8

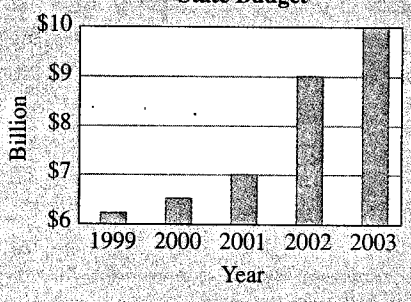
Descriptive Statistics—Data and Patterns

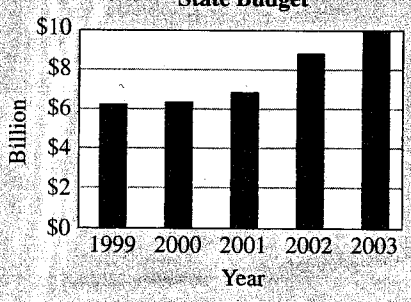
Action Proposed to Fight the Bloated Budget

During his reelection campaign, a state senator presented data illustrating the growth of the state budget over the years from 1973 to 2003. Using a large chart, he called attention to the fact that between 1993 and 2003, the state budget more than doubled. The senator said, "The state budget has been growing out of control. This irresponsible spending must be reined in and I'm the man to do it."

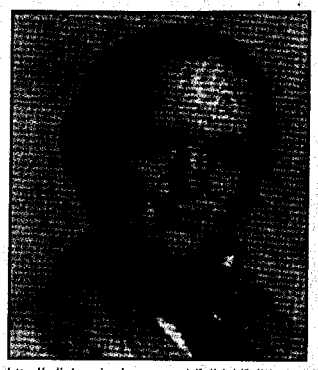
While the details have been changed, this story is based on an actual mailing from a state senator. The graph at the bottom and on the left shows the senator's chart which begins at 6 billion dollars instead of zero dollars. That design choice makes the budget increases look even larger than they are. (A similar choice of design for the chart was used in the actual mailing.) Although the senator may be correct in asserting that the budget growth is out of control, his constituents should examine the facts for themselves. In contrast, the graph below on the right shows the same data, but the vertical scale of this graph begins at zero dollars rather than at $6 billion. Which graph more fairly presents the state budget growth?

JupiterMedia/Alamy





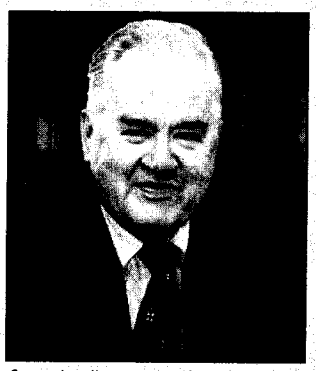
WILLIAM PLAYFAIR (1759–1823) invented many standard graph forms used today, including bar charts, line graphs, and pie charts. He was the son of James Playfair, a minister. William had an older brother John, a mathematician and geologist, who led the family when their father died. John taught William that, "whatever can be expressed in numbers may be expressed in a line." William lived during a time of great revolution in ideas and governments, both in Europe and America. He saw the value of persuading the common people instead of merely the aristocracy. He discovered that a chart with a good visual design provided more information and had more impact than long intricate arguments full of calculations and tables. Playfair also found ways to graph data in order to exaggerate the point that he was trying to make. His most famous graph, which showed the rise in the price of wheat versus the rise in wages, was misleading. He applied insights from the humanities, mathematics, engineering, and geography when constructing his charts. His work was so effective that it soon permeated all forms of communication.



http://edinburgh-places.co.uk/folk/oldfolk/william-playfair.html

These days, visual displays are one of our primary methods for communicating quantitative information. Books, newspapers, magazines, and television present a wide variety of visuals to inform, impress, or persuade the viewer. Few people realize that bar graphs, line graphs, and pie charts are the result of one person's genius. In addition to inspiring William in this area, his brother John also became famous for his alternative to Euclid's parallel postulate: "Through a given point only one line can be drawn parallel to a given line."

JOHN WILDER TUKEY (1915–2000), like Playfair before him, also created new types of graphs. His parents recognized that he was a prodigy when he was very young, and they decided to school him at home. Doing so was convenient because both of his parents were teachers. Tukey's formal education did not begin until he entered Brown University, where he earned a degree in chemistry. He wrote his doctoral thesis in mathematics at Princeton on an abstract topic considered to have little practical value. One of his early theorems generalized the *Ham-Sandwich Theorem*, which says that given any three regions in space (imagine ham, cheese, and bread—a ham sandwich) there exists a plane such that equal volumes of each region lie on either side of the plane (that is, the plane cuts the sandwich in half). During World War II, Tukey changed the direction of his work to graphical displays and statistics. He was extremely prolific and made important contributions in a variety of fields, ranging from astrophysics to global pollution. Most of the graphic forms developed during this century, such as stem-and-leaf plots and box-and-whisker plots, were created by Tukey. However, these graphs are only a small part of Tukey's work. He invented the term *bit* (short for *binary digit*) in 1946, and he was the first to use the word *software* in print in 1958.



Source: http://www-groups.dcs.stand.ac.uk/~history/PictDisplay/Tukey.html

Nearly all of the graphs and charts that we commonly use to communicate and analyze information are the inventions of two people, Playfair and Tukey, one, born in the 18th century, with no background in mathematics and one, born in the 20th century, with a doctorate in one of the most abstract areas of mathematics.

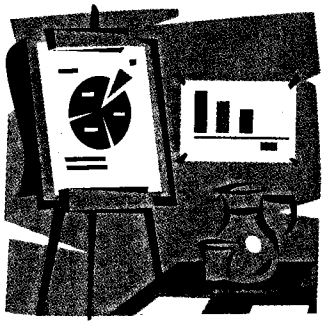
OBJECTIVES

In this chapter you will learn:

1. how to create and interpret graphs that display and compare data.
2. which types of graphs are appropriate for representing various types of data, and
3. how to recognize the ways graphs distort information.

8.1 Organizing and Picturing Data

INITIAL PROBLEM



Suppose that you have to give a sales report showing the sales figures of each of three districts with markets that are roughly equal in size. In 2003, district A had $135,000 in sales, district B had $85,000 in sales, and district C had $115,000 in sales. How might you present these data to clearly compare each district with the others?

A solution of the Initial Problem is on page 489.

Statistics is the science (and art) of making sense out of data. Our world is filled with numerical information. The trick is to organize that information in some meaningful way. This chapter describes how to organize sets of numbers, called **data sets**, into sensible visual patterns and charts. It also shows how the eye may be misled by such visual representations.

OBTAINING DATA

Sometimes as researchers we are provided with numerical data to organize and analyze. On the other hand, there are times when we need to seek out data. One of the best sources of data is the World Wide Web. Excellent print sources of data include the *Statistical Abstract of the United States*, published annually by the government, as well as newspapers and magazines. In particular, the *Wall Street Journal* and *Investor's Business Daily* are respected sources for financial data.

Another method for obtaining data is to run a **designed experiment** in which the researcher controls as many variables as possible. Typical of this type of data collection is a clinical study in which one group of patients receives a drug and another, comparable, group of patients receives a placebo.

A third means of obtaining data is the **observational study**, in which the researcher observes the objects of study in their natural setting and records the relevant data. For example, a kindergarten class could be observed at play and the size of the playgroups recorded.

In this book, we will devote most of our attention to a fourth method of obtaining data, the survey. In a **survey**, the researcher selects a sample from the entire group in which the researcher is interested, called the **population**, and measures the variables of interest with a questionnaire or by interviewing the subjects. For example, a political researcher might conduct a survey to find out whether residents of a particular county support a ballot measure. This may sound easy, but in order to obtain useful survey data, the researcher must carefully select the sample and carefully word the survey questions. We discuss methods for selecting a representative sample in Chapter 9.

Once we have collected our data, the next question is, "What do these data tell us?" By using **exploratory data analysis**, we can begin to answer that question. The goal of exploratory data analysis is to take an initial look at the data's big picture to see what patterns may emerge or to get an indication of further research needed. The researcher might notice trends not anticipated when the data collection was planned, or the data may indicate that there was some flaw in the data-collection procedure. One excellent way to carry out an exploratory data analysis is to present the data in a pictorial form. Pictures can often convey information more rapidly and more effectively than words or a list of numbers. Thus, in addition to allowing us to explore the data, pictorial representations can also be used to communicate our findings and their significance to others.

We can represent the data pictorially using a computer, a graphing calculator, or a drawing. In this section, we will discuss some common types of graphs and charts, and the principles discussed will apply whether the graphs are hand-drawn or computer-generated.

Tidbit

It has been estimated that each year there are between 1 trillion and 3 trillion charts printed. This is on the order of 400 for each human on Earth.

DOT PLOTS

Consider the following list of numbers:

80, 74, 87, 62, 96, 87, 71, 93, 32, 76, 26, 81, 84, 54, 70, 87, 89, 71, 95, 67.

What can we say about these raw data? Initially, we can say very little. There is no context to give meaning to these numbers and there is no clear trend or pattern to them. Actually, the most noticeable characteristic of this collection of numbers may be that they are boring. To provide some context, suppose these numbers are test scores from an economics class. We can see that there are 20 scores between 100 and 0 and that the highest score is 96, while the lowest score is 26.

As a first step, we can organize these data by putting the numbers in order. Usually we arrange numbers from lowest to highest, which gives us the following list:

Economics 101 Test Scores

26, 32, 54, 62, 67, 70, 71, 71, 74, 76, 80, 81, 84, 87, 87, 87, 89, 93, 95, 96.

Another way to represent these data is to draw a graph. A **dot plot**, such as the one shown in Figure 8.1, could be used to get an initial graphical picture of the data.

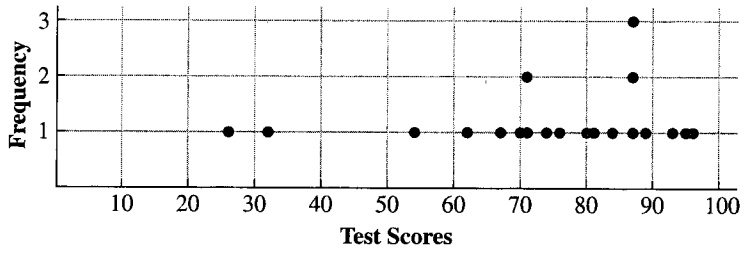


Figure 8.1

To create a dot plot for the test score data, we let the horizontal axis represent test scores, choosing a convenient scale: 10, 20, 30, and so on. The vertical axis indicates the frequency of scores—that is, the number of times each score occurs. Then we plot one dot for each test score in the data set. Notice that one dot is placed above 26, since there is one 26 in the data set, whereas the three dots above 87 represent the three 87s in the data set. Notice how a dot plot can be used to quickly identify the numbers that occur most frequently (the tallest column of dots) as well as gaps in the data and scores that are widely separated from others. So, a quick glance at this dot plot tells us that 87 was the most common economics test score and that there were a couple of low scores in the 20s and 30s.

STEM-AND-LEAF PLOTS

Another popular method of arranging data is to use a stem-and-leaf plot. To make a **stem-and-leaf plot** of the data from the Economics 101 test scores, we first group the scores by 10s (Table 8.1). The labels on the left-hand side of the vertical line identify the rows. Notice that the tens digits are also written on the right-hand side of Table 8.1. For example, in the 80s row, the 8 is repeated in each score on the right-hand side. Also notice that there are blank rows because there are no scores in the 0s, 10s, or 40s. To present the list of scores more efficiently, we can eliminate the repeated digits (Table 8.2). Notice that the zeros in 10-90 were dropped from the left-hand column in Table 8.1. This column in Table 8.2 is called the **stem** of the plot. The ten's digits are dropped from the right-hand side of Table 8.1 and the ones digits are arranged in order in Table 8.2. These digits are called the **leaves** of the plot. In Table 8.2, the top row in the stem-and-leaf plot, 9 | 3 5 6, indicates that the three scores in the 90s were 93, 95, and 96.

Tidbit

Stem-and-leaf plots were invented by the father of modern data analysis, John Tukey. He wrote "If we are going to make a mark, it may as well be a meaningful one. The simplest—and most useful—meaningful mark is a digit."

Table 8.1

90s	93, 95, 96
80s	80, 81, 84, 87, 87, 87, 89
70s	70, 71, 71, 74, 76
60s	62, 67
50s	54
40s	
30s	32
20s	26
10s	
0s	

Table 8.2

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

This visual way of arranging the data allows us to see at a glance the relative sizes of each category. For example, we can see that there are more scores in the 80s than in any other category. Notice that most of the data between 54 and 96 form a **cluster** of scores. There is a large **gap** between 54 and the scores 32 and 26. Scores such as 32 and 26 that are separated from the others by large gaps are called **outliers**. In discussing a stem-and-leaf plot, the terms *cluster*, *gap*, and *outlier* are imprecise terms that might be interpreted differently by different people. However, they often reveal useful information about the data, as illustrated in the following example.

EXAMPLE 8.1 Jack has delivered 10 pizzas in the first 2 hours of his shift. The prices of pizzas Jack delivered, in increasing order, were

$9.20, $10.50, $10.70, $10.80, $10.80, $12.00, $12.10, $12.20, $12.20, $12.30.

Make a stem-and-leaf plot of these data. Identify any clusters, gaps, and outliers, and suggest an interpretation for them.

Table 8.3

12	0 1 2 2 3
11	
10	5 7 8 8
9	2

SOLUTION In creating the stem-and-leaf plot, it might be convenient to use the dollar amounts as the stem and tens of cents as the leaves. For example, in the stem-and-leaf plot shown in Table 8.3, 10 | 5 represents the pizza price of $10.50. Notice that there are two clusters separated by a gap. Also, $9.20 may be considered to be an outlier since it stands apart from the rest of the numbers. We might guess that the small pizzas cost about $9, medium pizzas about $10.70, and large pizzas cost a little over $12. If this interpretation is correct, roughly the same numbers of medium and large pizzas were ordered. ■

Most graphs are designed to show only the general pattern of the data but not all of the data. One advantage of the stem-and-leaf plot is that all of the actual data is still contained in the graph and is displayed in a way that makes it easy to see. For instance, we could re-create the list of pizza prices from the stem-and-leaf plot.

HISTOGRAMS

We have seen that stem-and-leaf plots group data into categories. Another type of graph frequently used to display data that has been separated into categories is called a **histogram**. Histograms group data into intervals called **measurement classes** or **bins**.

We will illustrate the process by constructing a histogram for the test scores from the economics class pictured in Figure 8.1. We first need to define our measurement classes. The stem-and-leaf plot in Table 8.2 shows the data grouped into measurement classes of length 10, such as 80 to 89, or length 11, in the case of 90 to 100. We will use those same intervals for our histogram.

Next, we need to look at the number of test scores in each measurement class. We can see from the stem-and-leaf plot that three test scores were in the 90s, seven scores were in the 80s, and so on. The number of data points in each measurement class is called the **frequency** of the interval. The information about the number of data points in each measurement class can be organized into a **frequency table** such as Table 8.4.

Table 8.4

Interval	Frequency
90–100	3
80–89	7
70–79	5
60–69	2
50–59	1
40–49	0
30–39	1
20–29	1
10–19	0
0–9	0

Using the data in the frequency table, we can now create a histogram. First we mark off the horizontal axis, dividing it into the same intervals as in the frequency table. Then we mark off the vertical axis with the frequencies. To make the histogram, we draw a bar for each measurement class. The width of each bar is the length of the interval on the horizontal axis. The height of each bar is equal to the frequency of that measurement class. The completed histogram is shown in Figure 8.2. Although the horizontal grid lines make the graph easier to read, they may also make it more cluttered; their use is optional.

Tidbit

Not every graph was invented by Playfair and Tukey. Florence Nightingale made a radial histogram to show death rates in the Crimean War. Her graphs led to an improvement in medical care.

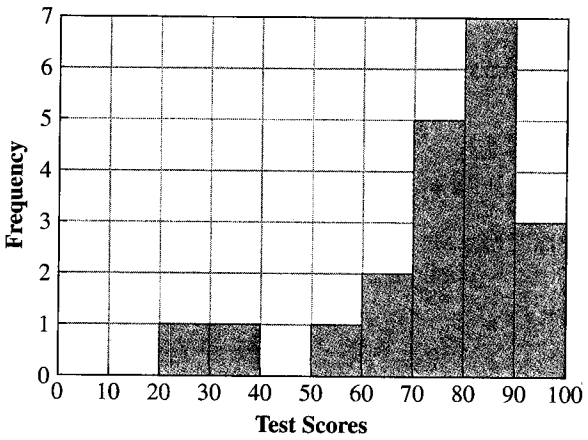


Figure 8.2

The histogram makes it easy to see the test results at a glance. It shows that while there were a few test scores in the 20s, 30s, and 50s, the bulk of the scores were in the 70s, 80s and 90s, and the greatest number of test scores was in the 80s. Notice that if you were given the histogram in Figure 8.2, you could use the heights of the bars to reconstruct the frequency table. However, you could *not* reconstruct the original list of test scores.

Sometimes a third column is added to a frequency table to create a **relative frequency table**. We know there are 20 economics test scores (either by counting them or by adding all of the frequencies listed in Table 8.4). Three scores fall in the measurement class 90–100, so the fraction of scores in that interval is $\frac{3}{20} = 0.15$. In other

words, 15% of the test scores fall in the interval 90–100. We say that 0.15 is the **relative frequency** of this measurement class. Table 8.5 shows a relative frequency table for the economics test score data. Notice that the third column in Table 8.5 lists the relative frequency of each measurement class, including 0.15 for the test scores in the interval 90–100.

Table 8.5

Interval	Frequency	Relative Frequency
90–100	3	0.15
80–89	7	0.35
70–79	5	0.25
60–69	2	0.10
50–59	1	0.05
40–49	0	0.00
30–39	1	0.05
20–29	1	0.05
10–19	0	0.00

Just as a frequency table was used to create a histogram, the relative frequency table can be used to create a relative frequency histogram. If the vertical scale on the graph represents the relative frequency, then the graph is called a **relative frequency histogram**. A relative frequency histogram for the economics test scores is shown in Figure 8.3.

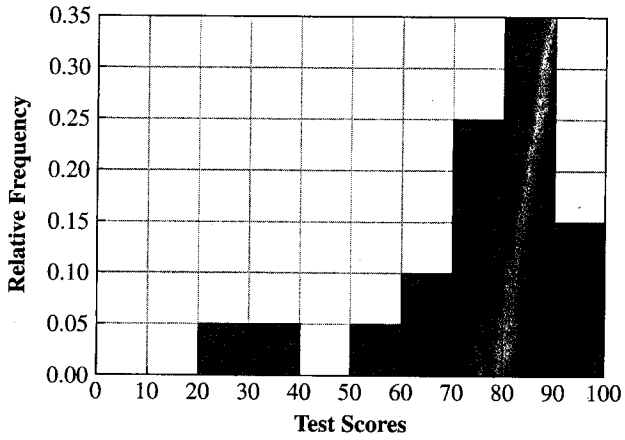


Figure 8.3

When constructing a histogram, we must carefully consider what size intervals would be best to use. If we group the data from the economics class into 5s rather than 10s, using the intervals of 30–34, 35–39, 40–44, and so on, we get the histogram in Figure 8.4. In this case, the intervals are a bit too small to show the general pattern clearly.

