	State	Violent Crime Rate	State	Violent Crime Rate
Alabama	444	Kentucky	279	North Dakota	78
Alaska	563	Louisiana	662	Ohio	351
Arizona	553	Maine	108	Oklahoma	503
Arkansas	424	Maryland	770	Oregon	292
California	593	Massachusetts	484	Pennsylvania	402
Colorado	352	Michigan	540	Rhode Island	285
Connecticut	311	Minnesota	268	South Carolina	822
Delaware	599	Mississippi	343	South Dakota	177
District of Columbia	1633	Missouri	539	Tennessee	717
Florida	770	Montana	352	Texas	579
Georgia	459	Nebraska	314	Utah	237
Hawaii	262	Nevada	638	Vermont	107
Idaho	255	New Hampshire	161	Virginia	291
Illinois	621	New Jersey	375	Washington	345
Indiana	357	New Mexico	740	West Virginia	234
Iowa	286	New York	496	Wisconsin	225
Kansas	377	North Carolina	470	Wyoming	274

Source: Federal Bureau of Investigation

In Problems 43 and 44, we compare data sets. A great way to compare two data sets is through back to back stem-and-leaf diagrams. The figure represents the number of grams of fat in 14 sandwiches served at McDonald's and 14 sandwiches served at Burger King. In the figure, 0|1|5 represents 10 for McDonald's and 15 for Burger King. Source: McDonald's Corporation, *A Full Serving of Nutrition Facts*, April 2003; Burger King Corporation, *Great Taste Table*, January 1999.



43. Academy Award Winners The following data represent the ages of the Academy Award winners for Best Actor and Best Actress in a leading role for the 30 years from 1975 to 2004.



Best Actor Ages					
38	60	30	40	42	37
76	39	52	45	35	61
43	51	32	42	54	52
37	38	31	45	60	45
40	36	47	29	43	37



Best Actress Ages					
41	35	31	41	33	31
74	33	49	38	61	21
41	25	80	42	28	33
35	45	49	39	34	24
25	33	35	35	28	30

- (a) Construct a back to back stem-and-leaf display.
- (b) Compare the two populations. What can you conclude from the back to back stem-and-leaf display?

- 44. Home Run Distances** In 1998, Mark McGwire of the St. Louis Cardinals set the record for the most home runs hit in a season by hitting 70 home runs. Three years later in 2001, Barry Bonds of the San Francisco Giants broke McGwire's record by hitting 73 home runs. The following data represent the distances of each player's home runs in his record-setting season.



Mark McGwire						
360	370	370	430	420	340	460
410	440	410	380	360	350	527
380	550	478	420	390	420	425
370	480	390	430	388	423	410
360	410	450	350	450	430	461
430	470	440	400	390	510	430
450	452	420	380	470	398	409
385	369	460	390	510	500	450
470	430	458	380	430	341	385
410	420	380	400	440	377	370



Barry Bonds						
420	417	440	410	390	417	420
410	380	430	370	420	400	360
410	420	391	416	440	410	415
436	430	410	400	390	420	410
420	410	410	450	320	430	380
375	375	347	380	429	320	360
375	370	440	400	405	430	350
396	410	380	430	415	380	375
400	435	420	420	488	361	394
410	411	365	360	440	435	454
442	404	385				

- (a) Construct a back to back stem-and-leaf display.
 (b) Compare the two populations. What can you conclude from the back to back stem-and-leaf display?

NW 45. Waiting Draw a dot plot of the waiting data from Problem 27.

46. Highway Repair Draw a dot plot of the highway repair data from Problem 28.

47. Pixar Corporation The data to the right represent the stock price for Pixar at the end of each month in 2004, adjusted for dividends and splits.

- (a) Construct a time-series plot of the data and comment on the trend.
 (b) During what month did the stock price increase the most? What might explain this?

Date	Closing Price	Date	Closing Price
1/04	66.39	7/04	68.24
2/04	65.76	8/04	77.72
3/04	64.46	9/04	78.90
4/04	68.27	10/04	80.42
5/04	67.85	11/04	90.67
6/04	69.51	12/04	85.61

48. Google Corporation The following data represent the stock price for Google at the end of each month from August 2004 to February 2005, adjusted for dividends and splits. Construct a time-series plot and comment on any trends. (**Note:** Google stock was traded publicly for the first time in August 2004.)



Date	Closing Price
8/04	102.37
9/04	129.60
10/04	190.64
11/04	181.98
12/04	192.79
1/05	187.99
2/05	185.18

49. College Enrollment The following data represent the percentage of recent high school graduates (graduated within 12 months before the given year-end) who enrolled in college in the fall. Construct a time-series plot of the data.



Year	Percent Enrolled	Year	Percent Enrolled
1988	58.9	1996	65.0
1989	59.6	1997	67.0
1990	60.1	1998	65.6
1991	62.5	1999	62.9
1992	61.9	2000	63.3
1993	62.6	2001	61.7
1994	61.9	2002	65.2
1995	61.9	2003	63.9

Source: U.S. Center for Education Statistics

- 50. IRS Audits** The following data represent the percentage of tax returns audited by the Internal Revenue Service. Construct a time-series plot of the data.

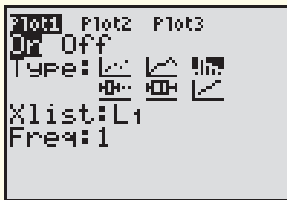


Year	Percent Audited	Year	Percent Audited
1988	1.57	1996	1.67
1989	1.29	1997	1.28
1990	1.04	1998	0.99
1991	1.17	1999	0.90
1992	1.06	2000	0.49
1993	0.92	2001	0.58
1994	1.07	2002	0.57
1995	1.67	2003	0.65

Source: U.S. General Accounting Office

Technology Step by Step

TI-83/84 Plus



Drawing Histograms and Stem-and-Leaf Plots

Histograms

Step 1: Enter the raw data in L1 by pressing STAT and selecting 1: Edit.

Step 2: Press 2nd Y = to access the StatPlot menu. Select 1: Plot1.

Step 3: Place the cursor on “ON” and press ENTER.

Step 4: Place the cursor on the histogram icon (see the figure) and press ENTER. Press 2nd QUIT to exit Plot 1 menu.

Step 5: Press WINDOW. Set Xmin to the lower class limit of the first class. Set Xmax to the lower class limit of the class following the class containing the largest value. For example, if the first class is 0–9, set Xmin to 0. If the class width is 10 and the last class is 90–99, set Xmax to 100. Set Xscl to the class width. Set Ymin to 0. Set Ymax to a value larger than the frequency of the class with the highest frequency.

Step 6: Press GRAPH.

Helpful Hints: To determine each class frequency, press TRACE and use the arrow keys to scroll through each class. If you decrease the value of Ymin to a value such as –5, you can see the values displayed on the screen more easily. The TI graphing calculators do not draw stem-and-leaf plots or dot plots.

MINITAB

Histograms

Step 1: Enter the raw data in C1.

Step 2: Select the **Graph** menu and highlight **Histogram** . . .

Step 3: Highlight the “simple” icon and press OK.

Step 4: Put the cursor in the “Graph variables” box. Highlight C1, and press Select. Click SCALE and select the Y-Scale Type tab. For a frequency histogram, click the frequency radio button. For a relative frequency histogram, click the percent radio button. Click OK twice.

Note: To adjust the class width and to change the labels on the horizontal axis to the lower class limit, double-click inside one of the bars in the histogram. Select the “binning” tab in the window that opens. Click the cutpoint button and the midpoint/cutpoint position radio button. In the midpoint/cutpoint box, enter the lower class limits of each class. Click OK.

Stem-and-Leaf Plots

Step 1: With the raw data entered in C1, select the **Graph** menu and highlight **Stem-and-Leaf**.

Step 2: Select the data in C1 and press OK.

Dot Plots

Step 1: Enter the raw data in C1.

Step 2: Select the **Graph** menu and highlight **Dotplot**.

Step 3: Highlight the “simple” icon and press OK.

Step 4: Put the cursor in the “Graph variables” box. Highlight C1 and press Select. Click OK.

Excel Histograms

Step 1: Enter the raw data in column A.

Step 2: Select **Tools** and **Data Analysis . . .**

Step 3: Select Histogram from the list.

Step 4: With the cursor in the Input Range cell, use the mouse to highlight the raw data. Select the Chart Output box and press OK.

Step 5: Double-click on one of the bars in the histogram. Select the Options tab from the menu that appears. Reduce the gap width to zero.

Excel does not draw stem-and-leaf or dot plots.

2.3 Graphical Misrepresentations of Data

Objective 1 Describe what can make a graph misleading or deceptive

1 Describe What Can Make a Graph Misleading or Deceptive

Often, statistics gets a bad rap for having the ability to manipulate data to support any position desired. One method of distorting the truth is through graphics. Sometimes graphics *mislead*; other times they *deceive*. We will call graphs misleading if they unintentionally create an incorrect impression. We consider graphs deceptive if they purposely create an incorrect impression. We have already discussed the power that graphical representations of data can have, so it is important to be able to recognize misleading and deceptive graphs. The most common graphical misrepresentation of data is accomplished through the manipulation of the scale of the graph.

EXAMPLE 1

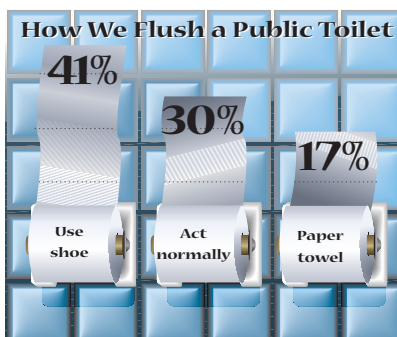
Misrepresentation of Data

Problem: The bar graph illustrated in Figure 18 is a *USA Today* type graph. A survey was conducted by Impulse Research for Quilted Northern Confidential in which individuals were asked how they would flush a toilet when the facilities are not sanitary. What’s wrong with the graphic?

Approach: We need to compare the vertical scales of each bar to see if they accurately depict the percentages given.

Solution: First, it is unclear whether the bars include the roll of toilet paper or not. In either case, the roll corresponding to “use shoe” should be 2.4 ($= 41/17$) times longer than the roll corresponding to “paper towel.” If we include the roll of toilet paper, then the bar corresponding to “use shoe” is less than double the length of “paper towel.” If we do not include the roll of toilet paper, then the bar corresponding to “use shoe” is almost exactly double the length of the bar corresponding to “paper towel.” The vertical scaling is incorrect.

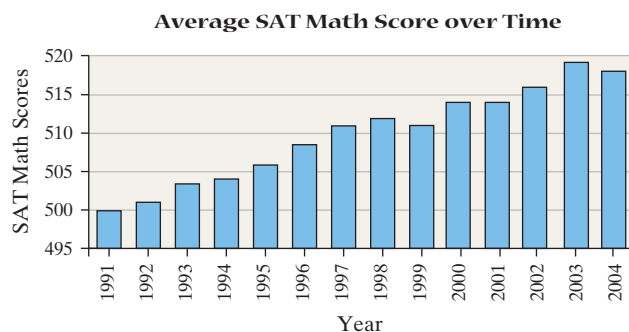
Figure 18



EXAMPLE 2**Misrepresentation of Data by Manipulating the Vertical Scale**

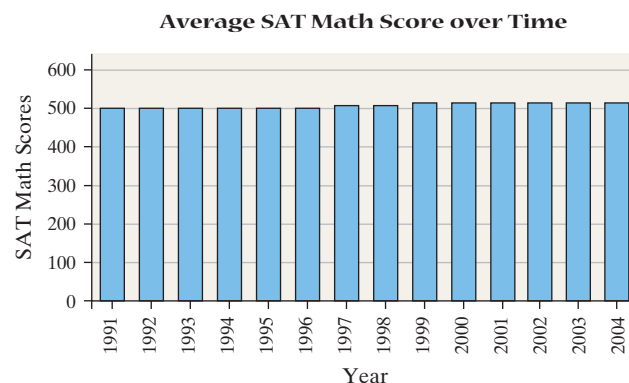
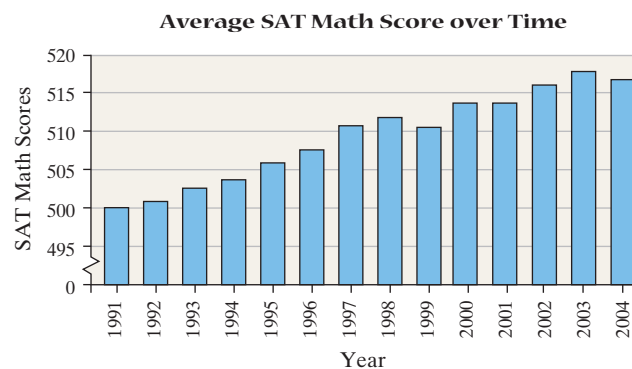
Problem: The bar graph shown in Figure 19 depicts the average SAT Math scores of college-bound seniors for the years 1991–2004, based on data from the College Board. Determine why this graph might be considered misrepresentative.