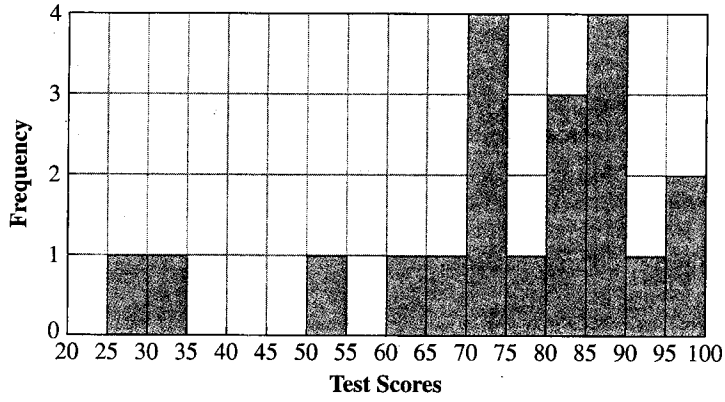


Figure 8.4

On the other hand, if we group the data into 20s using the intervals 20–39, 40–59, 60–79, and 80–100, we get the histogram in Figure 8.5. In this case, intervals are so large that they obscure much of the information. For example, it is now impossible to see that only two scores were in the 60s.

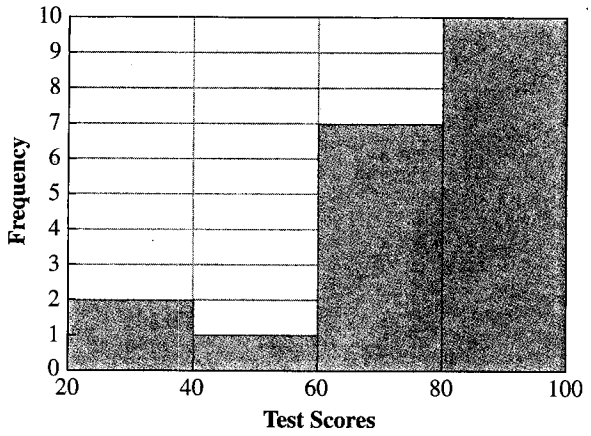


Figure 8.5

The histograms in Figures 8.2, 8.4, and 8.5 represent the same data yet give very different views of that data. Thus, when creating a histogram, choosing an appropriate interval size is an important first step.

EXAMPLE 8.2 Make a histogram of the pizza data from Example 8.1:

$9.20, $10.50, $10.70, $10.80, $10.80, $12.00, $12.10, $12.20, $12.20, $12.30.

SOLUTION Because the prices range between $9 and $13, it might be convenient to group the prices into measurement classes whose length is one dollar. Then pizzas costing $9.00 to $9.99 go into the first measurement class, pizzas costing $10.00 to $10.99 go into the second measurement class, etc. (Figure 8.6). We count the number of data points in each measurement class to determine the height of each bar. For example, in the class $9.00–9.99, we have only one pizza, so that bar is one unit tall.

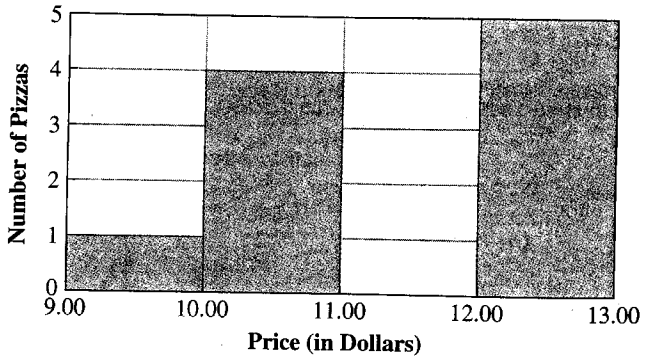


Figure 8.6

If we want to get a finer picture of these data we could group the pizza prices into smaller measurement classes, say measurement classes whose length is 50 cents, as shown in Figure 8.7.

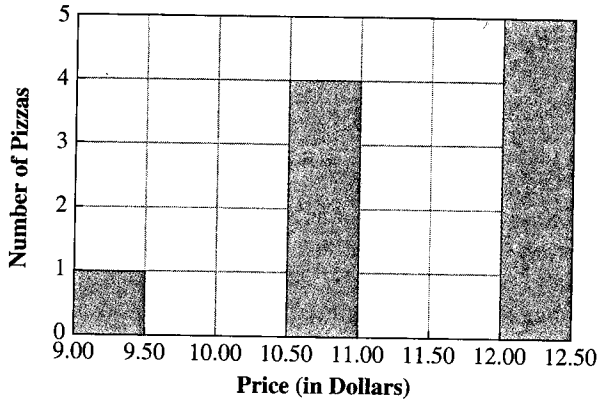


Figure 8.7

An outlier becomes apparent in the histogram in Figure 8.7. It might represent a choice of a small pizza with no toppings that was very cheap, but not very popular. ■

BAR GRAPHS

The bars in a histogram represent the frequency of data in the given measurement classes. However, the bars in a graph may also be used to represent information other than measurement classes or intervals; that is, histograms are a special case of a large class of graphs called **bar graphs** or **bar charts**. A bar graph is any graph in which the height or length of bars is used to represent frequencies or quantities.

The data listed in Table 8.6 suggest the financial rewards of an education. The table shows the median (middle) annual earnings of year-round, full-time workers, both male and female, aged 21–64, according to the 2000 census.

Table 8.6

Educational Attainment	1999 Median Income
Not a high-school graduate	$21,322
High-school graduate	$27,351
Some college	$31,988
Bachelor's degree	$42,877
Advanced degree	$55,242

Source: www.census.gov/hhas/income/earnings/callus.both.html.

A bar graph of the data in Table 8.6 is displayed in Figure 8.8. The different levels of education are represented on the horizontal axis, and the median income levels are shown on the vertical axis.

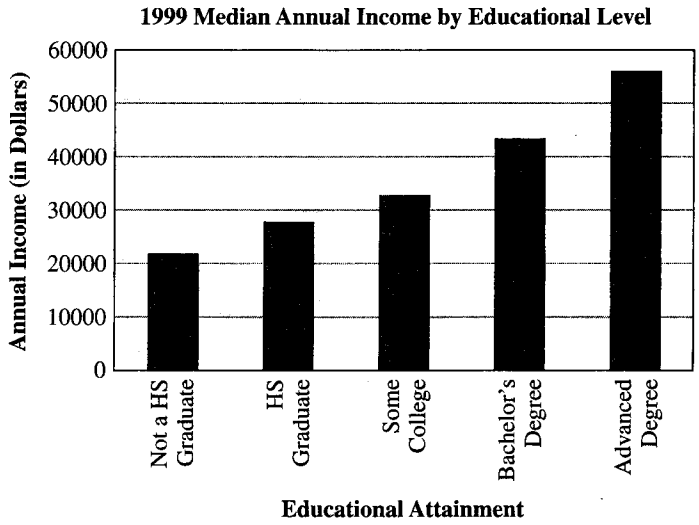


Figure 8.8

Each bar in the bar graph corresponds to an educational level and the height of each bar is proportional to the income level associated with that educational level. Notice how the heights of the bars provide a quick visual summary of the data. So, for example, we can see that income generally increases with educational level, and the highest income level is associated with advanced degrees.

When we make a bar graph, the vertical scale should be just a bit larger than the largest value. In Figure 8.8, the vertical scale goes from 0 to 60,000 because the largest income value listed in the table was $55,242. Starting the vertical scale at 0 gives a fair comparison of the amounts. The widths of the bars are chosen to allow the graph to fit in the available space and to make the graph easy to read and interpret. This scaling generally happens automatically when using computer software to create a bar graph.

Note that this information does not *prove* that more education will increase your income, although it is highly suggestive. It is possible that this merely reflects the possibility that people who have greater opportunities for education also have greater opportunities to make a lot of money.

Bar graphs frequently appear in newspapers, magazines, and other news media. The next example provides one illustration of how a bar graph can present a lot of information in a compact format that allows readers to analyze the data quickly.

EXAMPLE 8.3 The two bar graphs in Figure 8.9, which appeared in a newspaper article, provide details about the 2002 wildfire season and the states most affected. Based on the information in the graphs, which state suffered the greatest loss in terms of acres burned? Which state had the greatest number of fires?

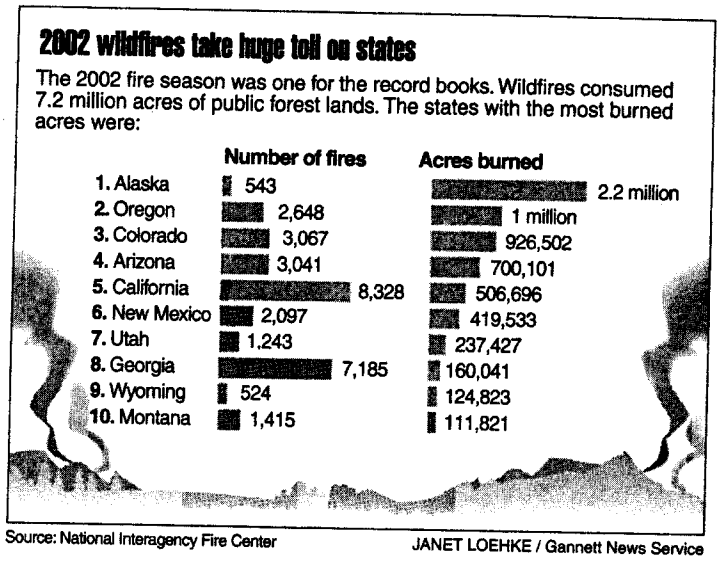


Figure 8.9

Source: Statesman Journal Newspaper, April 12, 2003. Permission courtesy of Gannett News Service

SOLUTION Notice that the bars in these graphs are horizontal rather than vertical and that the scale of the graph is omitted. Instead, the numerical value that would determine the length of a bar is indicated at the end of the bar, a common practice for bar graphs in the media. Based on the lengths of the bars, it can easily be seen that California, with 8328, had the greatest number of wildfires. Alaska, with 2.2 million acres burned, suffered by far the greatest loss in terms of acreage, although it had a relatively small *number* of wildfires. ■

Bar graphs are often used to show trends over time. The bar graph shown in Figure 8.10 illustrates how funding for a university increased over the period from 1999 through 2003.

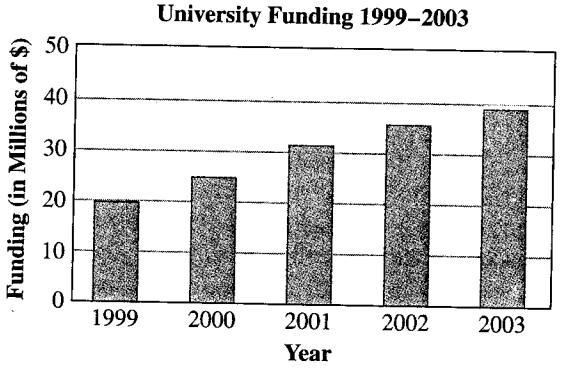


Figure 8.10

EXAMPLE 8.4 Use the bar graph in Figure 8.10 to determine the level of funding for the university for each year from 1999 through 2003. Which year showed the greatest increase over the previous year?

SOLUTION Reading across the horizontal lines and estimating heights of bars whenever necessary, we can see that in 1999, the university's funding was $20 million, in 2000 it was $25 million, in 2001 it was $31 million, in 2002 it was $35 million, and in 2003 it was $38 million. Looking for the greatest increase in bar heights, we find that the year that showed the greatest gain over the previous year was 2001, with an increase of approximately $6 million over 2000. ■

LINE GRAPHS

We saw in Figure 8.10 that a bar graph can illustrate how a quantity changes over time. A **line graph**, in which data points are connected by line segments, can also show how a variable behaves over time. As an example, we have plotted the same data for the university's funding levels in Figure 8.11 that we showed in the bar graph in Figure 8.10. Each dot in Figure 8.11 represents the level of funding for a particular year and is the same height as the corresponding bar in the bar graph of Figure 8.10. We complete the line graph by connecting consecutive dots with line segments. Note that a "line" graph is not necessarily a straight line. Line graphs are particularly useful in showing trends and variation over time.

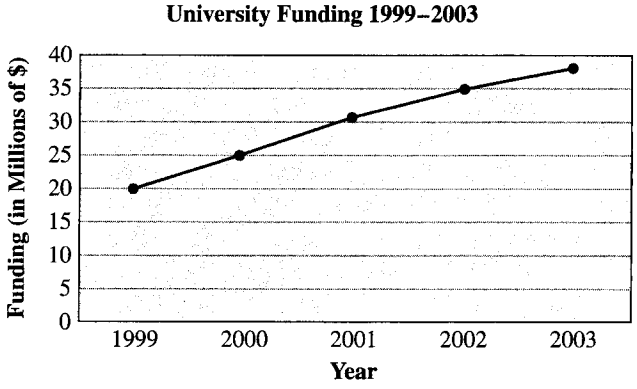


Figure 8.11

It is clear from the bar graph in Figure 8.11 that the trend in funding is up; that is, funding is increasing. However, a closer examination reveals more about the trend in funding. Notice that the line segments on the right are less steep than the line segments on the left part of the graph. This change in steepness (or slope) indicates that funding was increasing a little faster prior to 2001 than after 2001.

Line graphs are frequently used in the media to show historical information. The following example provides one illustration.

EXAMPLE 8.5 A newspaper used the line graph in Figure 8.12 to show the trend in college completion over the years 1965 through 2000 among persons in the United States aged 25–29. Based on the graph, approximately what percentage of adults aged 25–29 had completed at least 4 years of college in 1980? In 2002? What can you conclude about the trend in the percentages? Was the percentage increasing, decreasing, or remaining constant over the years 1990 through 1992?

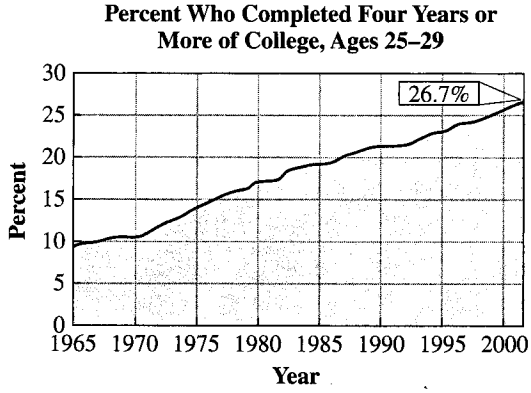


Figure 8.12

SOLUTION In this line graph, the line segments are short enough to make the graph resemble a curve. Reading from the graph, it appears that in 1980, approximately 17.5%

of persons in the 25–29-year-old age bracket had completed 4 or more years of college. In 2002, the percentage had increased to 26.7%. Note that in order to emphasize the most recent statistic, the percentage for 2002 was highlighted. Clearly, the general trend is up, although the graph has a few dips. In general, then, more young people are completing 4 or more years of college as time goes by. However, for the period of time between 1990 and 1992, the graph is basically flat, meaning that the percentage remained constant during those years. ■

PIE CHARTS

Sometimes it is useful to see a comparison of the percentages of a whole. **Pie charts** (also called **circle graphs**) are often used to show relative proportions of quantities by using wedge-shaped portions of the interior of a circle. You may have seen a pie chart showing where your tax dollar goes, for example. Pie charts are especially useful for displaying information about budgets because they provide a way to visualize the ratios between the various expenditures. The pie chart in Figure 8.13 shows the average number of hours of sleep a certain group of adults get during a night.

Sleep Times for Adults

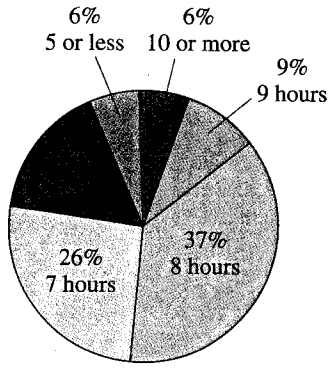


Figure 8.13

It is clear from the pie chart that most people in this group sleep 7 or 8 hours a night. Also, a surprisingly high percentage of people (6%, or more than 1 of 17 people) get 5 hours or less of sleep each night.

We can construct a pie chart by first determining what portions of a circle each part should represent. In the chart in Figure 8.13, the percentages 6%, 6%, 9%, 16%, 26%, and 37% are represented. If we wanted to draw this pie chart by hand, we would first need to calculate the size of sector (the piece of the pie), corresponding to each category. For example, we observe that 26% of the adults in the group got 7 hours of sleep. So, the part of the pie chart corresponding to 7 hours of sleep should be 26% of the pie. Remember that there are 360° in a circle. Thus, the angle corresponding to 7 hours of sleep should be 26% of 360°, or $0.26 \times 360° \approx 94°$.

Performing the same calculations to find the sizes of all of the pieces of the pie, we find that angles of the pieces are $6\% \times 360° \approx 22°$, $6\% \times 360° \approx 22°$, $9\% \times 360° \approx 32°$, $16\% \times 360° \approx 58°$, $26\% \times 360° \approx 94°$, and $37\% \times 360° \approx 133°$. (*Note:* Due to rounding, the sum of these six angles is greater than 360°, but close enough for measuring angles with a protractor.) After calculating the angles, we can construct the pie chart by hand using a compass to draw a circle and then measuring the appropriate angles with a protractor.

Tidbit

A 2001 poll conducted by the National Sleep Foundation found that the average American adult sleeps 7 hours on weekdays, but that may not be a bad thing. Some recent research suggests that 8 hours of sleep per night may not actually be optimal. In a 6-year study of more than 1 million adult men and women, researchers discovered that those who slept 8 hours per night or less than 4 hours per night had a lower survival rate than those who slept 6 or 7 hours per night. However, there is no evidence that changing your sleep habits will actually increase your longevity.

CHOOSING AN APPROPRIATE GRAPH

We can create pie charts, as well as many other graphs discussed in this section, on a computer by using a graph and chart feature of a word-processing program or by using a spreadsheet program. Whether creating graphs by hand or on a computer, we still need to decide what type of graph best illustrates the information we want to present. Table 8.7 summarizes this section's graphs and the uses for which they are best suited.

Table 8.7

Type of Chart/Graph	Use
Stem-and-leaf plot	Displays numerical data grouped into categories (all the data can be recovered from the plot)
Histogram	Displays data grouped into measurement classes
Bar chart	Displays data grouped by category, can be used to show trends
Line graph	Displays trends and variation
Pie chart	Displays and compares percentages of a whole

EXAMPLE 8.6 It has often been reported that workers in the United States work more than workers in other industrialized nations. Table 8.8 shows the total average number of hours worked in one year by employees in the United States and by workers in four other countries. What type of graph would convey this information in a visual way most effectively?

Table 8.8

Country	Number of Hours Worked Per Year
United States	1966
Japan	1889
Great Britain	1731
France	1656
Germany	1574

Source: www.interbiznet.com/hunt/archives/020207.html.

SOLUTION Because we will not be showing a trend over time or percentages of a whole, we can rule out a line graph and a pie chart, respectively. We would like to easily compare the number of hours worked, so a bar graph might be a good choice. In this case, the "categories" would be the countries. The heights of the bars in the graph will represent the hours worked. One possible bar graph for these data is shown in Figure 8.14.