Approach: We need to look at the graph for any characteristics that may mislead a reader, such as manipulation of the vertical scale.

Figure 19

Solution: The graph in the figure may lead a reader to believe that SAT math scores have increased substantially since 1991. While SAT math scores have been increasing, they have not doubled or tripled, as may be inferred from the graph (since the bar for 1997 is three times as high as the bar for 1991). We notice in the figure that the vertical axis begins its labeling at 495 instead of 0. This type of scaling is common when the smallest observed data value is a rather large number. It is not necessarily done purposely to confuse or mislead the reader. Often, the main purpose in graphs is to discover a trend, rather than the actual differences in the data. The trend is clearer in Figure 19 than in Figure 20, where the vertical axis begins at 0.

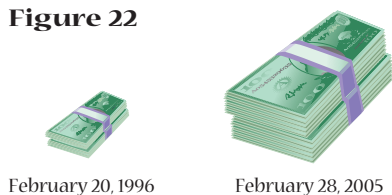
Often, instead of beginning the axis of a graph at 0 as in Figure 20, the graph is begun at a value slightly less than the smallest value in the data set. However, special care must be taken to make the reader aware of the vertical-axis scaling. Figure 21 shows the proper construction of the graph of the SAT scores, with the graph beginning at 495. The symbol \sphericalangle is used to signify that the graph has a gap in it.

Figure 20**Figure 21****Now Work Problem 5.**

In addition to vertical-axis scaling, figures can be misleading through pictures. Consider the next example.

EXAMPLE 3**Misleading Graphs**

Problem: The Dow Jones Industrial Average (DJIA) is a collection of 30 stocks from the stock market that are thought to be representative of the U.S. economy. It includes companies such as Intel, Wal-Mart, and General Motors. Had you invested \$10,000 in the DJIA on February 20, 1996, it would have been worth \$20,000 on February 28, 2005. To illustrate this investment, a brokerage firm might create the graphic shown in Figure 22. Describe how this graph is misleading.

Figure 22

Approach: Again, we look for characteristics of the graph that seem to manipulate the facts, such as an incorrect depiction of the size of the graphics.

Solution: The graphic on the right of the figure has been doubled in length, width, and height, causing an eightfold increase in the size, thereby misleading the reader into thinking that the size of the investment increased by eight times instead of by two times.

Now Work Problem 15.

There are many ways to create graphs that mislead. Two popular texts written about ways that graphs mislead or deceive are *How to Lie with Statistics* (W. W. Norton & Company, Inc., 1982), by Darrell Huff, and *The Visual Display of Quantitative Information* (Graphics Press, 2001), by Edward Tufte.

We conclude this section with some guidelines for constructing good graphics.

Characteristics of Good Graphics

- Label the graphic clearly and provide explanations if needed.
- Avoid distortion. Don't lie about the data.
- Avoid three dimensions. Three-dimensional charts may look nice, but they distract the reader and often result in misinterpretation of the graphic.
- Do not use more than one design in the same graphic. Sometimes graphs use a different design in a portion of the graphic to draw attention to this area. Don't use this technique. Let the numbers speak for themselves.

MAKING AN INFORMED DECISION



Tables or Graphs?

You work for the school newspaper. Your editor approaches you with a special reporting assignment. Your task is to write an article that describes the “typical” student at your school, complete with supporting information. To write this article, you have to survey at least 40 students and ask them to respond to a questionnaire. The editor would like to have at least two qualitative and two quantitative variables that describe the typical student. The results of the survey will be presented in your article, but you are unsure whether you should present tabular or graphical summaries, so you decide to perform the following experiment.

1. Develop a questionnaire that results in obtaining the values of two qualitative and two quantitative variables. Administer the questionnaire to at least 40 students on your campus.
2. Summarize the data in both tabular and graphical form.
3. Select 20 individuals. (They don't have to be students at your school.) Give the tabular summaries

to 10 individuals and the graphical summaries to the other 10. Ask each individual to study the table or graph for 5 seconds. After 1 minute, give a questionnaire that asks various questions regarding the information contained in the table or graph. For example, if you summarized age data, ask the individual which age group had the highest frequency. Record the number of correct answers for each individual. Which summary results in a higher percentage of correct answers, the tables or the graphs? Write a report that discusses your findings.

4. Now use the data collected from the questionnaire to create a couple of misleading graphs. Again, select 20 individuals. Give 10 individuals the misleading graphs and 10 individuals the correct graphs. Ask each individual to study each graph for 5 seconds. After 1 minute has elapsed, give a questionnaire that asks various questions regarding the information contained in the graphs. Record the number of correct answers for each individual. Did the misleading graphs mislead? Write a report that discusses your findings.

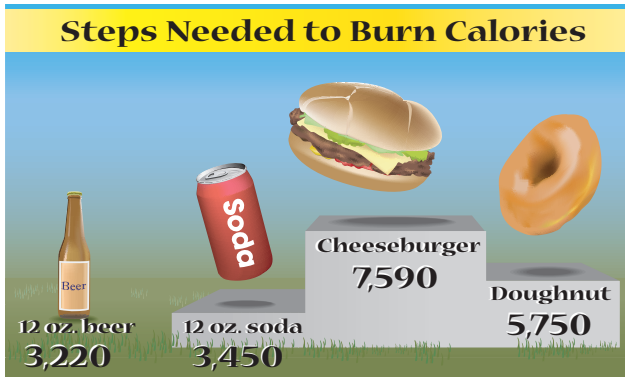
2.3 ASSESS YOUR UNDERSTANDING

Applying the Concepts

1. Inauguration Cost The following is a *USA Today* type graph. Explain how it is misleading.

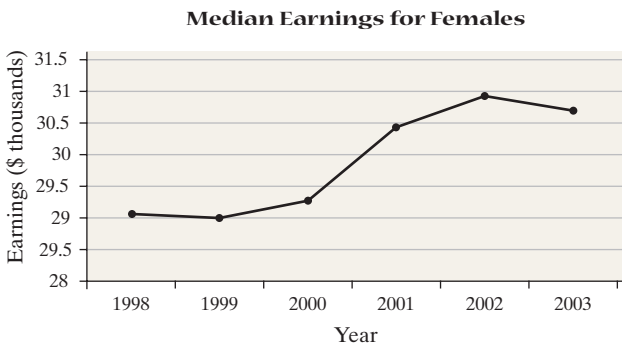


2. Burning Calories The following is a *USA Today* type graph.



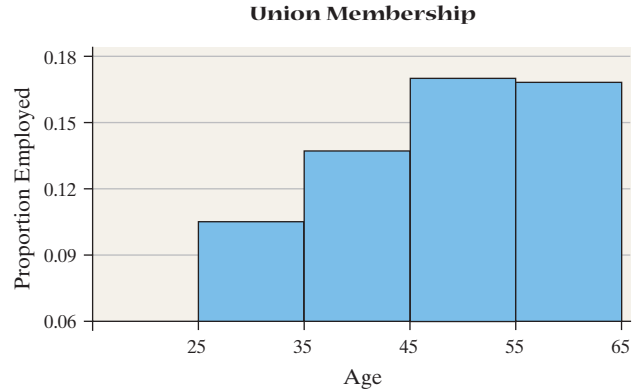
- (a) Explain how it is misleading.
- (b) What could be done to improve the graphic?

3. Median Earnings The following graph shows the median earnings for females from 1998 to 2003. (Source: U.S. Census Bureau, Income, Poverty, and Health Insurance Coverage in the United States, 2003)



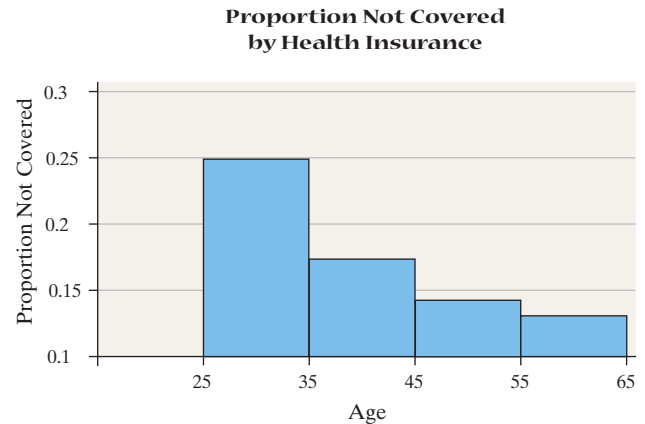
- (a) How is the graph misleading? What does the graph seem to convey?
- (b) Redraw the graph so that it is not misleading. What does the new graph seem to convey?

4. Union Membership The following relative frequency histogram represents the proportion of employed people aged 25 to 64 years old who were members of a union. (Source: U.S. Bureau of Labor Statistics)



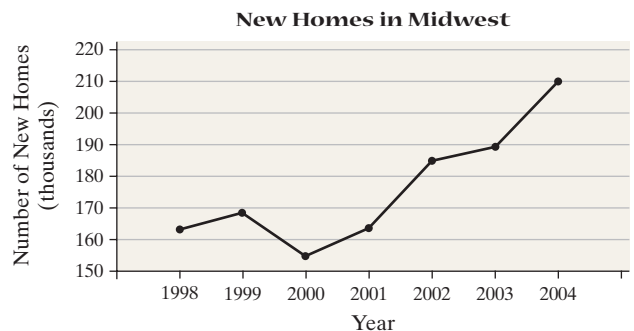
- (a) Describe how this graph is misleading. What might a reader conclude from the graph?
- (b) Redraw the histogram so that it is not misleading.

5. Health Insurance The following relative frequency histogram represents the proportion of people aged 25 to 64 years old not covered by any health insurance in 2002. (Source: U.S. Census Bureau)



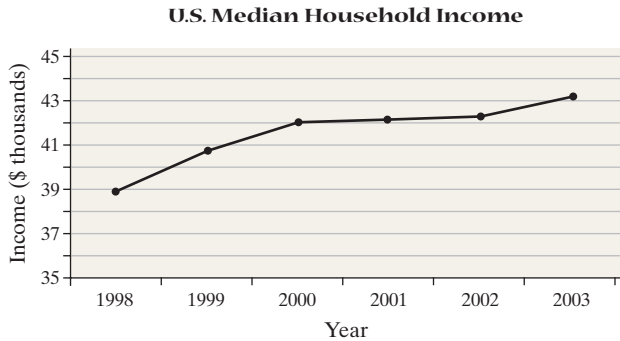
- (a) Describe how this graph is misleading. What might a reader conclude from the graph?
- (b) Redraw the histogram so that it is not misleading.

6. New Homes The following time-series plot shows the number of new homes built in the Midwest from 1998 to 2004. (Source: U.S. Census Bureau)



- (a) Describe how this graph is misleading.
- (b) What is the graph trying to convey?

7. **Median Income** The following time-series plot shows the median household income for the years 1998 to 2003. (Source: U.S. Census Bureau)



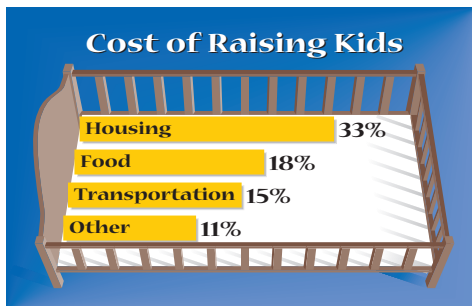
- (a) Describe how this graph is misleading.
- (b) What is the graph trying to convey?

8. **National Debt** The following graphic is a *USA Today* type graph.




How many times larger should the graphic for 1999 be than the 1900 graphic?

9. **Cost of Kids** The following is a *USA Today* type graph based on data from the Department of Agriculture. It represents the percentage of income a middle-income family will spend on their children.



- (a) How is the graphic misleading?
- (b) What could be done to improve the graphic?

10. **SAT Combined Scores** The following table gives the average SAT combined scores for the years 1999–2003.




Year	Avg. SAT Combined Score
1999	1016
2000	1019
2001	1020
2002	1020
2003	1026

Note: Beginning with the tests administered in March 2005, the SAT test consists of three parts for a total possible score of 1800.

- (a) Construct a misleading time-series plot that indicates the average SAT combined score has risen sharply over the given time period.
- (b) Construct a time-series plot that is not misleading.