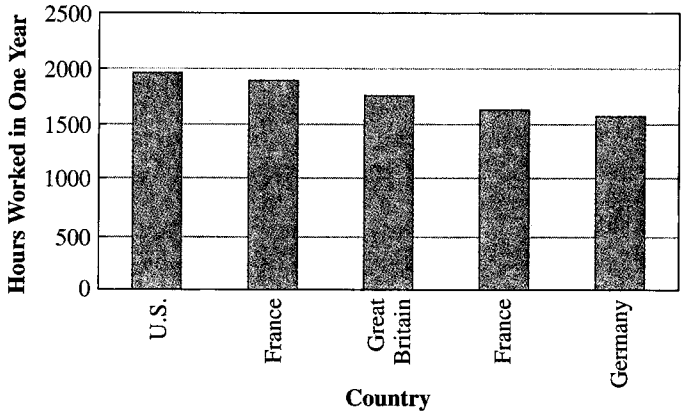
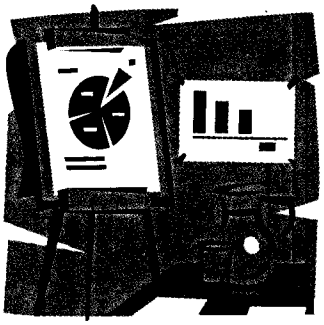


Figure 8.14

Note that while the differences in the heights of the bars may not appear significant at first glance, they may be significant to the workers. An average worker in the United States works 77 (or 1966 − 1889) more hours per year than a Japanese worker, which corresponds to about 2 weeks of work, assuming a 40-hour work week. The difference between a U.S. worker and a German worker is even more dramatic. A typical U.S. worker works 392 (or 1966 − 1574) more hours in 1 year, which is the equivalent of nearly 10 weeks of work! ■

SOLUTION OF THE INITIAL PROBLEM



Suppose that you have to give a sales report showing the sales figures of each of three districts with markets that are roughly equal in size. In 2003, district A had $135,000 in sales, district B had $85,000 in sales, and district C had $115,000 in sales. How might you present these data to clearly compare each district with the others?

SOLUTION Although the sales data could be presented as a bar chart, it is better to use a pie chart to show the comparisons because it will show the proportion of sales for each of the districts. To do this, we first compute the total sales:

$$\$135{,}000 + \$85{,}000 + \$115{,}000 = \$335{,}000.$$

Then, we find what portion of a circle (360°) each of the district sales represents.

$$\text{District A: } \frac{135{,}000}{335{,}000} \times 360° \approx 0.403 \times 360° = 40.3\% \times 360° \approx 145°$$

$$\text{District B: } \frac{85{,}000}{335{,}000} \times 360° \approx 0.254 \times 360° = 25.4\% \times 360° \approx 91°$$

$$\text{District C: } \frac{115{,}000}{335{,}000} \times 360° \approx 0.343 \times 360° = 34.3\% \times 360° \approx 124°$$

Sketching a circle with these angles at the center gives the pie chart in Figure 8.15.

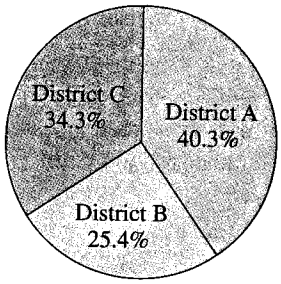
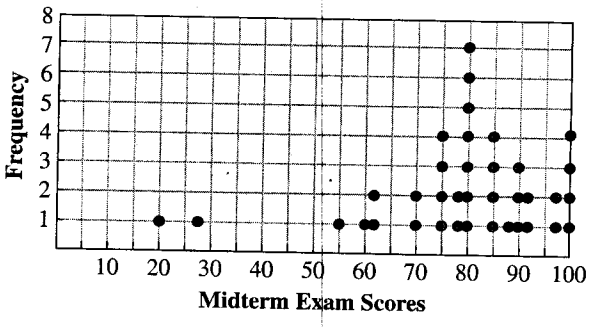


Figure 8.15

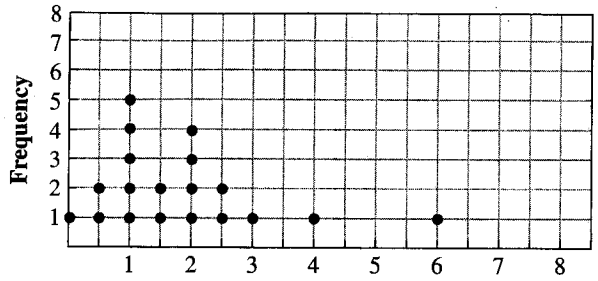
PROBLEM SET 8.1

1. Consider the following dot plot of midterm exam scores in a biology class.



a. What does the horizontal axis represent?
b. What does the vertical axis represent?
c. How many exam scores were there?
d. What was the most frequent exam score?
e. What were the high score and the low score?

2. Consider the following dot plot of the number of hours a class of second-graders spent watching television after school.



Number of Hours Spent Watching Television after School

   a. What does the horizontal axis represent?
   b. What does the vertical axis represent?
   c. How many students were in the second-grade class?
   d. What was the most frequent number of hours spent watching television?
   e. What was the largest amount of time spent watching television? What was the smallest amount of time spent watching television?

3. In the 2002 Salt Lake City Winter Olympic Games, athletes or teams from different countries earned medals. The total number of medals earned by each medal-winning country is summarized in the following stem-and-leaf plot.

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

Source: www.saltlake2002.com/main.html.

   a. How many countries received medals?
   b. How many medals were awarded?
   c. What is the largest number of medals earned by athletes from a single country?

4. The heights, in feet and inches, for the 2003–2004 Dallas Mavericks are given in the following stem-and-leaf plot.

7	0 6
6	2 3 6 6 7 8 8 9 9
5	11

Source: www.espn.com.

   a. How many players were on the team roster?
   b. List the height of each player.
   c. How tall was the tallest player for the Dallas Mavericks? How tall was the shortest player?

5. A sample of starting salaries for recent graduates of a university's accounting program is as follows. Salaries have been rounded to the nearest hundred dollars.

$29,500	$29,500	$27,800	$27,800
$27,100	$31,000	$28,200	$25,600
$26,800	$35,400	$29,400	$26,800
$28,800	$28,200	$30,300	$30,200

   a. Make a stem-and-leaf plot of the rounded salaries. Show the leaves in terms of 100s.
   b. Most of the data values cluster between what two salaries?
   c. Describe any gaps in the data. Do there appear to be any outliers? If so, where?

6. In 1798, the English scientist Henry Cavendish measured the density of the Earth in an experiment with a torsion balance. He made 29 repeated measurements with the same instrument and obtained the following data in grams per cubic centimeter. (*Source: Annals of Statistics*, 5: 1055–1078, 1977.)

5.50	5.61	4.88	5.07
5.26	5.55	5.36	5.29
5.58	5.65	5.57	5.53
5.62	5.29	5.44	5.34
5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34
5.46	5.30	5.75	5.68
5.85			

   a. Make a stem-and-leaf plot of Cavendish's experimental data. Show the leaves in terms of hundredths.
   b. Most of the data values cluster between what two densities?
   c. Describe any gaps in the data. Do there appear to be any outliers? If so, where?

Problems 7 and 8

A baseball player's batting average gives the ratio of hits to times at bat and is usually expressed in decimal form, rounded to three decimal places.

7. Make a stem-and-leaf plot of the following American League batting champion averages. Is there a general pattern or shape to your graph? What observations might you make about the batting averages based on your graph?

AMERICAN LEAGUE BATTING CHAMPIONS (1978–2003)		
1978	Carew	0.333
1979	Lynn	0.333
1980	Brett	0.390
1981	Lansford	0.336
1982	Wilson	0.332
1983	Boggs	0.361
1984	Mattingly	0.343
1985	Boggs	0.368
1986	Boggs	0.357
1987	Boggs	0.363
1988	Boggs	0.366
1989	Puckett	0.339
1990	Brett	0.328
1991	Franco	0.341
1992	Martinez	0.343
1993	Olerud	0.363
1994*	O'Neill	0.359
1995	Martinez	0.356
1996	Rodriguez	0.358
1997	Thomas	0.347
1998	Williams	0.339
1999	Garciaparra	0.357
2000	Garciaparra	0.372
2001	Suzuki	0.350
2002	Ramirez	0.349
2003	Mueller	0.326

*Shortened season (baseball strike)
Source: www.espn.com.

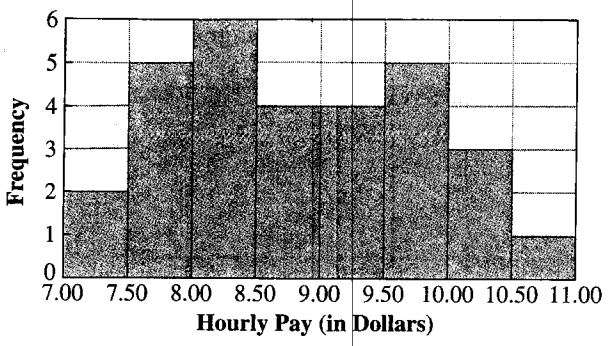
8. Make a stem-and-leaf plot of the following National League batting champion averages. Is there a general pattern or shape to your graph? What observations might you make about the batting averages based on your graph?

NATIONAL LEAGUE BATTING CHAMPIONS (1978–2003)		
1978	Parker	0.334
1979	Hernandez	0.344
1980	Buckner	0.324
1981	Madlock	0.341
1982	Oliver	0.331
1983	Madlock	0.323
1984	Gwynn	0.351
1985	McGee	0.353
1986	Raines	0.334
1987	Gwynn	0.370
1988	Gwynn	0.313
1989	Gwynn	0.336
1990	McGee	0.335
1991	Pendleton	0.319
1992	Sheffield	0.330
1993	Galarraga	0.370
1994*	Gwynn	0.394
1995	Gwynn	0.368
1996	Gwynn	0.353
1997	Gwynn	0.372
1998	Walker	0.363
1999	Walker	0.379
2000	Helton	0.372
2001	Walker	0.350
2002	Bonds	0.370
2003	Pujols	0.359

*Shortened season (baseball strike)
Source: www.espn.com.

Compare the stem-and-leaf plots you made for problems 7 and 8. Do the batting averages cluster around different values in the two graphs?

9. The following histogram shows the frequencies of hourly pay for a group of students who worked during the summer.



a. Use the histogram to make a frequency table for the hourly pay values.

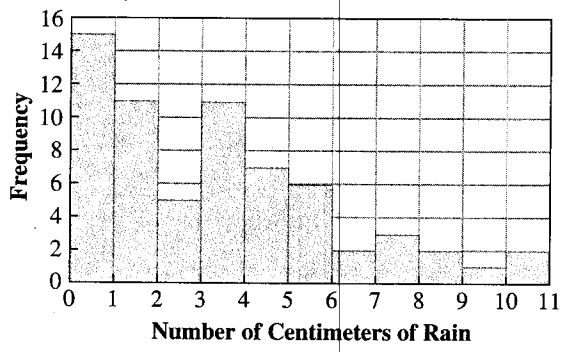
b. How many students worked during the summer?

c. What were the most frequent hourly pay range and the least frequent?

d. What percentage of students earned at least \$9.00 per hour?

e. What percentage of students earned between \$8.00 and \$10.00?

10. The following histogram shows the number of centimeters of rain during roughly a 2-month period for a village near the equator in the rain forest.



a. Use the histogram to make a frequency table for the number of centimeters of rain.

b. For how many days was rain measured?

c. On how many days were there at least 3 centimeters but less than 4 centimeters of rain?

d. What percentage of days had at least 7 centimeters of rain?

e. What percentage of days had at least 1, but less than 5 centimeters of rain?

11. Suppose students in a fifth-grade class were asked to record the number of hours of television they watch each week, with the following results: 0.5, 29.9, 25, 25.6, 24.3, 16.2, 28.5, 1, 27, 16.8, 17.8, 24.5, 24, 25.5, 26.5, 26, 15, 9.5, 14.5, 16, and 16.5. In order to create a histogram for this data, you will first need to decide how big to make the measurement classes or bins.

a. If bins have a length of 3, how many bins will be needed? Create a frequency table using bins of length 3.

b. If bins have a length of 5, how many bins will be needed? Create a frequency table using bins of length 5.

c. If bins have a length of 10, how many bins will be needed? Create a frequency table using bins of length 10.

d. Consider the frequency tables from parts (a) through (c). Which one, in your opinion, is the best display of the data? Give a reason for your choice and create a histogram from the frequency table of your choice.

12. Suppose the ages (in years) of the students in a particular third-grade class are as follows: 8.00, 8.08, 8.10, 8.13, 8.15, 8.18, 8.20, 8.23, 8.26, 8.27, 8.28, 8.29, 8.31, 8.32, 8.36, 8.38, 8.45, 8.49, 8.49, 8.50, 8.53, 8.62, 8.74, 8.87, and 8.99. In order to create a histogram for this data, you will first need to decide how big to make the measurement classes or bins.

a. If bins have a length of 0.5, how many bins will be needed? Create a frequency table using bins of length 0.5.

b. If bins have a length of 0.25, how many bins will be needed? Create a frequency table using bins of length 0.25.

c. If bins have a length of 0.10, how many bins will be needed? Create a frequency table using bins of length 0.10.

d. Consider the frequency tables from parts (a) through (c). Which one, in your opinion, is the best display of the data? Give a reason for your choice and create a histogram from the frequency table of your choice.

Problems 13 and 14

A teacher gives an 80-point test to his class, with the following scores: 30, 32, 35, 40, 44, 47, 48, 50, 51, 52, 55, 56, 57, 60, 61, 62, 62, 63, 64, 65, 66, 67, 67, 70, 72, 72,

75, and 80. The teacher carries out a data analysis of all test scores, including a frequency table and histogram. He considers two options for the bins: group the data into bins of length 10 or into bins of length 8.

13. Suppose the teacher groups the data into bins of length 10.

a. Create a frequency table.

b. Create a histogram using bins of length 10.

c. Describe any trends you observe in the data.

d. Give a reason why this choice of bin length is or is not appropriate.

14. Suppose the teacher groups the data into bins of length 8.

a. Create a frequency table.

b. Create a histogram using bins of length 8.

c. Describe any trends you observe in the data.

d. Give a reason why this choice of bin length is or is not appropriate.

Problems 15 and 16

A record number of participants ran the New York Marathon on November 2, 2003. The 26.2-mile course leads runners through the five boroughs that make up New York City. The male and female runners in the 2003 marathon fell into the following age categories.

Age in Years	Men	Women
10–19	51	35
20–29	3014	2906
30–39	8483	4612
40–49	6954	2909
50–59	3469	1050
60–69	856	170
70–79	117	22
80–89	3	0
90–99	1	0

Source: www.ingnycmarathon.org/results/topfinishers.html.

15. **a.** Find the percentage of male runners who were between the ages of 30 and 39.

b. Find the percentage of male runners who were age 60 or older.

c. Construct a relative frequency histogram for the ages of the male runners in the 2003 New York Marathon.

d. Most of the values cluster between what two ages?

16. **a.** Find the percentage of female runners who were between the ages of 30 and 39.

b. Find the percentage of female runners who were age 60 or older.

c. Construct a relative frequency histogram for the ages of the female runners in the 2003 New York Marathon.

d. Most of the values cluster between what two ages?

Problems 17 and 18

What's the state of your state's health? Some states are healthier to live in than others. A 2002 study by The United Health Foundation used data from government agencies and health organizations to rate the states on 17 statistical measures of health. Included were things such as smoking, violent-crime rates, motor-vehicle death rates per 100,000 miles driven, incidence of major infectious diseases, life expectancy at birth, and access to health care. The higher scores represent "healthier" states. (*Note:* Scores presented in the tables indicate the percentage a state is above or below the national norm. For example, a state with a score of 20 is 20 percent above the national average.)

Rank	State	Score	Rank	State	Score
1	New Hampshire	23.9	26	Wyoming	2.7
2	Minnesota	21.8	27	Ohio	1.7
3	Massachusetts	18.5	28	Maryland	0.8
4	Utah	17.9	29	Michigan	0.6
5	Connecticut	16.6	30	Alaska	0.2
6	Vermont	15.8	31	Illinois	−0.9
7	Iowa	14.5	32	New York	−2.6
7	Colorado	14.5	32	Missouri	−2.6
9	North Dakota	14.0	34	Arizona	−3.7
10	Maine	13.8	35	Delaware	−3.9
11	Washington	13.5	36	North Carolina	−5.3
11	Wisconsin	13.5	37	Texas	−5.6
13	Rhode Island	11.8	38	Nevada	−5.8
14	Hawaii	11.6	39	Kentucky	−7.6
15	Nebraska	10.5	40	Georgia	−8.8
16	South Dakota	9.7	41	West Virginia	−8.9
17	Oregon	9.3	42	New Mexico	−10.1
18	Virginia	8.7	43	Florida	−12.2
19	New Jersey	8.6	44	Tennessee	−12.3
20	Idaho	7.8	45	Alabama	−12.8
21	Kansas	6.7	46	Oklahoma	−13.3
22	Indiana	4.3	47	Arkansas	−14.9
23	Pennsylvania	3.8	48	South Carolina	−16.4
24	Montana	3.7	49	Mississippi	−22.2
24	California	3.7	50	Louisiana	−23.9

Source: America's Health: United Health Foundation State Health Rankings 2002 Edition. © 2002 United Health Foundation.

17. a. Complete the following frequency table for the 50 states using the 2002 state health scores.

Interval	Frequency
−30.0 to −20.1	
−20.0 to −10.1	
−10.0 to −0.1	
0 to 9.9	
10 to 19.9	
20 to 29.9	

b. Construct a histogram.

c. How many states ranked at least 10 percent above the national average?

18. a. Complete the following frequency table for the 50 states using the 2002 state health scores.

Interval	Frequency
−30.0 to −25.1	
−25.0 to −20.1	
−20.0 to −15.1	
−15.0 to −10.1	
−10.0 to −5.1	
−5.0 to −0.1	
0 to 4.9	
5 to 9.9	
10 to 14.9	
15 to 19.9	
20 to 24.9	
25 to 29.9	

b. Construct a histogram.

c. What percentage of states ranked below the national average?

d. Compare the histogram from part (b) to the histogram created in the previous problem. Which size interval best reveals the pattern in the data? Explain.

Problems 19 and 20

Yellowstone National Park is home to over 500 geysers. Geysers are hot springs that erupt periodically. Old Faithful is a geyser that erupts more frequently than other geysers. However, it is not the largest or most regular geyser. Park rangers keep a log book of geyser activity at the Old Faithful Visitors Center. Consider the following Old Faithful eruption-duration data taken for a 3-day period in August 1998 and in August 2003.

August 1998 Eruption Durations (Minutes:Seconds)			August 2003 Eruption Durations (Minutes:Seconds)		
3:35	4:10	4:10	4:34	1:56	4:12
4:00	4:15	4:49	4:10	3:53	1:48
4:06	4:06	4:04	1:50	4:23	4:48
2:54	4:23	4:30	4:24	1:51	2:01
4:28	4:10	4:24	4:28	5:00	4:45
3:54	4:03	4:21	2:20	4:08	1:53
4:30	1:47	4:09	4:45	4:06	4:31
3:47	4:36	4:15	2:13	4:30	1:47
3:52	4:10	4:25	4:36	4:22	4:48

Source: www.geyserstudy.org/g_logs.htm.