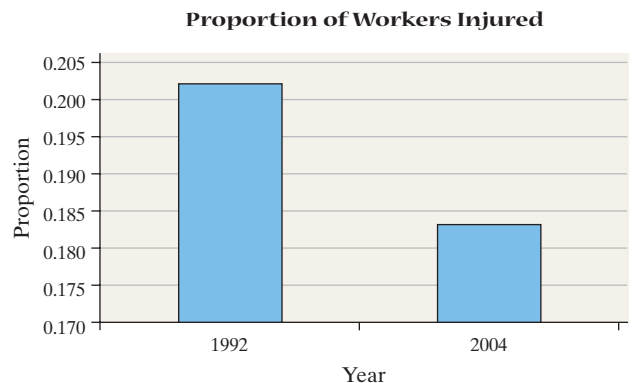
11. **Engineering Degrees** The following table gives the number of bachelor's degrees in engineering that were awarded from 1999 to 2003.



Year	Degrees Awarded
1999	62,372
2000	63,731
2001	65,113
2002	67,301
2003	70,949

- (a) Construct a misleading graph indicating that the number of degrees awarded has more than doubled since 1999.
- (b) Construct a graph that is not misleading.

12. **Worker Injury** The safety manager at Klutz Enterprises provides the following graph to the plant manager and claims that the rate of worker injuries has been reduced by 67% over a 12-year period. Does the graph support his claim? Explain why or why not.



- 13. Health Care Expenditures** The following data represent health care expenditures as a percentage of the U.S. gross domestic product (GDP) from 1997 to 2002. Gross domestic product is the total value of all goods and services created during the course of the year.



Year	Health Care as a Percent of GDP
1997	13.1
1998	13.1
1999	13.2
2000	13.3
2001	14.1
2002	14.9

Source: U.S. Health Care Financing Administration

- (a) Construct a time-series plot that a politician would create to support the position that health care expenditures, as a percentage of GDP, are increasing and must be slowed.
- (b) Construct a time-series plot that the health care industry would create to refute the opinion of the politician.
- (c) Construct a time-series plot that is not misleading.
- 14. Motor Vehicle Accidents** The following data represent the number of motor vehicle accidents and the traffic death rates (number of deaths per 100,000 licensed drivers) from 1998 to 2002.



Year	Motor Vehicle Deaths (in thousands)	Traffic Death Rate (per 100,000 licensed drivers)
1998	43.5	22.4
1999	43.0	22.3
2000	43.0	22.0
2001	43.7	22.1
2002	44.0	22.0

Source: National Safety Council

- (a) Construct a time-series graph to support the belief that the roads are becoming less safe.
- (b) Construct a time-series graph to support the belief that the roads are becoming safer.

- 15. Distance Learning** Between 1995 and 2001, the enrollment in college-level distance education courses nearly doubled.

- (a) Construct a graphic that is not misleading to depict this situation.
- (b) Construct a misleading graphic to depict this situation.

- 16. Overweight** Between 1980 and 2002, the number of adolescents in the United States who were overweight nearly tripled.

- (a) Construct a graphic that is not misleading to depict this situation.
- (b) Construct a misleading graphic to depict this situation.

CHAPTER 2 Review

Summary

Raw data are first organized into tables. Data are organized by creating classes into which they fall. Qualitative data and discrete data have values that provide clear-cut categories of data. However, with continuous data the categories, called classes, must be created. Typically, the first table created is a frequency distribution, which lists the frequency with which each class of data occurs. Another type of distribution is the relative frequency distribution.

Once data are organized into a table, graphs are created. For data that are qualitative, we can create bar charts and pie charts. For data that are quantitative, we can create histograms or stem-and-leaf plots.

In creating graphs, care must be taken not to draw a graph that misleads or deceives the reader. If a graph's vertical axis does not begin at zero, the symbol \neq should be used to indicate the gap that exists in the graph.

Vocabulary

- | | | |
|---|-------------------------------------|----------------------------------|
| Raw data (p. 54) | Histogram (p. 72) | Split stems (p. 79) |
| Frequency distribution (p. 55) | Class (p. 73) | Dot plot (p. 80) |
| Relative frequency (p. 56) | Lower and upper class limit (p. 73) | Uniform distribution (p. 80) |
| Relative frequency distribution (p. 56) | Class width (p. 73) | Bell-shaped distribution (p. 80) |
| Bar graph (p. 57) | Open ended (p. 73) | Skewed right (p. 80) |
| Pareto chart (p. 58) | Stem-and-leaf plot (p. 76) | Skewed left (p. 80) |
| Side-by-side bar graph (p. 58) | Stem (p. 76) | Time-series data (p. 81) |
| Pie chart (p. 60) | Leaf (p. 76) | Time series plot (p. 81) |

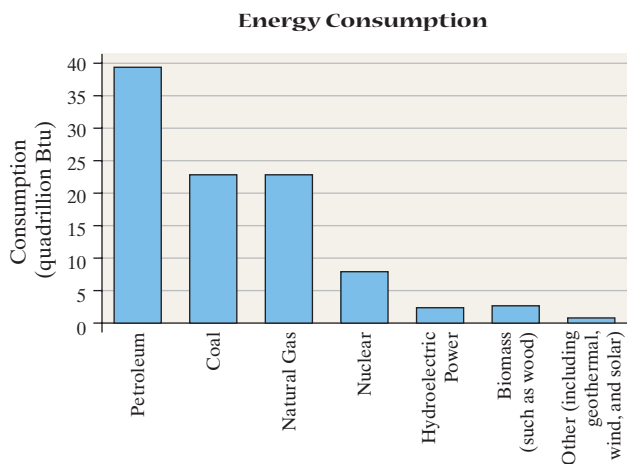
Objectives

Section	You should be able to . . .	Example	Review Exercises
2.1	1 Organize qualitative data in tables (p. 55)	1, 2	3(a), 4(a), 7(a) and (b), 8(a) and (b)
	2 Construct bar graphs (p. 57)	3 through 5	3(c) and (d), 4(c) and (d), 7(c), 8(c)
	3 Construct pie charts (p. 60)	6	3(e), 4(e), 7(d), 8(d)
2.2	1 Organize discrete data in tables (p. 71)	1	9(a) and (b), 10(a) and (b)
	2 Construct histograms of discrete data (p. 72)	2	9(c) and (d), 10(c) and (d),
	3 Organize continuous data in tables (p. 73)	3	11(a) and (b), 12(a) and (b), 13 (a) and (b), 14(a) and (b)
	4 Construct histograms of continuous data (p. 75)	4, 5	5(b) and (c), 6(b) and (c), 11(c) and (d); 12(c) and (d), 13(c) and (d), 14(c) and (d)
	5 Draw stem-and-leaf plots (p. 76)	6 through 8	15, 16
	6 Draw dot plots (p. 80)	9	9(g), 10(g)
	7 Identify the shape of a distribution (p. 80)	10	11(c), 12(c), 13(c), 14(c), 15, 16(b)
	8 Draw time-series graphs (p. 81)	11	17, 18
2.3	1 Describe what can make a graph misleading or deceptive (p. 93)	1 through 3	19, 20, 21

Review Exercises

1. Energy Consumption The following bar chart represents the energy consumption of the United States (in quadrillion Btu) in 2003.

(Source: Energy Information Administration)

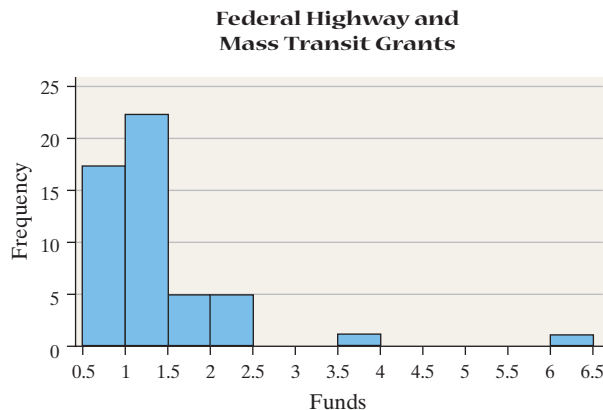


- Approximately how much energy did the United States consume from natural gas?
- Approximately how much energy did the United States consume from biomass?
- Approximate the total energy consumption of the United States in 2003.

- Which category has the lowest frequency?
- Is it appropriate to describe the shape of the distribution as skewed right? Why or why not?

2. Highway Funds The following frequency histogram represents the dollar amount each state and the District of Columbia gets back in federal highway and mass transit grants for each dollar of gasoline tax that their motorists pay into the federal highway trust fund, based on data obtained from the Federal Highway Administration.

- Determine the class width.
- Identify the classes.
- Which class has the highest frequency?



3. Weapons Used in Homicide The following frequency distribution represents the cause of death in homicides for the year 2002.



Type of Weapon	Frequency
Firearms	9369
Knives or cutting instruments	1767
Blunt objects (clubs, hammers, etc.)	666
Personal weapons (hands, fists, etc.)	933
Strangulation	143
Fire	104
Other weapon or not stated	1176

Source: Crime in the United States, 2002, FBI, Uniform Crime Reports

- Construct a relative frequency distribution.
- What percentage of homicides was committed using a blunt object?
- Construct a frequency bar graph.
- Construct a relative frequency bar graph.
- Construct a pie chart.

4. U.S. Greenhouse Emissions The following frequency distribution represents the total greenhouse emissions in millions of metric tons in 2003 in the United States.



Gas	Emissions
Carbon dioxide	1600.8
Methane	164.1
Nitrous oxide	87.3
Hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride	39.1

Source: Energy Information Administration

- Construct a relative frequency distribution.
- What percent of emissions was due to carbon dioxide?
- Construct a frequency bar graph.
- Construct a relative frequency bar graph.
- Construct a pie chart.

5. Live Births The following frequency distribution represents the number of live births (in thousands) in the United States in 2003 by age of mother.



Age of Mother (years)	Births (thousands)
10–14	7
15–19	415
20–24	1032
25–29	1087
30–34	976
35–39	468
40–44	101

Source: National Center for Health Statistics

- Construct a relative frequency distribution.
- Construct a frequency histogram. Describe the shape of the distribution.
- Construct a relative frequency histogram.
- What percentage of live births was to mothers aged 20 to 24?
- What percentage of live births was to mothers of age 30 or older?

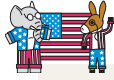
6. Suicides The following frequency distribution represents the number of suicides in the United States in 2001, by age.

Age	Suicides	Age	Suicides
5–9	7	55–59	1985
10–14	272	60–64	1332
15–19	1611	65–69	1212
20–24	2360	70–74	1220
25–29	2389	75–79	1219
30–34	2681	80–84	973
35–39	3176	85–89	538
40–44	3459	90–94	200
45–49	3260	95–99	28
50–54	2682		

Source: National Center for Health Statistics

- Construct a relative frequency distribution.
- Construct a frequency histogram. Describe the shape of the distribution.
- Construct a relative frequency histogram.
- What percentage of suicides from the distribution was aged 40 to 44?
- What percentage of suicides from the distribution was 24 years or younger?

- 7. Political Affiliation** A sample of 100 randomly selected registered voters in the city of Naperville was asked their political affiliation: Democrat (D), Republican (R), or Independent (I). The results of the survey are shown in the table below.



D	R	D	R	D	R	D	D	R	D
R	D	D	D	R	R	D	D	D	D
R	R	I	I	D	R	D	R	R	R
I	D	D	R	I	I	R	D	R	R
D	I	R	D	D	D	D	I	I	R
R	I	R	R	I	D	D	D	D	R
D	I	I	D	D	R	R	R	R	D
D	R	R	R	D	D	I	I	D	D
D	D	I	D	R	I	D	D	D	D
R	R	R	R	R	D	R	D	R	D

- Construct a frequency distribution of the data.
- Construct a relative frequency distribution of the data.
- Construct a relative frequency bar graph of the data.
- Construct a pie chart of the data.
- What appears to be the most common political affiliation in Naperville?

- 8. Educational Attainment** The Metra Train Company was interested in knowing the educational background of its customers. The company contracted a marketing firm to conduct a survey with a random sample of 50 commuters at the train station. In the survey, commuters were asked to disclose their educational attainment. The following results were obtained:



No high school diploma	Some college	Advanced degree	High school graduate	Advanced degree
High school graduate	High school graduate	High school graduate	High school graduate	No high school diploma
Some college	High school graduate	Bachelor's degree	Associate's degree	High school graduate
No high school diploma	Bachelor's degree	Some college	High school graduate	No high school diploma
Associate's degree	High school graduate	High school graduate	No high school diploma	Some college
Bachelor's degree	Bachelor's degree	Some college	High school graduate	Some college
Bachelor's degree	Advanced degree	No high school diploma	Advanced degree	No high school diploma
High school graduate	Bachelor's degree	No high school diploma	High school graduate	No high school diploma
Associate's degree	Bachelor's degree	High school graduate	Bachelor's degree	Some college
Some college	Associate's degree	High school graduate	Some college	High school graduate

- 9. Family Size** A random sample of 60 couples married for 7 years were asked to disclose the number of children they have. The results of the survey are shown to the right.