19. Consider the eruption data from August 1998.

a. Create a histogram using measurement classes 1:00 to 1:59, 2:00 to 2:59, 3:00 to 3:59, and 4:00 to 4:59.

b. Create a histogram using measurement classes 1:30 to 1:59, 2:00 to 2:29, . . ., 4:30 to 4:59.

c. Compare the histograms created in parts (a) and (b). Which size interval best reveals the pattern in the data? Explain.

20. Consider the eruption data from August 2003.

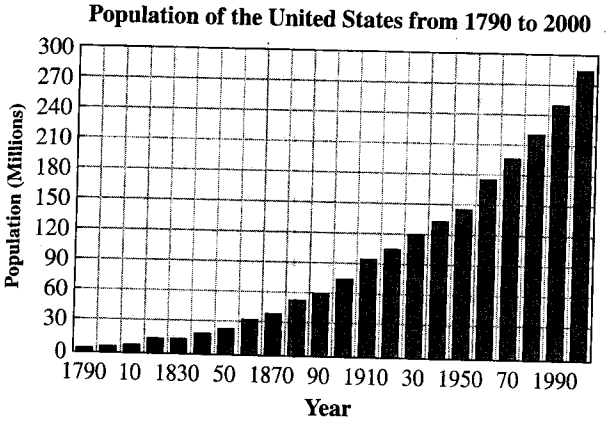
a. Create a histogram using measurement classes 1:00 to 1:59, 2:00 to 2:59, 3:00 to 3:59, 4:00 to 4:59, and 5:00 to 5:59.

b. Create a histogram using measurement classes 1:30 to 1:59, 2:00 to 2:29, . . ., 5:00 to 5:29.

c. Compare the histograms created in parts (a) and (b). Which size interval best reveals the pattern in the data? Explain.

d. Compare the "best" histogram from part (c) to the "best" histogram from part (c) in problem 19. Describe the similarities and differences between the two graphs.

21. The following bar graph shows how the population of the United States changed from 1790 to 2000.



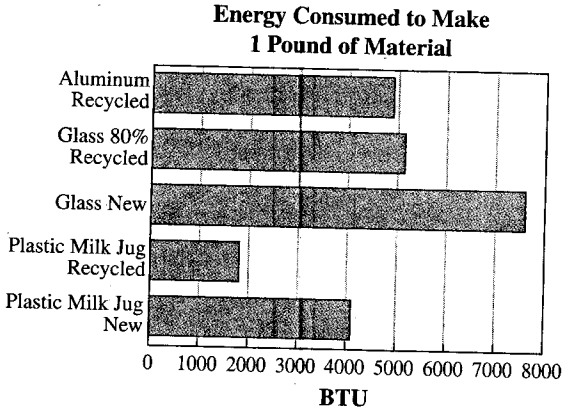
a. Estimate the population of the United States in 1790, 1890, 1990, and 2000.

b. What was the change in population from 1790 to 1890?

c. In what year were there approximately 30 million people in the United States?

d. Between what two decades did the United States experience the greatest increase in population?

22. The following bar graph shows the energy consumed (in BTUs) to make 1 pound of a material.



a. Use the bar graph to estimate the BTUs required to make 1 pound of each type of material.

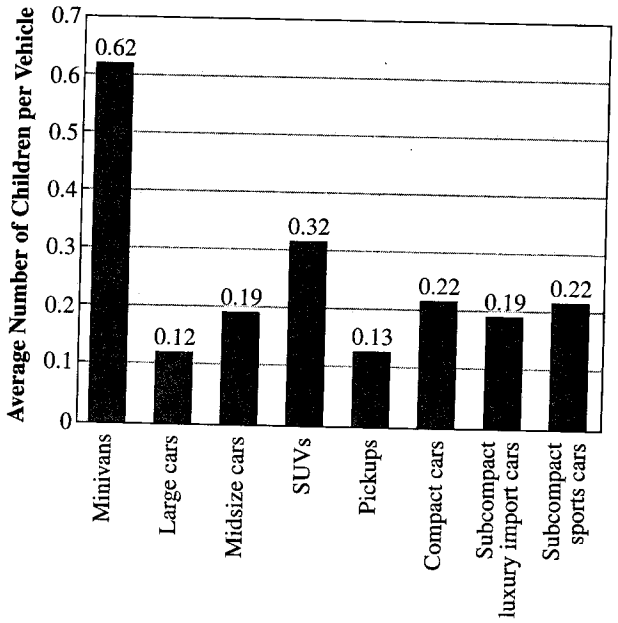
b. Approximately how many times more energy is used to make 1 pound of recycled aluminum than 1 pound of recycled plastic milk jug?

c. What percentage of energy is saved if 1 pound of glass that is 80% recycled is made rather than 1 pound of new glass?

d. What conclusions can you make about the energy required to make 1 pound of the various materials?

23. The data used in the following graph are from the Fatality Analysis Reporting System (FARS) for the years 1995 to 2001. The FARS was created in 1975 to assist in identifying traffic safety problems and motor-vehicle safety problems. The bar graph shows the average number of children per vehicle in a collision or rollover involving a fatality for eight types of 1995- to 2001-model vehicles. For a multivehicle crash involving a fatality, FARS included all the children in all the vehicles in the crash regardless of whether a fatality occurred in a vehicle containing a child. The total number of child fatalities in 1995- to 2001-model vehicles was 365 for minivans, 131 for large cars, 474 for midsize cars, 609 for SUVs, 391 for pickups, 292 for compact cars, 282 for subcompact luxury import cars, and 494 for subcompact sports cars.

Average Number of Children per Vehicle Type in an Accident Involving a Fatality, 1995 to 2001



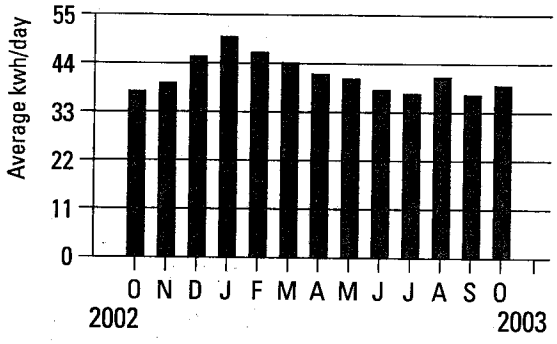
a. Which two types of vehicles carried the most children on average?

b. Which two types of vehicles were involved in the greatest total number of child fatalities?

c. Compare the average number of children in minivans to the average number of children in SUVs and subcompact sports cars. Minivans carry how many times more children, on average, than the other two vehicle types? For which type of vehicle would you expect there to be the greatest number of child fatalities?

d. Compare child fatalities for minivans to child fatalities for SUVs and subcompact sports cars. Compare your result to part (c). What do you notice?

24. PacifiCorp provides electricity to more than 1.5 million customers in the United States. In Oregon, Washington, Wyoming, and California, PacifiCorp operates as Pacific Power. The monthly bills Pacific Power sends its customers include a graph called the "Energy Usage Comparison Chart." One such chart follows. The chart gives the average kilowatt-hours used per day (kwh/day) for each month. A kilowatt-hour is a measure of electrical energy, and one kilowatt-hour is approximately the amount of energy used by one 100-watt light bulb for 10 hours.



a. What was the average kilowatt-hours per day used by this homeowner during the month of December 2002?

b. During what month and year did the homeowner use an average of 44 kwh/day?

c. If there were 29 days in the billing cycle for October 2003, approximately how many kilowatt-hours were used during the month of October?

d. In which month was the homeowner's average kilowatt-hours per day the greatest?

e. The average number of kilowatt-hours used per day fell from January 2003 through July 2003, and then suddenly increased. What might be one possible cause of the increase?

25. In 2002, the five most populous nations in the world were as follows:

China	1281 million
India	1050 million
United States	287 million
Indonesia	217 million
Brazil	174 million

Source: Population Reference Bureau.

a. Draw a bar graph to represent this information.

b. Could you create a pie chart with the population totals? Explain.

26. The Arizona State Legislature established the Homeless Youth Intervention Program, which began on January 1, 2000. In June 2003, the homeless youth population in Arizona had the following composition. (*Source:* www.azchildren.org.)

45% Anglo

33% Hispanic

12% African American

4% Native American

a. Draw a bar graph to represent this information.

b. Could you create a pie chart with the youth population percentages? Explain.

27. According to the Food Marketing Institute, the average weekly grocery cost per person in the United States is as follows. Make a bar graph to illustrate this relationship.

Size of Household	Average Weekly Cost per Person (in Dollars)
1 person	52
2 people	40
3 or 4 people	30
5 or more people	24

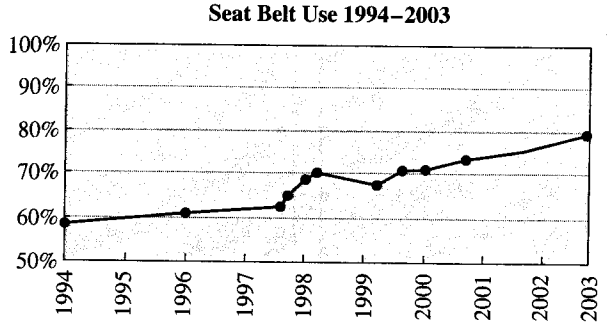
Source: Food Marketing Institute.

28. The Centers for Disease Control and Prevention reported the use of cigarettes in the year 2002 as follows. Make a bar graph to illustrate this relationship.

Age	Cigarette Use (Percent of Population)
12–13 years	3.2
14–15 years	11.2
16–17 years	24.9
18–25 years	40.8
26–34 years	32.7
35 years and over	23.4

Source: www.cdc.gov.

29. According to the National Center for Statistics and Analysis (NCSA), seat-belt use nationwide has generally increased from 1994 to 2003, as shown in the following line graph.



a. Estimate the percentage of people nationwide who used their seat belt during each of the years from 1994 to 2003.

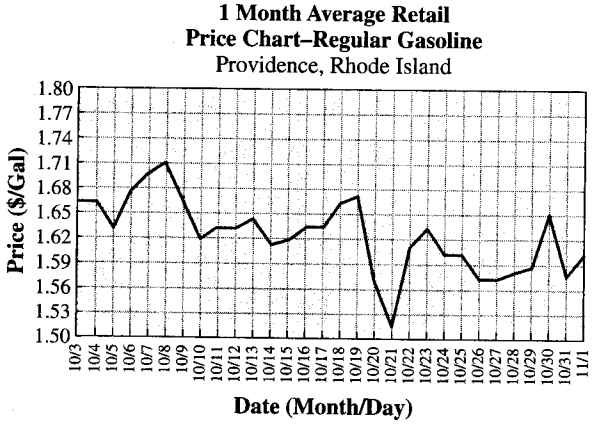
b. During which year did the largest increase in seat-belt use occur?

c. During which year(s) did the use of seat-belts decrease?

d. Find the change in the percentage of seat-belt use from 1994 to 2003.

e. In which year(s) did approximately 75% of nationwide use their seat belts?

30. Regular gasoline prices fluctuate daily. The following line graph shows the price in dollars per gallon for regular gasoline in Providence, Rhode Island, from October 3 to November 1, 2003.



a. What was the greatest price charged for regular gasoline, and on what day did it occur?

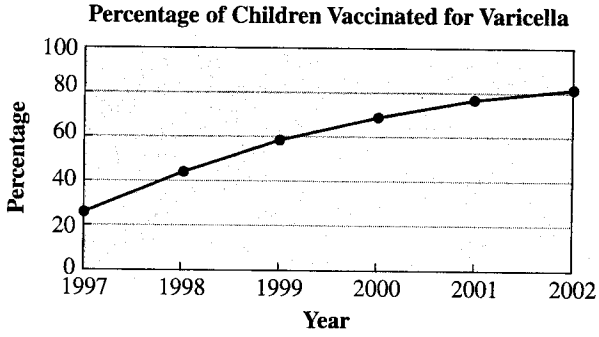
b. What was the smallest price charged for regular gasoline, and on what day did it occur?

c. Between which 2 days did the greatest increase in price occur?

d. Between which 2 days did the greatest decrease in price occur?

e. On which day(s) did regular gasoline cost $1.66 per gallon?

31. The varicella vaccine was licensed on March 17, 1995, by the U.S. Food and Drug Administration. Varicella is a common, contagious illness; it is also known as chickenpox. Before the vaccine was available, the 4 million annual cases of varicella caused 10,000 hospitalizations and 100 deaths. The following line graph shows the percentage of children aged 19 to 35 months in the United States who were vaccinated with the varicella vaccine from 1997 to 2002.

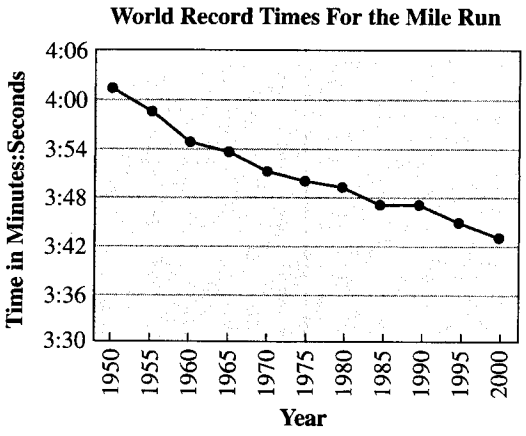


a. Estimate the percentage of children aged 19 to 35 months who were vaccinated with the varicella vaccine each year.

b. Estimate the increase in the percentage of children vaccinated with the varicella vaccine for each pair of consecutive years. What do you notice?

c. Describe the trend in the percentage of children immunized. Discuss whether this trend can continue indefinitely.

32. The following line graph displays the world-record times for the mile run over a 49-year period. The current world-record holder is Hicham El Guerrouj of Morocco. His time of 3 minutes and 43.13 seconds was set in 1999 and is unbroken today.



a. Estimate the world record time for the mile run in 1950, 1970, and 1990.

b. The world-record time for the mile run was reduced by how many seconds in the decade from 1950 to 1960? After 1960, how many more years did it take to achieve the same time reduction as in the decade from 1950 to 1960?

c. Describe the trend in the world-record times for the mile run. Discuss whether this trend can continue indefinitely.

33. The approximate number of immigrants admitted to the United States in each of the years from 1990 to 2002 is given in the following table.

Year	Number of Immigrants
1990	1,536,000
1991	1,827,000
1992	974,000
1993	904,000
1994	804,000
1995	720,000
1996	916,000
1997	798,000
1998	654,000
1999	647,000
2000	850,000
2001	1,064,000
2002	1,064,000

Source: U.S. Immigration and Naturalization Service, Statistical Yearbook.

a. Make a line graph to represent these data.

b. Identify the year in which the largest increase in admitted immigrants occurred, compared to the previous year, and the year in which the largest decrease occurred.

c. Identify the time period(s) that experienced significant changes in the number of immigrants admitted to the United States.

d. Discuss the trend in the data.

34. The per-capita personal income in the United States from 1991 to 2001, according to the U.S. Department of Commerce, is given in the following table.

Year	Per-Capita Personal Income
1991	20,023
1992	20,960
1993	21,539
1994	22,340
1995	23,255
1996	24,270
1997	25,412
1998	26,893
1999	27,880
2000	29,760
2001	30,413

Source: www.bea.doc.gov.

a. Make a line graph to represent these data.

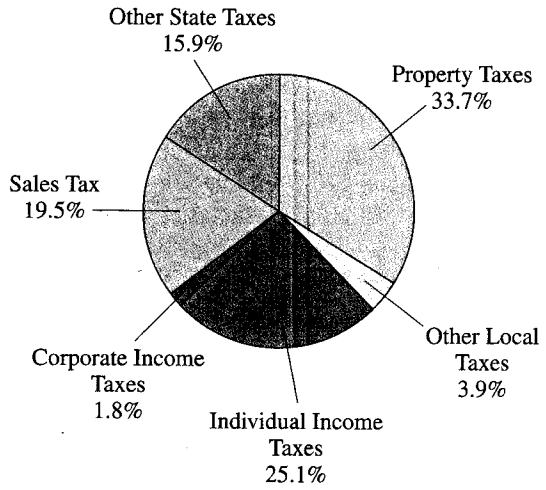
b. Identify the year in which the smallest increase in per-capita income occurred.

c. During what time period did per-capita personal income grow the fastest?

d. Discuss the trend in the data.

35. According to the Office of Fiscal and Program Review in the Maine State Legislature, 2002 state and local taxes in Maine came from six sources, as indicated in the following pie chart.

Maine State and Local Tax Mix



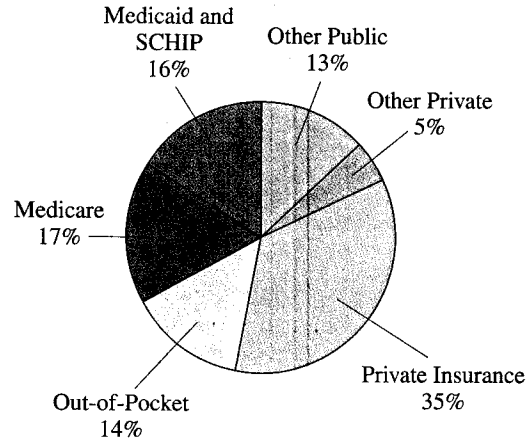
a. If the state and local taxes in 2002 totaled $4,277,900,000, find the dollar amount of taxes that came from each source.

b. According to the pie chart, what are the two main sources of state and local taxes? What source contributed the least to state and local taxes?

c. Find the degree measure of each sector in the pie chart. Round to the nearest tenth of a degree.

36. The Centers for Medicare and Medicaid Services have determined that the nation's health-care dollars in 2001 came from six sources, as shown in the following pie chart. (SCHIP is State Children's Health Insurance Program.)

Sources for the Nation's Health-Care Dollars



a. Use the pie chart to determine how many health-care dollars out of every million came from each of the six sources.

b. According to the pie chart, what is the main source of health-care funding?

c. Find the degree measure of each sector in the pie chart. Round to the nearest tenth of a degree.

37. According to the Progressive Grocer, April 2003, consumers in grocery stores spend $100 in the following way.

HOW $100 IS SPENT	
Perishables	$50.42
Beverages	$10.71
Misc. grocery	$ 5.34
Nonfood grocery	$ 9.03
Snack foods	$ 6.25
Main meal items	$ 8.25
Health & beauty care	$ 3.96
General merchandise	$ 3.39
Pharmacy/unclassified	$ 2.65

Source: www.progressivegrocer.com

a. Find the percentage of $100 spent in each category.

b. If each category will be represented by a sector in a circle, find the degree measure of each sector. Round to the nearest tenth of a degree.

c. Construct a pie chart to illustrate how the typical U.S. consumer spends $100 at the grocery store.

38. The following table shows the main causes of death for people aged 25 to 40 in the United States in 2001.

Causes	Number of Deaths
Unintentional injuries	27,784
Malignant neoplasms	20,563
Diseases of the heart	16,486
Suicide	11,705
Homicide	9472
Human immunodeficiency virus (HIV)	7968
Liver diseases	3723
Diabetes mellitus	2553
Influenza and pneumonia	1322
Other	31,781
Total	133,357

Source: Centers for Disease Control and Prevention, Atlanta, GA.

a. Find the percentage of deaths attributed to each cause.

b. If each category will be represented by a sector in a circle, find the degree measure of each sector. Round to the nearest tenth of a degree.

c. Construct a pie chart to illustrate the main causes of death graphically.