- Construct a frequency distribution of the data.
- Construct a relative frequency distribution of the data.
- Construct a frequency histogram of the data. Describe the shape of the distribution.
- Construct a relative frequency histogram of the data.
- What percentage of couples married 7 years has two children?
- What percentage of couples married 7 years has at least two children?
- Draw a dot plot of the data.



0	0	3	1	2	3
3	4	3	3	0	3
1	2	1	3	0	3
4	2	3	2	2	4
2	1	3	4	1	3
0	3	3	3	2	1
2	0	3	1	2	3
4	3	3	5	2	0
4	2	2	2	3	3
2	4	2	2	2	2

10. Waiting in Line The data to the right represent the number of cars that arrived at a McDonald's drive-through between 11:50 A.M. and 12:00 noon each Wednesday for the past 50 weeks.



1	7	3	8	2	3	8	2	6	3
6	5	6	4	3	4	3	8	1	2
5	3	6	3	3	4	3	2	1	2
4	4	9	3	5	2	3	5	5	5
2	5	6	1	7	1	5	3	8	4

- (a) Construct a frequency distribution of the data.
- (b) Construct a relative frequency distribution of the data.
- (c) Construct a frequency histogram of the data. Describe the shape of the distribution.
- (d) Construct a relative frequency histogram of the data.
- (e) What percentage of the time did exactly three cars arrive between 11:50 A.M. and 12:00 noon?
- (f) What percentage of the time did three or more cars arrive between 11:50 A.M. and 12:00 noon?
- (g) Draw a dot plot of the data.

11. Crime Rate by State The following data represent the crime rate (per 100,000 population) for each state in 2002.

In (a)–(d), start the first class at a lower class limit of 2000 and maintain a class width of 400.

- (a) Construct a frequency distribution.
- (b) Construct a relative frequency distribution.

- (c) Construct a frequency histogram. Describe the shape of the distribution.
- (d) Construct a relative frequency histogram.
- (e) Repeat (a)–(d), using a class width of 1000. In your opinion, which class width provides the better summary of the data? Why?



State	Crime Rate	State	Crime Rate	State	Crime Rate
Alabama	4465.2	Kentucky	2902.6	North Dakota	2406.2
Alaska	4309.7	Louisiana	5098.1	Ohio	4107.3
Arizona	6386.3	Maine	2656.0	Oklahoma	4743.2
Arkansas	4157.5	Maryland	4747.4	Oregon	4868.4
California	3943.7	Massachusetts	3094.2	Pennsylvania	2841.0
Colorado	4347.8	Michigan	3874.1	Rhode Island	3589.1
Connecticut	2997.2	Minnesota	3535.1	South Carolina	5297.3
Delaware	3939.0	Mississippi	4159.2	South Dakota	2278.7
District of Columbia	8022.3	Missouri	4602.4	Tennessee	5018.0
Florida	5420.6	Montana	3512.9	Texas	5189.6
Georgia	4507.2	Nebraska	4256.7	Utah	4452.4
Hawaii	6043.7	Nevada	4497.5	Vermont	2530.0
Idaho	3172.5	New Hampshire	2220.0	Virginia	3140.3
Illinois	4016.4	New Jersey	3024.2	Washington	5106.8
Indiana	3750.0	New Mexico	5077.8	West Virginia	2515.2
Iowa	3448.2	New York	2803.7	Wisconsin	3252.7
Kansas	4087.0	North Carolina	4721.4	Wyoming	3580.9

Source: *Crime in the United States, 2002*. FBI, Uniform Crime Reports.

12. Towing Capacity The data on page 103 represent the towing capacity (in pounds) for selected sport utility vehicles (SUVs).

Start the first class at a lower class limit of 3000 and maintain a class width of 1000:

- (a) Construct a frequency distribution.
- (b) Construct a relative frequency distribution.
- (c) Construct a frequency histogram. Describe the shape of the distribution.
- (d) Construct a relative frequency histogram.



SUV	Towing Capacity	SUV	Towing Capacity	SUV	Towing Capacity
Acura MDX	4,500	GMC Yukon	8,700	Land Rover Range Rover	7,700
BMW X5	6,000	GMC Yukon XL	12,000	Lincoln Navigator	8,800
Buick Rendezvous	3,500	Honda Passport	4,500	Mitsubishi Montero	5,000
Chevrolet Blazer	5,600	Hummer	8,300	Nissan Pathfinder	5,000
Chevrolet Suburban	12,000	Infiniti QX4	5,000	Pontiac Aztek	3,500
Chevrolet Tahoe	8,700	Isuzu Axiom	4,500	Suzuki XL-7	3,000
Dodge Durango	7,650	Isuzu Rodeo	4,500	Toyota 4Runner	5,000
Ford Escape	3,500	Jeep Cherokee	5,000	Toyota Highlander	3,500
Ford Excursion	10,000	Jeep Grand Cherokee	6,500	Toyota Land Cruiser	6,500
Ford Expedition	8,100	Jeep Liberty	5,000		
GMC Jimmy	5,900	Land Rover Discovery	7,700		

Source: Manufacturers

- 13. Diameter of a Cookie** The data to the right represent the diameter (in inches) of a random sample of 34 Keebler Chips Deluxe™ Chocolate Chip Cookies.



In (a)–(d), start the first class at a lower class limit of 2.2000 and maintain a class width of 0.0200:

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- Construct a frequency histogram. Describe the shape of the distribution.
- Construct a relative frequency histogram.
- Repeat (a)–(d) using a class width of 0.0400. In your opinion, which class width provides the better summary of the data? Why?

2.3414	2.3010	2.2850	2.3015	2.2850	2.3019	2.2400
2.3005	2.2630	2.2853	2.3360	2.3696	2.3300	2.3290
2.2303	2.2600	2.2409	2.2020	2.3223	2.2851	2.2382
2.2438	2.3255	2.2597	2.3020	2.2658	2.2752	2.2256
2.2611	2.3006	2.2011	2.2790	2.2425	2.3003	

Source: Trina S. McNamara, student at Joliet Junior College

- 14. Home Sales** The data to the right represent the closing price (in U.S. dollars) of homes sold in a midwest city.



Start the first class at a lower class limit of 85,000 and maintain a class width of 10,000.

- Construct a frequency distribution.
- Construct a relative frequency distribution.
- Construct a frequency histogram. Describe the shape of the distribution.
- Construct a relative frequency histogram.