39. People are fired for many reasons. According to one source, the reasons can be broken down as shown in the following table. Create a pie chart to depict this data.

REASONS FOR BEING FIRED	
Incompetence	39%
Inability to get along with others	17%
Dishonesty or lying	12%
Negative attitude	10%
Lack of motivation	7%
Failure to follow instructions	7%
Other reasons	8%

Source: Robert Half International, Menlo Park, CA.

40. The American Dietetic Association reported that children aged 8 to 12 years old said their top role models were as follows: mother (23%), father (17.4%), unsure or no role model (13.2%), and sports celebrity (8.3%). Create a pie chart to represent these data. Notice that the percentages provided do not add to 100%. Create a fifth category labeled "other" for the remaining percentage.
Source: www.eatright.org/Public/index.cfm.

41. In the year 2000, there were approximately 221.1 million people aged 15 and older in the United States. Of those, 120.2 million were married; 41 million were widowed, divorced, or separated; and 59.9 million had never been married. (*Source:* www.census.gov.) For each of the following, if the graph would be an appropriate display for these data, create the graph. If not, explain why.

a. histogram

b. bar graph

c. line graph

d. pie chart

42. In a July 2002 report by the U.S. Department of Labor, the average hourly earnings for selected occupations were as follows: engineers ($32.73), physicians ($51.66), dentists ($35.51), elementary-school teachers ($30.75), secretaries ($14.77), food-service workers ($7.41), and registered nurses ($24.57). (*Source:* www.bls.gov.) For each of the following, if the graph would be an appropriate display for these data, create the graph. If not, explain why.

a. histogram

b. bar graph

c. line graph

d. pie chart

43. The following table gives the number of live births per 1000 people in the United States from January 2002 through December 2002. For each of the following, if the graph would be an appropriate display for these data, create the graph. If not, explain why.

Month	Births per 1000 People
January	13.6
February	13.8
March	13.6
April	13.8
May	13.9
June	13.9
July	14.6
August	14.7
September	14.7
October	14.1
November	13.4
December	13.4

Source: National Vital Statistics Reports, Vol. 52, No. 5, September 29, 2003.

a. histogram **b.** bar graph
c. line graph **d.** pie chart

44. The Bureau of Labor Statistics sponsors an annual demographic survey. The numbers of families earning less than $25,000 in 2001 are listed in the following table.

Total 2001 Income	Number of Families
Under $2500	1,268,000
$2500–$4999	637,000
$5000–$7499	893,000
$7500–$9999	1,176,000
$10,000–$12,499	1,561,000
$12,500–$14,999	1,629,000
$15,000–$17,499	2,133,000
$17,500–$19,999	1,913,000
$20,000–$22,499	2,365,000
$22,500–$24,999	1,992,000

Source: U.S. Census Bureau.

For each of the following, if the graph would be an appropriate display for these data, create the graph. If not, explain why.

a. histogram **b.** bar graph
c. line graph **d.** pie chart

45. Each year, the U.S. Department of Justice reports the total number of victimizations per 1000 people aged 12 and over. The following table contains the totals for the years from 1991 through 2001. Decide what type of graph would convey this information in a visual way most effectively, and create the graph. Write a paragraph that summarizes the information and describes important features of the graph.

Year	Number of Victimizations per 1000 people Aged 12 and over
1991	48.4
1992	47.9
1993	49.1
1994	51.2
1995	46.1
1996	41.6
1997	38.8
1998	36.0
1999	32.1
2000	27.4
2001	24.7

Source: U.S. Department of Justice.

46. The Center on Hunger and Poverty reported that as of August 2002, the 10 states with the largest hunger rates were those shown in the following table. The hunger rate is defined as the percentage of households that were in a "state of hunger," that often felt pain because of lack of food. Decide what type of graph would convey this information most effectively in a visual way, and create the graph. Write a paragraph that summarizes the information and describes important features of the graph.

State	Hunger Rate
Oregon	6.2
Washington	5
New Mexico	4.6
Utah	4.4
Texas	4.4
Idaho	4.3
Alaska	4.3
Florida	4
Oklahoma	3.9
Tennessee	3.9

Source: Salem Statesman Journal, August 16, 2002.

Extended Problems

47. Check your local newspaper and cut out an example of each of the following types of graphs studied in this section: histogram, bar graph, line graph, and pie chart. For each graph, explain what the graph illustrates. Is the graph easy to read? Does it help clarify the information in the article? Would the graph have been more informative if it had been presented in another form? For each graph, present the information using a different type of graph. Is the new graph a better representation of the information? Explain.

48. The U.S. Census Bureau is the leading source of data and information about the people of the United States and the economy. The most recent national census was conducted in the year 2000. Go to the U.S. Census Bureau website at www.census.gov, and select a state you would like to study. Browse that state's tables of data such as demographic, social, economic, housing, or employment characteristics. Select four different sets of data, and create graphs that best display those data sets. Be sure your graphs are labeled. In addition to the graphs you construct, write a short report about the state you selected and refer to the information provided by the graphs.

49. Research the history of graphs. William Playfair has been credited with inventing many of the types of graphs we use today. Find and print several of the innovative graphs William Playfair invented. In each case, describe the points that Playfair tried to make and explain how he used the graph to make those points. Use search keywords "William Playfair graphs" on the Internet. Write a short report on your findings, and include several graphs as examples.

50. All of the graphs that were created in this section could have been created using a computer program such as Excel, a graphing calculator, or graphing applets on the Internet. Search the Bureau of Labor and Statistics at http://stats.bls.gov/ to find a topic of interest to you. Create two different graphs using a computer or graphing calculator. Describe any trends or clusters that are apparent in the graphs.

51. Research the history of the federal debt. The actual dollar amounts for the federal debt from the years 1791 through today can be found at www.publicdebt.treas.gov. Record federal debt totals beginning in 1791 for every 10-year period through 2001. Make a line graph for the federal debt totals. Discuss any trends in the graph. What national or world events contributed to the greatest increases in the federal debt?

52. Vilfredo Pareto was an Italian economist who studied the distribution of wealth more than 100 years ago. He concluded that about 20% of the people controlled about 80% of society's wealth. This idea has come to be known as the **Pareto effect**. In more general terms, the Pareto effect is that a small portion of causes produce a large portion of effects. A Pareto chart is a modified bar graph with the bars arranged so that the tallest bars are on the left, and bars are arranged in decreasing order. In this way, data are arranged so that the vital, most influential factors are listed first in a prioritized order.

a. The 2004 NASA budget is $15,469 million and is divided in the following way. Use the budget information to construct a Pareto chart and describe how the Pareto effect applies to the NASA budget. If you were asked to trim the budget, where would you begin looking for cuts?

Category	Budgeted Amount (in Millions of Dollars)
Inspector general	26
Education	170
Aeronautics	959
Space flight	7782
Space science	4007
Earth science	1552
Biological and physical research	973

Source: www.nasa.gov.

b. The Lewiston Police Department in Lewiston, Idaho, lists annual crime statistics. The following annual major crimes were listed for 2002. Use the crime statistics to create a Pareto chart and describe how the Pareto effect applies to the crime in Lewiston. If you were going to spend grant money to try to reduce crime, in which areas would you spend the money?

Category	Major Crimes 2002
Homicide	1
Rape	6
Robbery	7
Assault	431
Burglary	107
Larceny	1072
Vehicle theft	61

Source: www.cityoflewiston.org.

8.2 Comparisons

INITIAL PROBLEM

You are the manager of a small refreshment stand near the beach that sells hot chocolate, ice cream, and hot dogs. You have to present monthly sales figures to the owner showing how the shop has done over the past year. The following table contains a record of the monthly sales, in dollars, of each item.

	Hot Chocolate	Ice Cream	Hot Dogs
Oct	400	330	220
Nov	470	240	200
Dec	630	200	270
Jan	600	110	190
Feb	670	90	180
Mar	570	120	210
Apr	490	220	250
May	280	370	270
Jun	130	460	310
Jul	70	620	330
Aug	80	660	340
Sep	240	450	260

How could you present these data to show clearly the sales trends of each item and to compare sales of the three items?

A solution of this Initial Problem is on page 511.

Graphs and charts can help us understand the patterns and relationships within a set of data. They can also reveal the nature of changes in quantities over a period of time. In addition, we can use charts and graphs to make comparisons between different, but related, sets of data. An effective visual presentation can not only show similarities or differences between sets of data, but can also help to explain why they exist. In the initial problem posed above, we want to provide a snapshot of the refreshment stand's sales and visualize the trends of the business. We also want to understand the relationships, if any, among the items sold. The visual representation we create may reveal trends and relationships that could be critical factors when decisions must be made about the business.

DOUBLE STEM-AND-LEAF PLOTS

In the previous section, we saw how to create a stem-and-leaf plot for a set of data. Stem-and-leaf plots and histograms may also be used to compare two different data sets. In Section 8.1, we created a stem-and-leaf plot for a set of scores on an economics test. Now suppose that two different sections of an economics course took the same test, and the scores for the first class were those used in the example of Section 8.1. Suppose that the scores of the two classes were (in order):

Economics 101 Test Scores—Class 1

26, 32, 54, 62, 67, 70, 71, 71, 74, 76, 80, 81, 84, 87, 87, 87, 89, 93, 95, 96

Economics 101 Test Scores—Class 2

34, 45, 52, 57, 63, 65, 68, 70, 71, 72, 74, 76, 76, 78, 83, 85, 85, 87, 92, 99

Notice that both classes had the same number of students. Which class did better on the test? The answer is not obvious from looking at the raw data.

In Section 8.1, we created a stem-and-leaf plot to represent the test scores for one class. These two data sets may be combined into one plot called a **double stem-and-leaf plot** (Table 8.9).

Table 8.9

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

Note that the stem is placed in the middle and the two sets of leaves are placed on either side of the stem, like branches on a tree trunk. Because more leaves are near the top left side of the stem than on the right, it appears that class 1 did somewhat better on the economics test than class 2.

COMPARISON HISTOGRAMS

The test data for the two economics classes may be used to create a **comparison histogram**, which may help clarify which section performed better on the test (Figure 8.16).

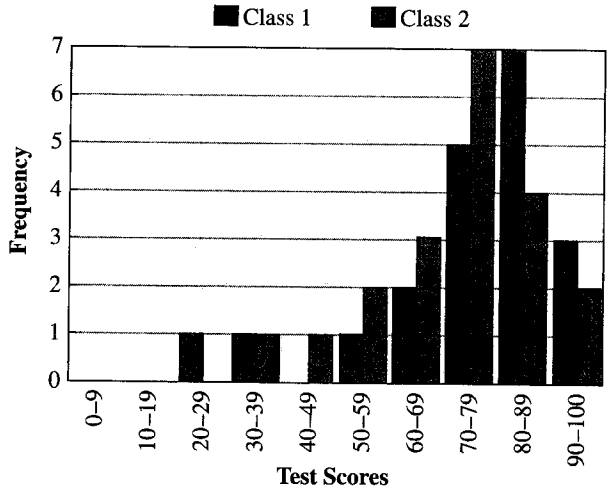


Figure 8.16

In Figure 8.16, we see that the histogram for class 1 peaks in the 80–89 bin, while class 2 has a peak in the 70–79 bin. So the largest group of students in class 1 had scores in the 80s, while the largest group of students in class 2 had scores in the 70s. We can

also see that more students in class 1 than in class 2 scored in the 90s, while fewer students in class 1 scored in the 50s and 60s. Later we will develop quantitative methods to compare these classes.

MULTIPLE-BAR GRAPHS

In the previous section, we learned that histograms are a special type of bar graph. A **multiple-bar graph**, also called a **comparison bar graph**, which involves two or more bars per category, may also be used to show relative strengths, as the next example illustrates.

EXAMPLE 8.7 Table 8.10 shows the number of male and female doctors in various specialties. Construct a comparison bar graph for the data given in the table.

Table 8.10

U.S. DOCTORS BY MEDICAL SPECIALTY AND GENDER, 2000		
	Female	Male
Family Practice	20,401	51,234
General Practice	2338	12,875
Internal Medicine	37,073	97,466
Pediatrics	30,322	32,063
Obstetrics/Gynecology	14,124	26,117

Source: Journal of the American Medical Association, as printed in the *Salem Statesman Journal*, August 22, 2003.

SOLUTION A comparison bar graph of the data is shown in Figure 8.17.

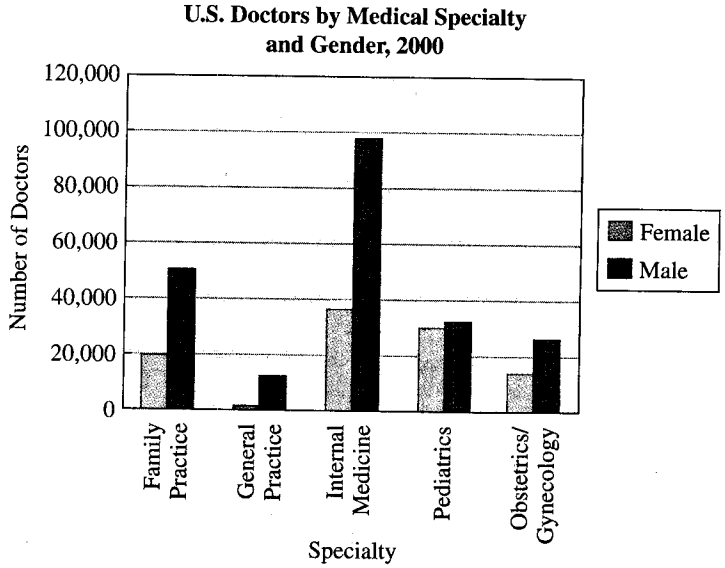


Figure 8.17

A number of observations can be made based on the comparison bar graph we created. For example, the graph indicates that the largest group of both male and female doctors practices internal medicine, although there are far more male doctors in this category than female doctors. The smallest number of doctors of both genders is in general practice. The numbers of male and female doctors are most nearly equal in the area of pediatrics. The comparison bar graph we have created is also called a **double-bar graph**.

Because they allow for quick comparison and analysis of data, double-bar graphs appear frequently in news media and reports. The graph in Figure 8.18 is a double-bar graph similar to one that appeared in a newspaper article about children's use of the Internet.

The Kids Are online
Children are leading the way to the Web. About 59 percent of youths ages 5 to 17 use the Internet—a higher rate than that of adults.

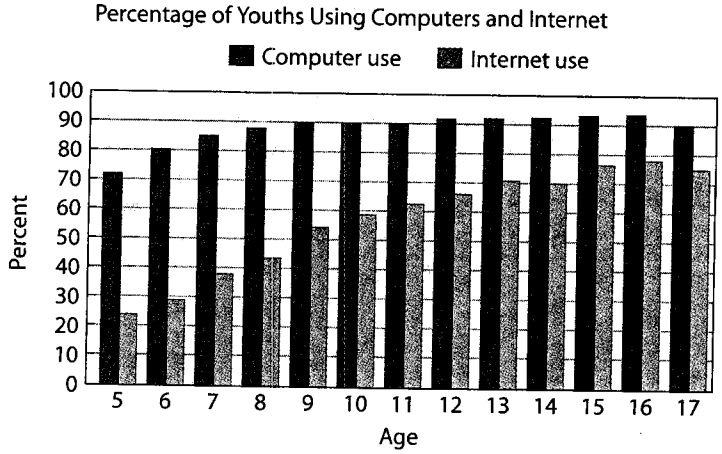


Figure 8.18

The double-bar graph indicates that the majority of kids in all age groups have access to computers. Older children use the Internet more often than younger children, but at least 50% of all children aged 9 or older use the Internet and nearly 80% of 16-year-olds use it.

MULTIPLE-LINE GRAPHS

In Section 8.1, we saw that line graphs can show trends over time. Like a multiple-bar graph, a **multiple-line graph**, also called a **comparison line graph**, which involves two

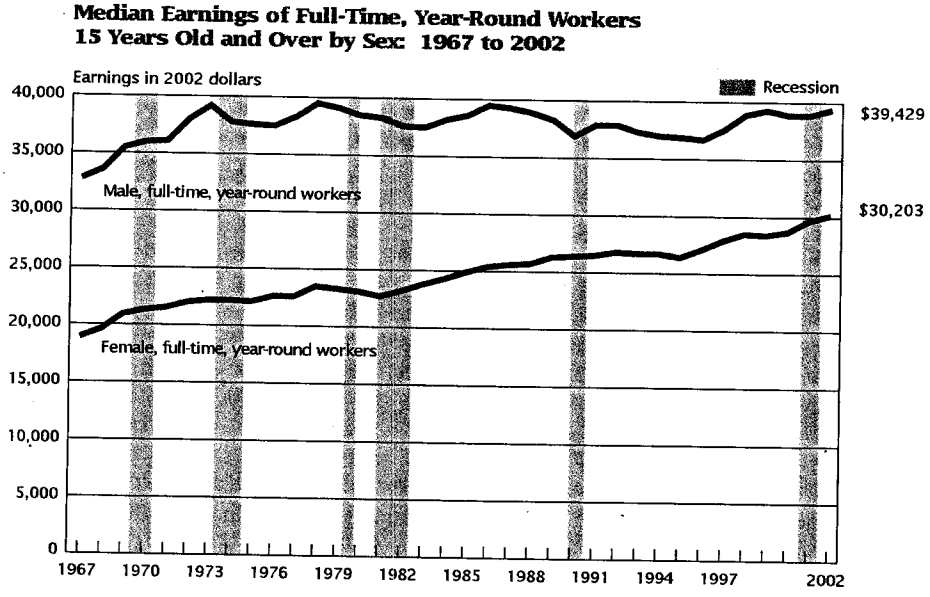


Figure 8.19

Source: www.census.gov/prod/2003pubs/p60-221.pdf.

or more lines, can be used to show relative strengths. For example, the **double-line graph** in Figure 8.19 compares median incomes for male and female full-time workers in the United States for the period 1967–2002.

The line graph makes it clear that while earnings have fluctuated over the years, the gap between men's and women's earnings has decreased from a difference of about $14,000 in 1967 to a difference of about $9000 in 2002.

EXAMPLE 8.8 Refer to the economics test scores pictured in the double stem-and-leaf plot of Table 8.9. Compare the two economics class test scores by using a double-line graph.

SOLUTION To create the double-line graph, we place the test intervals on the horizontal axis and the frequency of each interval on the vertical axis. Each dot on the graph represents the number of test scores in that interval. The double-line graph in Figure 8.20 shows that the two classes are clearly separated. The graphs of the classes appear similar, but the graph for class 1 is roughly 10 points higher than the graph for class 2.

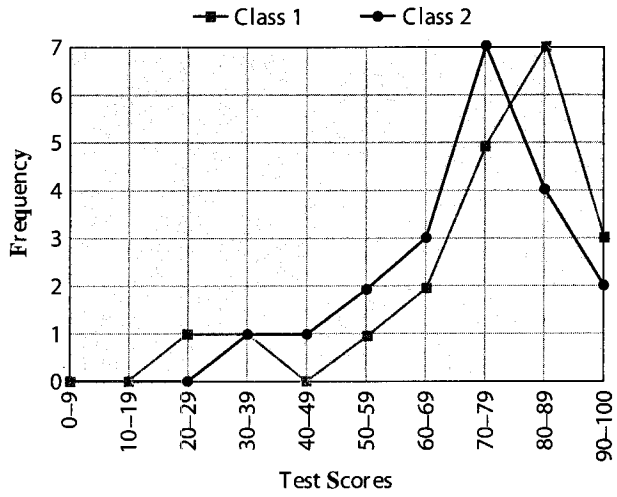


Figure 8.20

MULTIPLE PIE CHARTS

Recall from Section 8.1 that pie charts can show portions of a whole. A **multiple pie chart**, also called a **comparison pie chart**, consisting of a series of two or more pie charts, can show how the composition of a whole changes over time. Table 8.11 gives the percentage of U.S. residents in various age groups for the years 1900, 1950, and 2000.

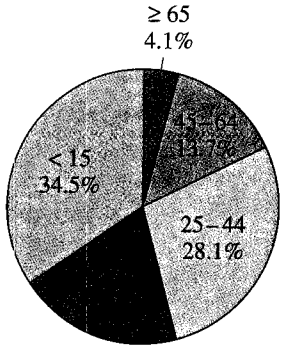
Table 8.11

PERCENTAGE OF THE TOTAL U.S. POPULATION BY AGE FOR THE YEARS 1900, 1950, AND 2000			
	1900	1950	2000
65 and over	4.1	8.1	12.4
45–64	13.7	20.3	22.0
25–44	28.1	30.0	30.2
15–24	19.6	14.7	13.9
15 or younger	34.5	26.9	21.4

Source: www.census.gov/prod/2002pubs/censr-4.pdf.

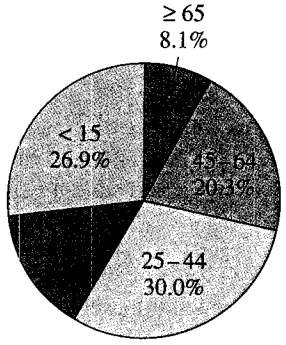
We can illustrate graphically how the makeup of the U.S. population has changed over 100 years by constructing multiple pie charts. Figures 8.21 (a) to (c) show three pie charts, one for each of the years 1900, 1950, and 2000.

Percentage of Total U.S. Population by Age in 1900



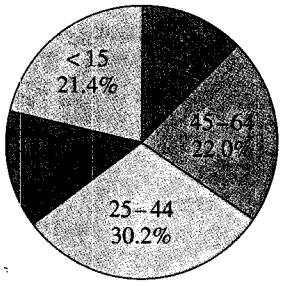
(a)

Percentage of Total U.S. Population by Age in 1950



(b)

Percentage of Total U.S. Population by Age in 2000



(c)

Figure 8.21

Notice how this series of pie charts illustrates a change in the demographics of the U.S. population. Over the past 100 years, the percentage of citizens aged 65 and older has increased significantly, while the percentage of those under the age of 15 has decreased dramatically, and the percentage aged 45–64 has remained relatively constant. These changes are more readily apparent in the pie charts than in the table of percentages.

PROPORTIONAL BAR GRAPHS

Bar graphs, especially proportional bar graphs, show relative amounts and trends simultaneously. In a **proportional bar graph**, all the bars are the same height, and each bar corresponds to 100% of a whole. Each bar is divided into pieces whose lengths correspond to the appropriate portions of the whole. Figure 8.22 shows a proportional bar graph that illustrates how the U.S. population has been distributed among four geographic regions of the country (West, South, Midwest, and Northeast) over a period of 100 years.

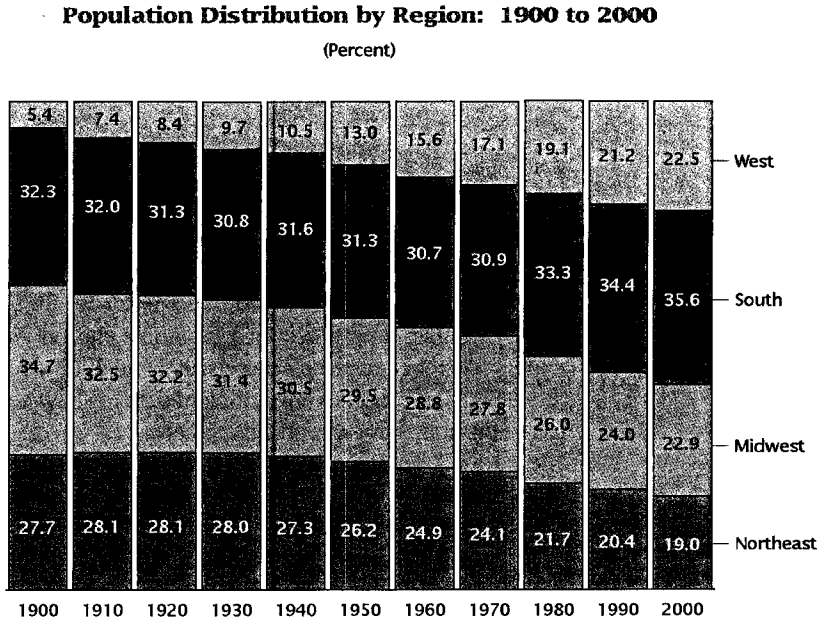


Figure 8.22

Source: www.census.gov/prod/2002pubs/censr-4.pdf.

Figure 8.22 shows that the percentages of the population residing in the Midwest and Northeast have decreased over the past 100 years, while the percentages of the population living in the West and the South have increased.

CHOOSING AN APPROPRIATE GRAPH

As we said earlier, software that can make striking, colorful graphs of any data is widely available. Using technology minimizes the work involved in creating graphs and leaves the creator free to focus on other issues, such as what type of graph is most appropriate for a particular purpose. The choice of comparison graph depends on the type of data and on the features of the data that you want to emphasize.

EXAMPLE 8.9 In March 2002 the New York–based Conference Board conducted a nationwide mail survey of 5000 people to determine their level of job satisfaction. They conducted a similar survey in 1995. Table 8.12 shows the areas addressed in the surveys and some results of the surveys. While the numbers in the table reveal that people were apparently less satisfied with their jobs in 2002 than in 1995, what type of graph might make the comparison between these 2 years more striking?

Table 8.12

QUESTION: TO WHAT EXTENT ARE YOU SATISFIED WITH EACH OF THE FOLLOWING ASPECTS OF YOUR PRESENT JOB?		
	1995	2002
Job Security	49%	50%
Wages	39%	37%
Vacation	56%	51%
Sick Leave	51%	47%
Health Plan	44%	40%
Interest in Work	65%	58%
People at Work	64%	58%

Source: National Family Opinion (NFO) World Group; The Conference (Associated Press).

SOLUTION Because the survey results have percentages for job components, a double-bar graph is an appropriate choice. We will let each bar represent one aspect of the job, and place bars for 1995 and 2002 side by side. Because only 2 years' results are given, there are not sufficient data to show a trend over a long period of time, so a line graph is not a good choice here.

A double-bar graph of these data is shown in Figure 8.23. From this visual representation of the data, we can see immediately that worker satisfaction decreased in nearly every category over the 7-year period. Workers were slightly more satisfied with their job security in 2002 than in 1995 and were only slightly less happy with their wages in 2002. However, in 2002, workers expressed much less satisfaction in the areas of "interest in work" and "people at work."

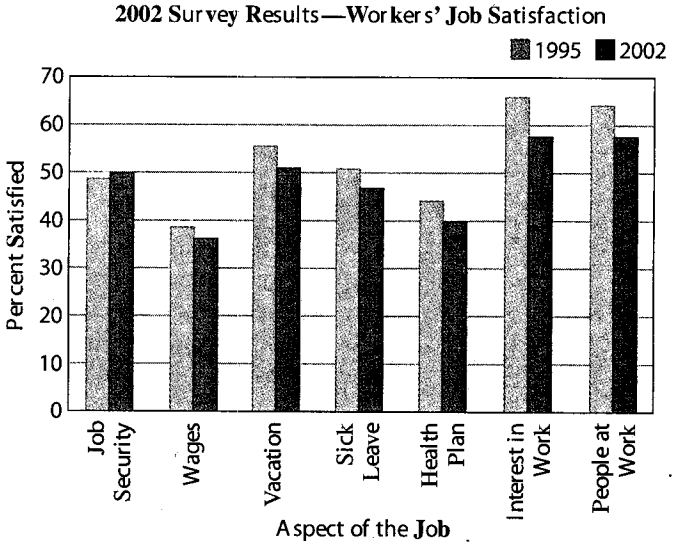


Figure 8.23

SOLUTION OF THE INITIAL PROBLEM

You are the manager of a small refreshment stand near the beach that sells hot chocolate, ice cream, and hot dogs. You have to present monthly sales figures to the owner showing how the shop has done over the past year. The following table contains a record of the monthly sales, in dollars, of each item.

	Hot Chocolate	Ice Cream	Hot Dogs
Oct	400	330	220
Nov	470	240	200
Dec	630	200	270
Jan	600	110	190
Feb	670	90	180
Mar	570	120	210
Apr	490	220	250
May	280	370	270
Jun	130	460	310
Jul	70	620	330
Aug	80	660	340
Sep	240	450	260

How could you present these data to show clearly the sales trends of each item and to compare sales of the three items?

SOLUTION The table lists the sales figures for the three products for the past year. We display these data using a multiple line graph in order to show trends in sales over the year (Figure 8.24).

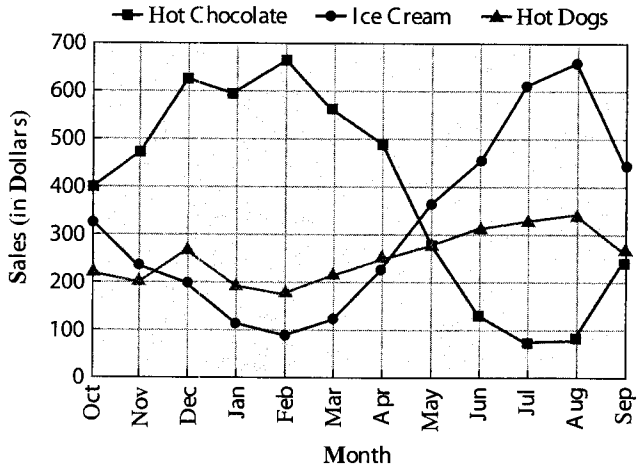


Figure 8.24

The graph suggests various interpretations. Hot chocolate seems to sell better during the colder months, while sales of ice cream increase during the warmer months. Sales of hot dogs do not seem to show as much variation.

PROBLEM SET 8.2

1. An instructor teaches two sections of chemistry. The results of an exam given to both classes are displayed in the following double stem-and-leaf plot.

Morning Class		Afternoon Class
	10	0 0
6 3	9	0 1 5 8
8 7 1	8	2 3 3 5 9 9
9 9 9 7 5 4 2 2	7	0 1 4 4 7 9
8 6 5 5 0	6	7 8
8 6 5	5	
9 7	4	
5 5	3	
	2	2
9	1	
	0	

a. How many students took the test in each class?

b. Give the high score and the low score for each class.

c. Compare the results for the two classes. What conclusions can you make?

2. Two health classes measured their body fat percentages to the nearest tenth. The results are given in the following double stem-and-leaf plot.

Health Class I		Health Class II
	25	7 8 9 9
	24	4 6
	23	1 3 3
	22	
	21	6 7
8 4 3	20	0 2
8 7 6 1	19	0 8 8
8 5 0	18	2
7 6 4 3 1	17	3 9
6 4 2 2 0 0	16	4 5 7
9 9 6 2	15	1 6 8

a. How many students measured their body fat in each class?

b. Give the high percentage and the low percentage for each class.

c. Compare the results for the two classes. What conclusions can you make?

3. Suppose that two fifth-grade classes take a reading test, yielding the following scores. (Scores are given in year.month equivalent form. One example is a score of 5.3, which means that the student is reading at the fifth-year, third-month level, where "year" means year in school.)

Class 1: 5.3, 4.9, 5.2, 5.4, 5.6, 5.1, 5.8, 5.3, 4.9, 6.1, 6.2, 5.7, 5.4, 6.9, 4.3, 5.2, 5.6, 5.9, 5.3, 5.8

Class 2: 4.7, 5.0, 5.5, 4.1. 6.8, 5.0, 4.7, 5.6, 4.9, 6.3, 7.2, 3.6, 8.1, 5.4, 4.7, 4.4, 5.6, 3.7, 6.2, 7.5

a. Make a double stem-and-leaf plot for the test scores from the two classes.

b. Describe any similarities and differences in the reading test results for the two classes.

c. Do the data have any outliers or other striking features?

4. A professor scheduled two sociology classes together for a joint midterm. The scores for the two classes follow:

Class 1: 85, 73, 84, 76, 73, 92, 64, 86, 84, 95, 66, 87, 63, 74, 84, 92, 76, 80, 86, 77, 91, 74, 76, 85

Class 2: 66, 74, 86, 84, 54, 82, 70, 86, 94, 88, 96, 83, 73, 78, 75, 83, 80, 74, 77, 82, 85, 73, 85, 80, 84, 76, 88

a. Make a double stem-and-leaf plot for the test scores from the two classes.

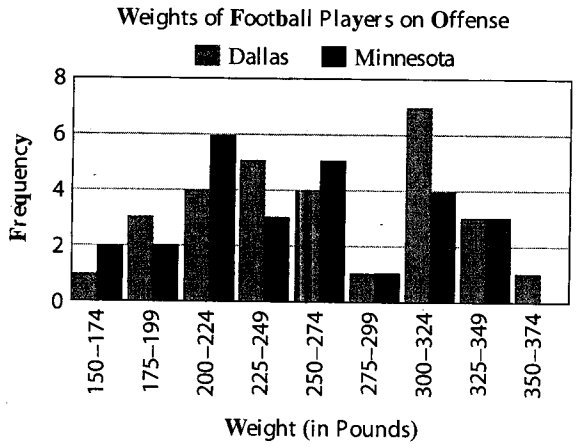
b. Describe any similarities and differences in the midterm test results for the two classes.

c. Do the data have any outliers or other striking features?

5. Babe Ruth was one of the greatest baseball players of all time. Among his many accomplishments were his lifetime and seasonal records for home runs. In his 15 years as a New York Yankee, Babe Ruth hit the following number of home runs per year: 54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, and 22. Next to Babe Ruth, the most productive home run hitter to wear a New York Yankee uniform was Mickey Mantle. In his 18 years as a Yankee, Mantle had the following home run totals: 13, 23, 21, 27, 37, 52, 34, 42, 31, 40, 54, 30, 15, 35, 19, 23, 22, and 18. Make a double stem-and-leaf plot of these data. How do Ruth and Mantle compare as hitters? (*Source:* www.baseball-reference.com.)

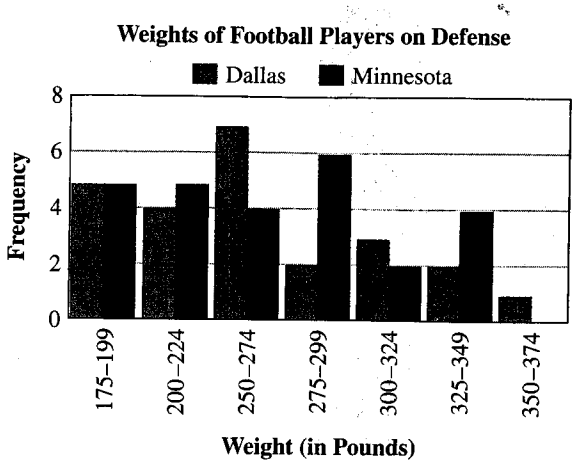
6. Seismologists use Richter-scale measurements to classify earthquakes. Earthquakes measuring less than 3.5 on the Richter scale are generally not felt, but they are recorded. On November 8, 2003, the following Richter-scale measurements were taken for earthquakes in California: 2.1, 1.3, 1.7, 2.0, 1.3, 1.1, 1.5, 2.6, 1.4, 1.0, 2.0, 2.0, 2.1, 1.7, 1.5, 0.9, 1.2, 2.9, and 0.8. On November 8, 2003, the following Richter-scale measurements were taken for earthquakes in Hawaii: 1.7, 1.5, 2.0, 3.3, 1.5, 1.7, 2.0, 2.0, 1.3, 1.4, 1.5, 1.6, 1.6, 1.4, 2.0, 1.9, 1.6, and 1.8. Make a double stem-and-leaf plot for these earthquake data. How did earthquakes compare for California and Hawaii? (*Sources:* Members of the Advanced National Seismic System, the U.S. Geological Survey, the North California Earthquake Data Center (NCEDC), and contributors to the NCEDC.)

7. The following comparison histogram displays the weights, in pounds, for offense players on the 2003 Dallas Cowboy football team and the 2003 Minnesota Viking football team.



   a. Find the length of the measurement classes.
   b. How many offense players play for each team?
   c. What percentage of Dallas' offense players weigh at least 250 pounds?
   d. What percentage of Minnesota's offense players weigh at least 250 pounds?
   e. What can you conclude about the weights of the offense players who play for Dallas compared with those who play for Minnesota?

8. The following comparison histogram displays the weights, in pounds, for defense players on the 2003 Dallas Cowboy team and the 2003 Minnesota Viking team.



   a. Find the length of the measurement classes.
   b. How many defense players play for each team?
   c. What percentage of Dallas' defense players weigh at least 250 pounds?
   d. What percentage of Minnesota's defense players weigh at least 250 pounds?
   e. What would you conclude about the weights of the defense players for Dallas compared with those who play for Minnesota?

9. Every year since 1928, the Academy of Motion Pictures Arts and Sciences has awarded an Oscar to the best actress in a leading role. The following list gives the ages of the actresses who received Oscars from 1928 to 1964: 22, 37, 30, 62, 32, 26, 31, 27, 26, 27, 30, 26, 29, 24, 39, 24, 29, 37, 30, 34, 34, 33, 28, 38, 45, 24, 26, 47, 41, 27, 39, 38, 28, 27, 31, 37, and 30. The following list gives the ages of the actresses who received Oscars from 1965 to 2004: 24, 34, 60, 61, 26, 35, 34, 34, 26, 37, 42, 41, 35, 31, 41, 33, 30, 74, 33, 49, 38, 61, 21, 41, 26, 81, 42, 29, 33, 35, 45, 49, 39, 34, 25, 25, 33, 33, 35, and 28. (*Source:* www.oscar.com.)
   a. Create a comparison histogram for this data.
   b. From 1928 to 1964, what percentage of actresses who won an Oscar for best actress in a leading role were at least 40 at the time of the award? Answer the same question for actresses who won from 1965 to 2004.
   c. What might you conclude about the ages of actresses winning best actress Oscars in the past 40 years compared with over 40 years ago?

10. The following table shows monthly rainfall amounts, in inches, over an 8-year period for Friday Harbor in the San Juan Islands.