138,820	149,143	99,000	115,000	157,216
169,541	140,794	136,924	124,757	149,380
135,512	153,146	136,833	128,429	136,529
147,500	120,936	95,491	115,744	119,900
89,900	102,696	149,634	123,103	126,630
140,269	183,000	133,646	121,225	121,524
146,439	182,000	110,128	109,520	104,640
124,760	134,305	111,220	121,795	170,072
136,550	115,595	155,507	152,600	130,000
152,537	163,165			

Source: Transamerica Intellitech

- 15. Eat Your Vegetables!** The data to the right represent the number of servings of vegetables per day that a random sample of forty 20- to 39-year-old females consumes. The data are based on a survey conducted by the U.S. Department of Agriculture. Construct a stem-and-leaf diagram of the data, and comment on the shape of the distribution.



1.7	2.7	0.3	3.5	0.7	1.4	5.1	3.9
0.2	2.1	4.1	5.8	3.8	0.4	6.1	0.7
2.4	11.1	3.5	6.7	2.3	4.9	5.9	0.4
3.3	0.8	7.6	10.2	5.8	2.6	0.6	3.2
0.5	2.4	4.9	2.3	8.3	6.0	5.3	3.5

16. Fertility Rate The following data represent the fertility rate (births per 1000 women aged 15 to 44) for each state and the District of Columbia.

(a) Round each observation to the nearest whole number and draw a stem-and-leaf diagram.

(b) Describe the shape of the distribution.

(c) Redraw the stem-and-leaf diagram using split stems. For example, data between 60 and 64 is one stem and data between 65 and 69 is a second stem. Does this stem-and-leaf diagram better summarize the data? Why?



State	Fertility Rate	State	Fertility Rate	State	Fertility Rate
Alabama	61.2	Kentucky	60.5	North Dakota	58.7
Alaska	73.5	Louisiana	65.4	Ohio	61.7
Arizona	77.8	Maine	49.8	Oklahoma	68.8
Arkansas	66.6	Maryland	60.6	Oregon	61.9
California	68.3	Massachusetts	56.7	Pennsylvania	56.4
Colorado	69.3	Michigan	60.7	Rhode Island	54.6
Connecticut	58.8	Minnesota	62.0	South Carolina	60.7
Delaware	62.2	Mississippi	65.7	South Dakota	68.3
District of Columbia	52.9	Missouri	62.1	Tennessee	62.2
Florida	62.5	Montana	60.3	Texas	77.1
Georgia	68.4	Nebraska	69.5	Utah	90.6
Hawaii	68.6	Nevada	72.5	Vermont	48.9
Idaho	73.8	New Hampshire	52.4	Virginia	61.9
Illinois	66.1	New Jersey	63.5	Washington	60.2
Indiana	64.8	New Mexico	70.7	West Virginia	57.0
Iowa	61.7	New York	59.8	Wisconsin	59.0
Kansas	68.7	North Carolina	65.4	Wyoming	63.6

Source: U.S. National Center for Health Statistics

17. Federal Minimum Wage Rates The following data represent the value of the minimum wage for the years 1980 to 2003.



Year	Minimum Wage	Year	Minimum Wage
1980	3.10	1992	4.25
1981	3.35	1993	4.25
1982	3.35	1994	4.25
1983	3.35	1995	4.25
1984	3.35	1996	4.75
1985	3.35	1997	5.15
1986	3.35	1998	5.15
1987	3.35	1999	5.15
1988	3.35	2000	5.15
1989	3.35	2001	5.15
1990	3.80	2002	5.15
1991	4.25	2003	5.15

Source: Economic Policy Institute

(a) Construct a time-series plot of the data.
 (b) Comment on the apparent trend.

18. Federal Minimum Wage Rates The following data represent the value of the minimum wage for the years 1980 to 2003 in constant 2003 dollars. Constant dollars are dollars adjusted for inflation.

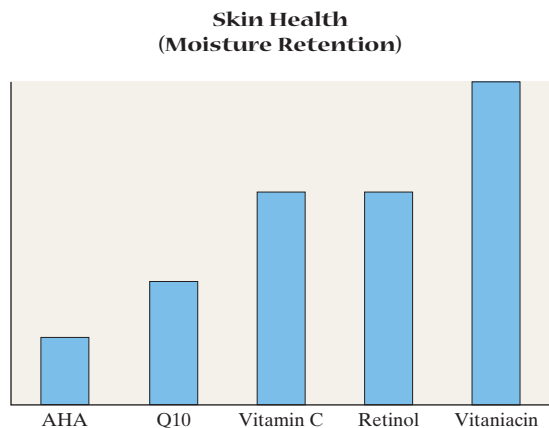


Year	Minimum Wage	Year	Minimum Wage
1980	6.55	1992	5.46
1981	6.48	1993	5.33
1982	6.11	1994	5.22
1983	5.87	1995	5.09
1984	5.64	1996	5.54
1985	5.46	1997	5.89
1986	5.36	1998	5.80
1987	5.19	1999	5.68
1988	5.01	2000	5.50
1989	4.80	2001	5.35
1990	5.19	2002	5.27
1991	5.60	2003	5.15

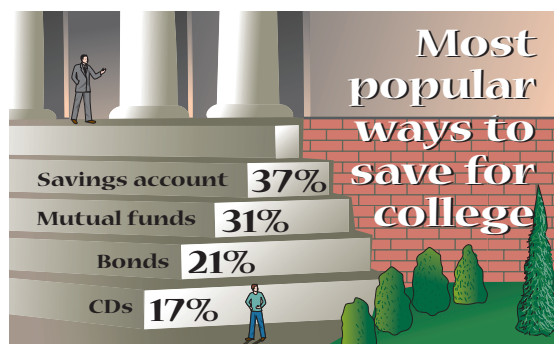
Source: Economic Policy Institute

(a) Construct a time-series plot of the data.
 (b) Comment on the apparent trend.
 (c) Compare this time-series plot with the one in Problem 17. Which graph is misleading? Why?

- 19. Misleading Graphs** The following graph was found in a magazine advertisement for skin cream. How is this graph misleading?



- 20. Misleading Graphs** The following is a *USA Today* type graph.



Do you think the graph is misleading? Why? If you think it is misleading, what might be done to improve the graph?

- 21. Misleading Graphs** In 2002 the average earnings of a high school graduate were \$27,280. At \$51,194, the average earnings of a recipient of a bachelor's degree were about 88% higher.

(Source: U.S. Census Bureau, Current Population Survey, 2003)

- Construct a graph that a college recruiter might create to convince high school students that they should attend college.
- Construct a graph that does not mislead.

THE CHAPTER 2 CASE STUDY IS LOCATED ON THE CD THAT ACCOMPANIES THIS TEXT.