Year							
2000	1999	1998	1997	1996	1995	1994	1993
2.88	6.24	4.99	4.99	4.47	1.95	1.35	2.16
1.45	4.53	1.78	2.76	2.45	4.51	2.04	0.45
2.02	1.94	2.73	3.56	0.77	1.94	2.01	2.10
1.28	1.29	0.48	0.94	3.08	1.66	1.13	1.45
2.82	0.66	2.53	1.27	2.56	0.39	0.71	2.09
1.36	1.94	0.96	3.09	0.51	0.84	1.44	1.48
0.47	0.31	1.34	1.14	0.40	1.30	0.95	1.14
0.41	1.31	0.10	0.03	0.34	2.98	0.66	0.58
1.06	0.26	0.31	2.23	1.66	0.56	1.78	0.33
2.58	3.75	1.37	5.08	4.08	4.04	3.00	1.83
3.42	4.22	9.04	2.34	3.38	9.62	3.39	1.74
3.59	5.29	6.85	3.45	8.09	4.39	4.01	2.61

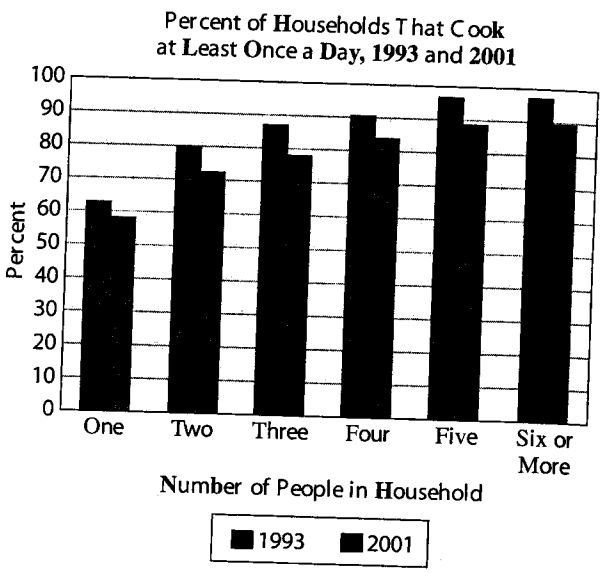
Source: www.sanjuanislander.com.

a. Create a comparison histogram for these data. Split the data into annual rainfall for the years 1993 to 1996 and annual rainfall for the years 1997 to 2000.

b. What percentage of months from 1997 to 2000 averaged at least 4 inches of rain? Answer the same question for the months from 1993 to 1996.

c. Compare the rainfall for each time period. What can you conclude?

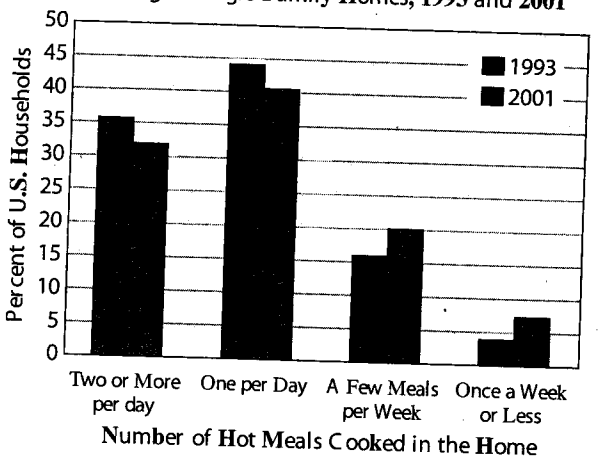
11. The Residential Energy Consumption Survey (RECS) collects data on household characteristics. The next multiple-bar graph uses survey data from 1993 and 2001 involving cooking trends and household sizes.



a. Estimate the change in the percentage of households that cook at least once a day from 1993 to 2001 for each of the household size classifications. What trend do you notice?

b. Which household size category experienced the greatest decrease in the percentage of households that cook at least once a day?

12. The Residential Energy Consumption Survey (RECS) collects data on household characteristics. The following double-bar graph uses survey data from 1993 and 2001 involving the number of meals cooked in single-family homes.



a. For single-family homes in 1993 and 2001, find the change in the percentage of U.S. households that cook two or more meals per day. Answer the same question for each of the categories.

b. Based on the double-bar graph, what can you say about the number of meals cooked in the home in 2001 compared with the number in 1993?

13. The percentages of the U.S. population, aged 18 or over, in each of four ethnic groups who were registered to vote in 2000 and the percentages of each ethnic group who actually voted in 2000 are given in the following table.

	Asian and Pacific Islander	Hispanic	Black	White
Percentage Registered	30.7	34.9	63.6	65.6
Percentage Voted	25.4	27.5	53.5	56.4

Source: U.S. Census Bureau.

a. Create a double-bar graph to display the data.

b. Which group of registered voters had the highest voter turn out? Which group of voters had the lowest percentage registered?

c. For each ethnic group, divide the percentage who voted by the percentage who were registered. How would you interpret this value? Which group had the largest value and which group had the smallest value?

14. The U.S. energy consumption for 1998 and 2002 by energy source is given in the following table. Energy consumption is measured in quadrillion BTUs.

Energy Source	1998	2002
Coal	21.7	22.2
Natural Gas	22.9	23.1
Petroleum	36.9	38.4
Nuclear Electric Power	7.1	8.1
Renewable Energy	6.5	5.9

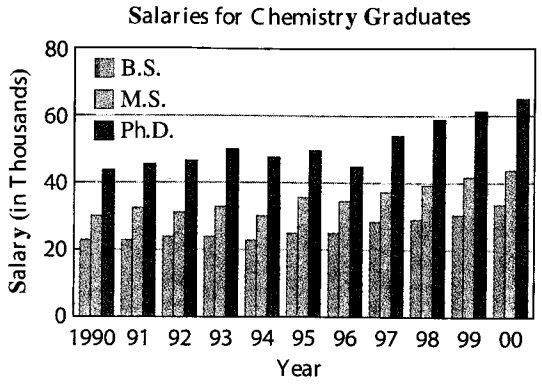
Source: www.eia.doe.gov.

a. Create a double-bar graph to display the data.

b. Find the change in energy consumption for each energy source from 1998 to 2002. Which energy source had the greatest increase? Were there any decreases?

c. For each of the energy-consumption changes you calculated in part (b), divide the change by the consumption value from 1998 and write as a percentage. How would you interpret these values? Which source had the largest value and which source had the smallest value?

15. The following graph displays the median salaries for 1990 through 2000 for new chemistry graduates who have earned a bachelor's degree (B.S.), a master's degree (M.S.), or a doctoral degree (Ph.D.). The salaries are for graduates with full-time positions and less than 1 year's technical work experience prior to graduation.



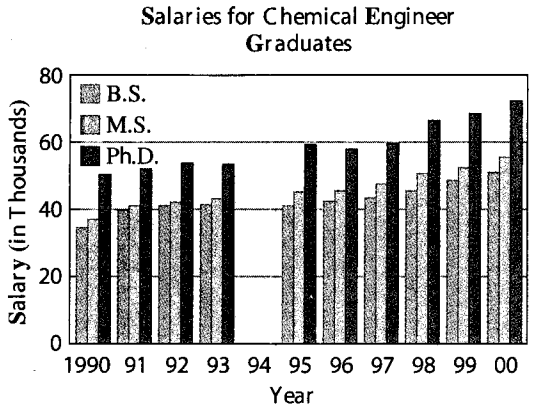
Source: Adapted from http://pubs.acs.org/cen/employment/7936/7936salarysurvey.html. Copyright 2001 American Chemical Society.

a. Estimate the salaries for chemistry graduates with a bachelor's degree in the years 1990, 1997, and 2000. What do you notice about the salaries for graduates with a bachelor's degree from 1990 through 1996?

b. Estimate the median salaries for chemistry graduates with a master's degree in the years 1990, 1995, and 2000. For each of the years given in the bar graph, find the difference between the salaries for graduates with a bachelor's degree and the salaries for graduates with a master's degree. What do you notice?

c. Estimate the median salaries for chemistry graduates with a doctoral degree in the years 1990, 1997, and 2000. During which time periods were salaries for graduates with a doctoral degree increasing?

16. The following graph displays the median salaries for new chemical engineering graduates from 1990 through 2000 who have earned a bachelor's degree (B.S.), a master's degree (M.S.), or a doctoral degree (Ph.D.). The salaries are for graduates with full-time positions and less than 1 year's technical work experience prior to graduation. (*Note:* The data from 1994 are not available.)



Source: Adapted from http://pubs.acs.org/cen/employment/7936/7936salarysurvey.html. Copyright 2001 American Chemical Society.

a. Estimate the salaries for chemical engineering graduates with a bachelor's degree in the years 1990, 1995, and 2000. Describe the trend in the salaries from 1990 through 2000. Estimate the median salary for the year 1994. Explain how you arrived at your estimate.

b. Estimate the median salaries for chemical engineering graduates with a master's degree in the years 1990, 1995, and 2000. For each of the years given in the bar graph, find the difference between the salaries for graduates with a master's degree and the salaries for graduates with a bachelor's degree. What do you notice?

c. Estimate the median salaries for chemical engineering graduates with a doctoral degree in the years 1990, 1997, and 2000. In which year did the salaries for chemical engineering graduates with a doctoral degree show the greatest increase compared with the previous year?

17. The Consumer Price Index (CPI) is prepared by the U.S. Bureau of Labor. A new CPI is released each month and provides a basis for comparing the changes in the cost of goods and services, and is often referred to as the **cost-of-living** index. As a reference point, the 1982 CPI is set to 100. The following table contains consumer price indices for three items for the months of October.

Item	1994	1997	2000	2003
Food/beverages	145.6	158.7	169.6	182.3
Apparel	135.2	134.9	132.8	121.5
Medical care	214.0	235.8	263.7	300.5

Source: www.bls.gov.

a. Create a multiple-bar graph to display the data.

b. In which 3-year period did the largest change occur in the CPI for food and beverages? Was it an increase or a decrease? Did the CPIs for other items experience their largest change in the same 3-year period?

c. How has the CPI for medical care changed over the years? Find the change in the CPI values given for medical care for each 3-year period. What do you notice?

d. Describe the changes in the CPI for apparel.

18. In the United States in 1990, the five languages most frequently spoken at home (other than English) were Spanish, French, German, Italian, and Chinese, in that order. The following table gives the total number of speakers of each language for the population 5 years old and older in the years 1990 and 2000.

Language	Number of Speakers in 1990	Number of Speakers in 2000
Spanish	17,339,172	28,101,052
French	1,702,176	1,643,838
German	1,547,099	1,382,613
Italian	1,308,648	1,008,307
Chinese	1,249,213	2,022,143

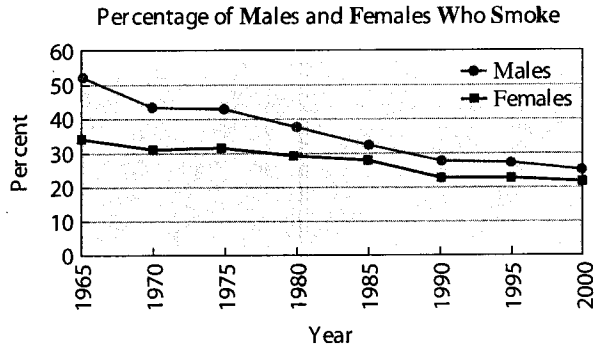
Source: www.census.gov.

a. Create a double-bar graph to display the data.

b. Which language experienced the greatest decrease in the number of speakers from 1990 to 2000? Which experienced the greatest increase?

c. How would the languages be ranked by the number of speakers in 2000?

19. In the United States in 2001, approximately 25.2 percent of men were smokers, while approximately 20.7 percent of women were smokers. The following double-line graph shows the percentage of men and women who were smokers from 1965 to 2000.

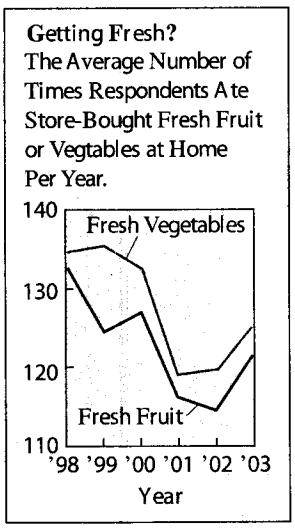


a. Describe the trend in the percentages of men and women who smoke.

b. Estimate the percentage of men who smoked in each 5-year period from 1965 to 2000. Do the same for the percentage of women who smoked.

c. During what 5-year period did the greatest decrease in the percentage of men who smoked occur? Answer the same question for women.

20. A study of the eating habits of 5000 people was conducted by the market-research firm NPD Group Inc. The following double-line graph shows the average number of times per year respondents ate store-bought fresh fruit or vegetables at home.



Source: Reprinted by permission of The Wall Street Journal, Copyright © 2003 Dow Jones & Company, Inc. All Rights Reserved Worldwide. License number 1345730455855.

a. Describe the trend in the average number of times per year store-bought fruits or vegetables were eaten at home from 1998 to 2003.

b. Estimate the average number of times per year respondents ate store-bought fresh fruit at home for each of the years from 1998 through 2003. Do the same for vegetables. Then add the average values for each year to get a total average for each year for fruits and vegetables combined.

c. Create a new line graph using the totals for each year from part (b). Compare the single-line graph with the double-line graph and explain which you prefer and why.

21. The following table contains the number of CDs and cassette tapes (in millions) sold from 1990 through 2001.

Year	CDs (in Millions)	Cassette Tapes (in Millions)
1990	286.5	442.2
1991	333.3	360.1
1992	407.5	366.4
1993	495.4	339.5
1994	662.1	345.4
1995	722.9	272.6
1996	778.9	225.3
1997	753.1	172.6
1998	847.0	158.5
1999	938.9	123.6
2000	942.5	76.0
2001	906.6	45.6

Source: Recording Industry Association of America, www.riaa.com/default.asp

a. Create a double-line graph to represent these data. Describe the trend in the numbers of CDs sold and the number of cassette tapes sold.

b. In approximately what year were total sales of CDs and cassette tapes the same?

c. Between which 2 years did CD total sales increase the most?

d. Between which 2 years did cassette tape sales decrease the most?

22. Collective bargaining agreements have set the minimum salaries for basketball players in the National Basketball Association (NBA) and the Women's National Basketball Association (WNBA), depending on the number of years of service. The following table contains the salary minimums for the 2003–2004 season.

Years of Service	WNBA Minimum Salary	NBA Minimum Salary
0	$30,000	$367,000
1	$30,000	$564,000
2	$30,000	$639,000
3	$30,000	$689,000
4	$42,000	$751,000
5	$42,000	$814,000
6	$42,000	$876,000

Source: www.nba.com and womensbasketballonline.com.

a. Create a double-line graph to represent these data. Describe the trend in the minimum salaries for the WNBA and for the NBA.

b. Between which 2 years do WNBA minimum salaries increase the most?

c. Between which 2 years do NBA minimum salaries increase the most?

23. The following table gives the median age at first marriage for females and males from 1890 through 1960.

MEDIAN AGE AT FIRST MARRIAGE		
Year	Females	Males
1890	22.9	26.1
1900	21.9	25.9
1910	21.6	25.1
1920	21.2	24.6
1930	21.3	24.3
1940	21.5	24.3
1950	20.3	22.8
1960	20.3	22.8

Source: U.S. Census Bureau.

a. Represent the data with a double-line graph.

b. Describe the trend in the median age at first marriage for females. Describe the trend in the median age at first marriage for males.

24. The following table gives the median age at first marriage for females and males from 1970 through 2000.

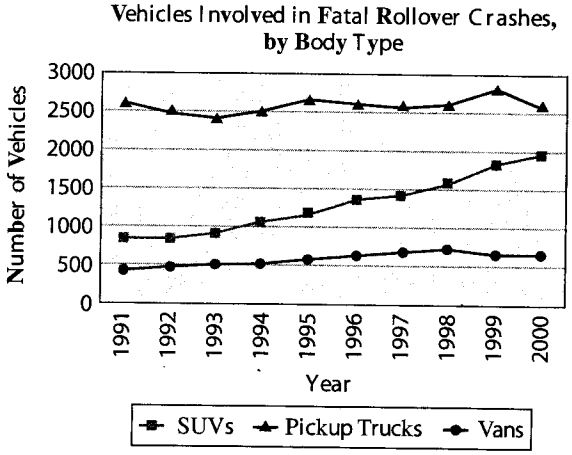
MEDIAN AGE AT FIRST MARRIAGE		
Year	Females	Males
1970	20.8	23.2
1980	22.0	24.7
1990	23.9	26.1
2000	25.1	26.8

Source: U.S. Census Bureau.

a. Combine these data with the data from the previous problem, and represent the data for 1890 to 2000 with a double-line graph.

b. Describe the trend in the median age at the first marriage for females for the time period 1890 to 2000. Describe the trend in the median age at first marriage for males for the time period 1890 to 2000.

25. On December 20, 2002, a $51.5 million settlement was reached with Ford Motor Company over allegations of deceptive trade practices relating to sales and advertising of the Ford Explorer and other SUVs. It was claimed that Ford's deceptive ads led consumers to believe that SUVs could be steered and handled like cars even in emergency situations, when the truth was that rollover crashes were more likely with SUVs. The next multiple-line graph shows the number of vehicles involved in fatal rollover crashes by body type.



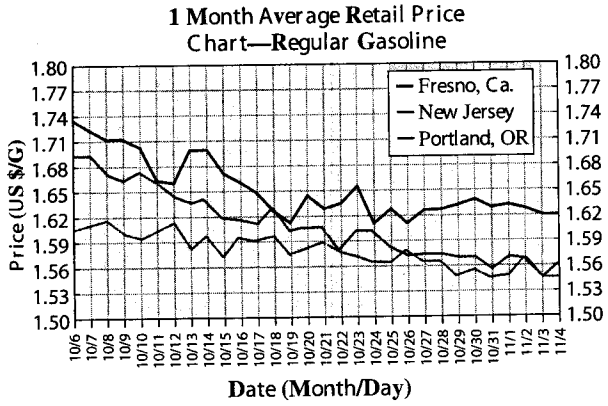
a. Estimate the number of pickup trucks involved in fatal rollover crashes in 1991, 1994, 1997, and 2000. Make the same estimates for SUVs and vans.

b. For each vehicle type, estimate the difference between the largest number of fatal rollover crashes in a year and the smallest number of fatal rollover crashes in a year.

c. Describe the general trend in the number of fatal rollover crashes for each type of vehicle.

d. Give one other reason, besides deceptive advertising, that could help explain the trend in fatal rollover crashes for SUVs.

26. In 2003, Oregon and New Jersey were the only two states that did not allow self-service gasoline stations. Supporters of self-service gas stations argued that pumping your own gas would result in lower prices. Compare the gasoline prices from October 6, 2003, through November 4, 2003, for New Jersey, Oregon, and California.



Source: Adapted from www.gasbuddy.com

a. Describe the general trend in regular gasoline prices from October 6, 2003, through November 4, 2003. How do prices compare for Oregon and New Jersey, where there is a ban on self-service gas stations, and Fresno, California, where consumers may pump their own gas?

b. Which location experienced the greatest drop in regular gasoline prices over the time period?

c. On what day did Fresno, California, experience the lowest price for regular gasoline? Answer the same question for Portland, Oregon, and New Jersey.

d. On what day(s) did Fresno, California, and Portland, Oregon sell regular gasoline for the same price?

e. On what day(s) did New Jersey and Portland, Oregon, sell regular gasoline for approximately the same price?

Problems 27 and 28

The following problems involve the calculation of a percentage change in a quantity. To find a percentage change, whether it is an increase or a decrease, find the difference between the new value and the original value, divide the difference by the original value, and multiply by 100%.

$$\text{percentage change} = \frac{\text{new value} - \text{original value}}{\text{original value}} \times 100\%.$$

27. Per-capita personal incomes for Alabama, California, and Oklahoma are given in the following table for the years 1998 through 2002. Per-capita personal income is the total personal income divided by the total midyear population. Midyear population estimates are from the Bureau of the Census.

State	1998	1999	2000	2001	2002
Alabama	22,054	22,972	23,521	24,477	25,128
California	28,163	29,856	32,149	32,655	32,996
Oklahoma	21,960	22,958	23,650	24,945	25,575

Source: U.S. Department of Commerce, Bureau of Economic Analysis, www.bea.doc.gov/bea/regional/spi/.

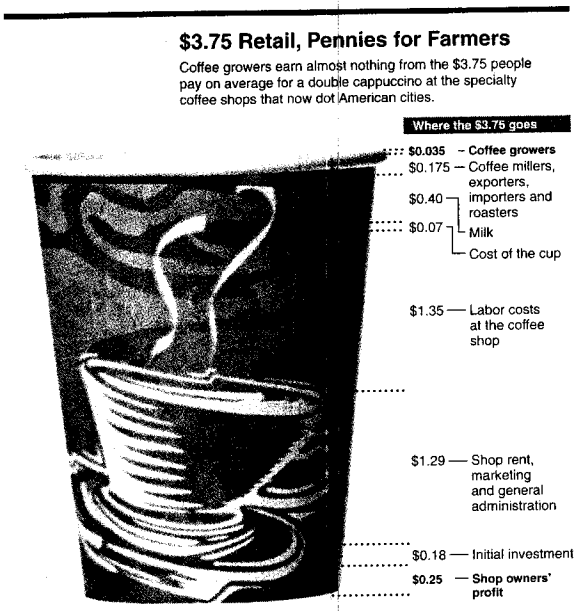
a. Create a multiple-line graph to depict the per-capita personal incomes for 1998 through 2002.

b. Find the percentage increase in per-capita personal income from 1998 to 2002 for each of the three states. Which state experienced the largest percentage increase in per-capita personal income? Which state experienced the smallest percentage increase?

c. During which year(s) did the per-capita personal income for Oklahoma residents exceed that of Alabama residents?

d. Find the percentage increase from 1999 to 2000, from 2000 to 2001, and from 2001 to 2002 for both Oklahoma and Alabama. Describe the trend in the percentage increases. Based on your observations, do you think the per-capita income for Alabama residents will exceed that of Oklahoma residents in the future? Explain.

21. Specialty coffee is a huge industry in the United States. The Colombian Coffee Federation is opening its own "high-end" Juan Valdez coffee shops in the United States. The first such store is in Manhattan. The following graphic appeared in the *New York Times* in November 2003. It shows where every penny of a $3.75 coffee goes.



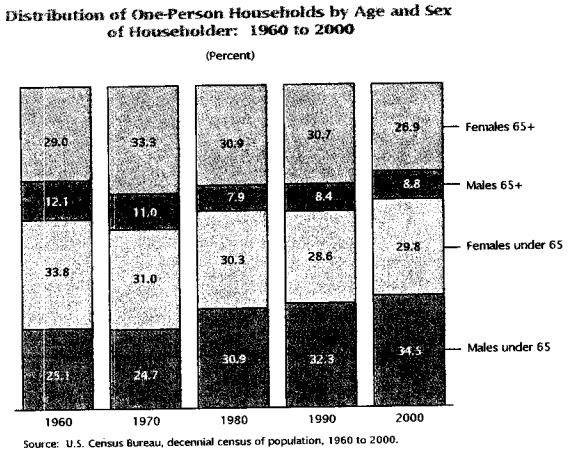
Source: The New York Times

a. Find the percentages of every $3.75 sale that goes to each of the eight categories. Round each value to the nearest tenth of a percent.

b. A quick breakdown of the cost of a cup of coffee might be shown in a pie chart. Use the percentages to find the number of degrees that would be associated with each sector of a pie chart and then create the pie chart. Round each value to the nearest tenth of a degree.

c. Summarize the information from the pie chart in a short paragraph. Where does the majority of the cost of the cup of coffee go? Which categories receive the smallest percentages of the cost?

22. The following proportional bar graph gives the distribution of one-person households by age and sex based on census data from 1960 through 2000. It shows what percentage of one-person households fall into each of four categories.



a. Describe the trend in the percentage of one-person households headed by males under 65. Describe the trend in the percentage of one-person households headed by males who were at least 65

b. Of all one-person households in each of the given years, find the percentages that were headed by males. Has the percentage steadily increased? Answer the same question for females.

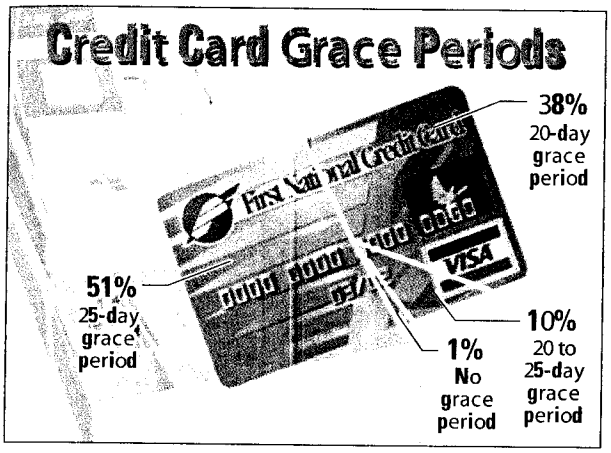
c. Describe the trend in the percentage of one-person households headed by females under 65. Describe the trend in the percentage of one-person households headed by females who were at least 65 years old.

23. Of the 2004 four-wheel drive minivans, the Chrysler Town & Country is reported to get 23 mpg on the highway. The Dodge Caravan also is reported to get 23 mpg on the highway, while the Toyota Sienna gets 24 mpg. *Source:* http://www.fueleconomy.gov.

a. Create a bar graph so that the differences between the highway gas mileages are downplayed.

b. Create a bar graph so that the differences between the highway gas mileages are emphasized.

24. The grace period for a credit card is the time between the statement date and the payment due date. A 2005 credit card survey found that grace periods vary, but the majority of credit cards have a 25-day grace period as shown in the following graph.



a. Describe what is misleading about this graph.

b. Create a single pie chart using the percentages given in the graph.

25. The recording industry has been battling Internet piracy of music and, in 2003, sued individuals in the United States for violating copyright laws by swapping files online. The following bar graph shows the number, in millions, of home computers used in file sharing, by source, in May 2003.

Share and Share Alike

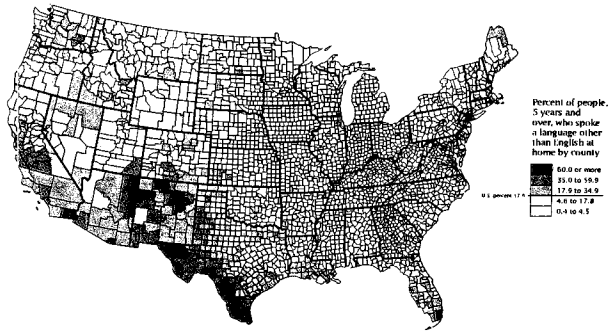
More than 11 million home computers in the United States were actively sharing files in May.

KeZeA	7.60 million
WinMX	2.30
IMesh	0.89
Morpheus	0.44
Grokster	0.30
LimeWire	0.05

a. For each category, compare the quantity represented by each bar to the length of the bar. What do you notice? Can you conclude that the bar lengths are proportional to the quantities they represent?

b. Suppose the length of the bar for the LimeWire category is 1 mm long. If a new bar graph were to be drawn so that the bars are proportional to the quantities they represent, what lengths that would be needed for each of the other bars?

26. Based on the results of the 2000 census, the following geographic map was created to show the percentage of people 5 years and older, by county, who spoke a language other than English at home. The national average is 17.9%.



http://www.census.gov/prod/2003pubs/c2kbr-29.pdf.

a. In what regions of the United States would you find the largest percentages of people 5 years and over who spoke a language other than English? In what regions of the United States would you find the lowest percentages?

b. Give possible reasons why the percentages are higher in certain regions of the country and lower in others.

28. Per-capita personal incomes for Pennsylvania, Nevada, and the United States are given in the following table for the years 1998 through 2002. Per-capita personal income is the total personal income divided by the total midyear population. Midyear population estimates are from the Bureau of the Census.

State	1998	1999	2000	2001	2002
Pennsylvania	27,469	28,619	29,504	30,752	31,727
Nevada	29,200	31,004	29,506	30,128	30,180
United States	27,203	28,546	29,469	30,413	30,941

Source: U.S. Department of Commerce, Bureau of Economic Analysis, www.bea.doc.gov/bea/regional/spi/.

a. Create a multiple-line graph to depict the per-capita personal incomes for 1998 through 2002.

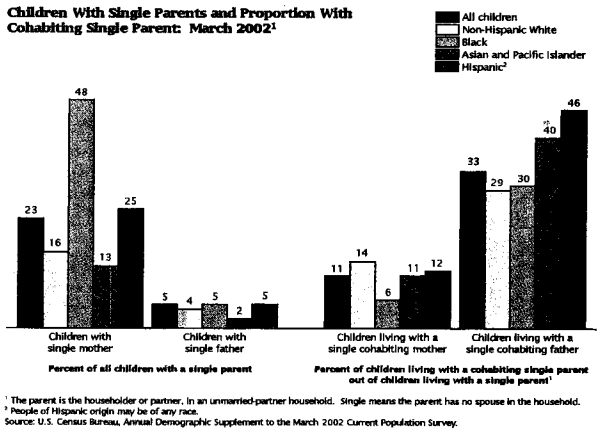
b. Find the percentage increase in per-capita personal income from 1998 to 2002 for Pennsylvania and Nevada. Which state experienced the larger percentage increase? Which state experienced the smaller percentage increase?

c. Which state's per-capita personal income most closely resembled the per-capita personal income for the United States from 1998 to 2002? For this state and the United States, find the difference in the per-capita personal income for each of the 5 years. Describe the trend.

d. Find the percentage increase or decrease from 1998 to 1999, from 1999 to 2000, from 2000 to 2001, and from 2001 to 2002 for Pennsylvania and Nevada. What do you observe?

Problems 29 and 30

The welfare of children is a concern to parents and to society in general. Based on current population surveys, the U.S. Census Bureau has produced a population report for 2002. In March 2002, approximately 72 million children under the age of 18 were living in the United States. The following multiple-bar graph, which shows where many of these children lived, was included in the Bureau's report.



Source: U.S. Census Bureau, Children's Living Arrangements and Characteristics: March 2002.

29. Consider the two multiple-bar graphs for the "Percent of all children with a single parent."

a. Can you redraw the bar graphs using two pie charts? Explain why or why not.

b. What do the numbers at the top of each bar represent?

c. For which ethnic group did the largest percentage of children live with a single mother? For which ethnic group did the smallest percentage of children live with a single mother?

d. For the two ethnic groups from part (c), create two comparison pie charts. Note that each pie chart will have three sectors. One sector will correspond to the percentage of children of that ethnic group who live with a single mother. Another will correspond to the percentage of children of that ethnic group who live with a single father. How will the third sector be labeled?

30. Consider the two multiple-bar graphs for the "Percent of children living with a cohabiting single parent out of children living with a single parent."

a. Can you redraw the two bar graphs using two pie charts? Explain why or why not.

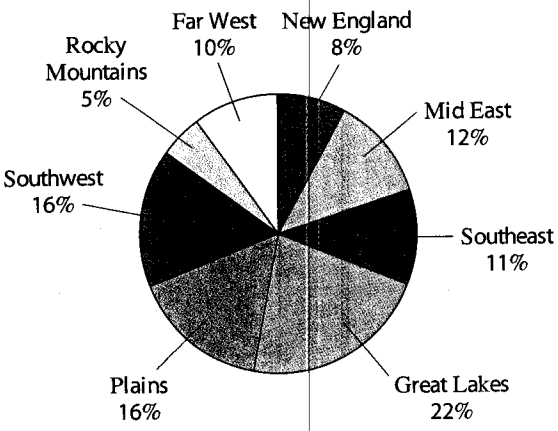
b. What do the numbers at the top of each bar represent?

c. For which ethnic group did the largest percentage of children live with a single cohabiting father? For which ethnic group did the smallest percentage of children live with a single cohabiting father?

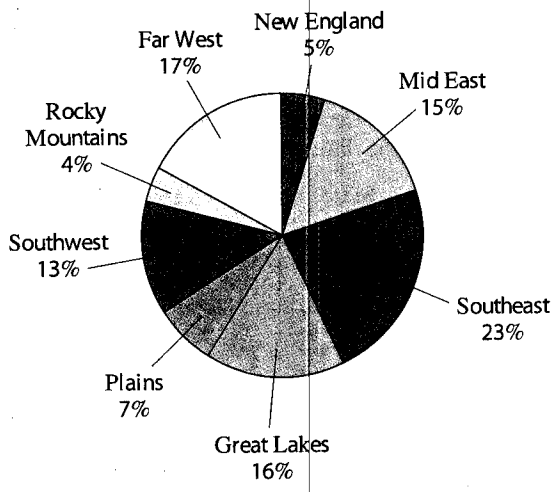
d. For the two ethnic groups from part (c), create two comparison pie charts. Note that each pie chart will have three sectors. One sector will correspond to the percentage of children of that ethnic group

who live with a single cohabiting father. Another will correspond to the percentage of children of that ethnic group who live with a single cohabiting mother. How will the third sector be labeled?

Percentage of School Districts by Region 2002–2003



Percentage of Student Enrolled Per Region 2002–2003



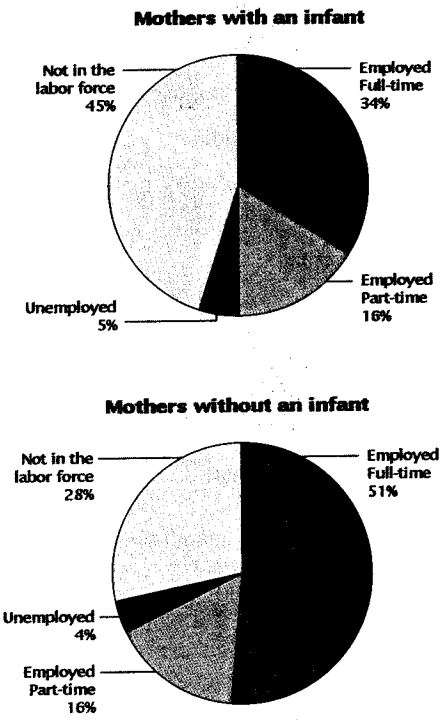
31. The percentage of school districts by region and the percentage of students enrolled in public school by region are given in the preceding pie charts.

a. If there were 15,215 operating school districts in the United States in the 2002–2003 school year, then approximately how many school districts were there in each region? Round your answers to the nearest whole number.

b. If there were 47,792,369 students enrolled in school in the United States in the 2002–2003 school year, then approximately how many students were enrolled in each region? Round your answers to the nearest whole number.

c. Compare the two pie charts. Which regions have a larger percentage of districts and a smaller percentage of students enrolled? Which regions have a smaller percentage of districts and a larger percentage of students enrolled?

32. The following pie charts summarize the labor-force participation of mothers, aged 15 to 44, with infants and without infants.



Source: U.S. Census Bureau, Current Population Survey, June 2002.

a. If there were 3,766,000 mothers with an infant in June 2002, then how many mothers were there in each employment category?

b. If there were 30,905,000 mothers without an infant in June 2002, then how many mothers were there in each employment category?

c. Compare the two pie charts. What can you conclude?

Problems 33 and 34

The week before Special Prosecutor Kenneth Starr testified before the House Judiciary Committee, which acted to impeach President Clinton in early 1999, a public opinion poll sought to determine how the American public would react. Presented in the following two problems are the results of a number of questions asked in a CNN/USA Today/Gallup Poll survey. Results are based on telephone interviews conducted November 13–15, 1998, with 1039 adults nationwide. The margin of error is plus or minus 3 percentage points.

33. The following three questions were posed in the survey. Create pie charts to display the results for each question. What might you conclude?

Question: Do you approve or disapprove of the way Bill Clinton is handling his job as president?

Approve	66%
Disapprove	31%
No opinion	3%

Question: What would you want your member of the House of Representatives to do?

Vote to impeach	31%
Vote not to impeach	66%
No opinion	3%

Question: If the House does vote to impeach President Clinton and sends the case to the Senate for trial, what would you want your senators to do?

Vote in favor of convicting	30%
Vote against convicting	68%
No opinion	2%

34. The following two questions were also posed in the survey. Create pie charts to display the results for each question. What might you conclude?

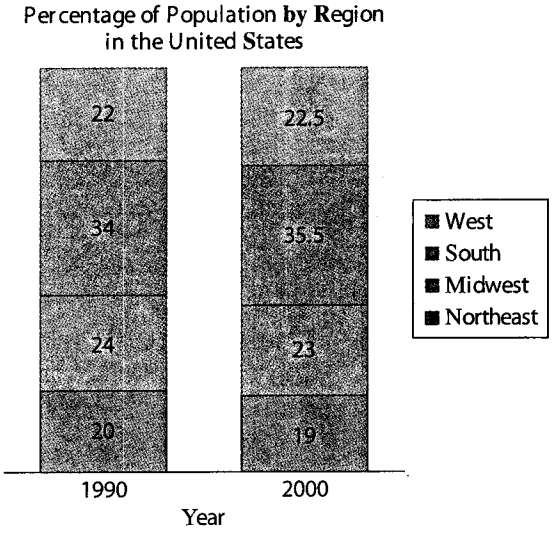
Question: Which would you prefer?

Continue hearings	26%
Censure and stop hearings	35%
Drop altogether	39%

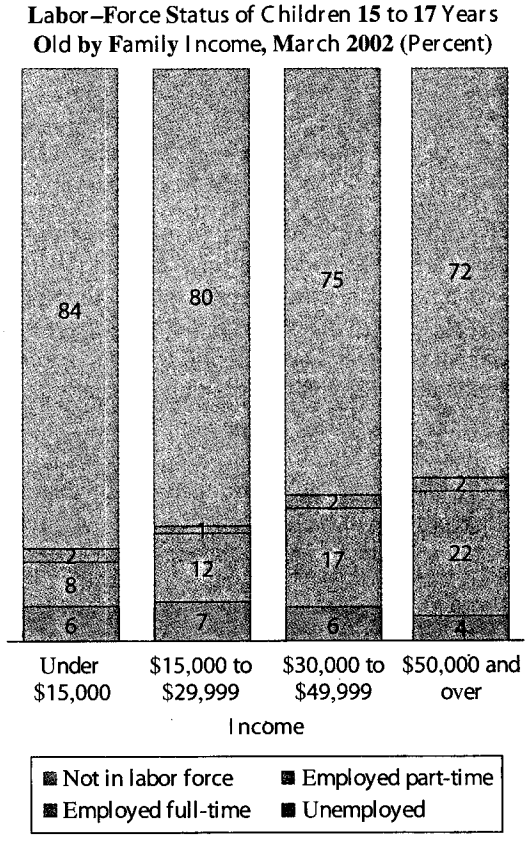
Question: Do you approve of the decision to hold these hearings?

Strongly approve	22%
Moderately approve	18%
Moderately disapprove	24%
Strongly disapprove	35%
No opinion	1%

35. The next proportional bar graph gives the population distribution for the United States by region for 1990 and 2000. What conclusion would you make about the percentage of the population living in each region in 1990 and in 2000?

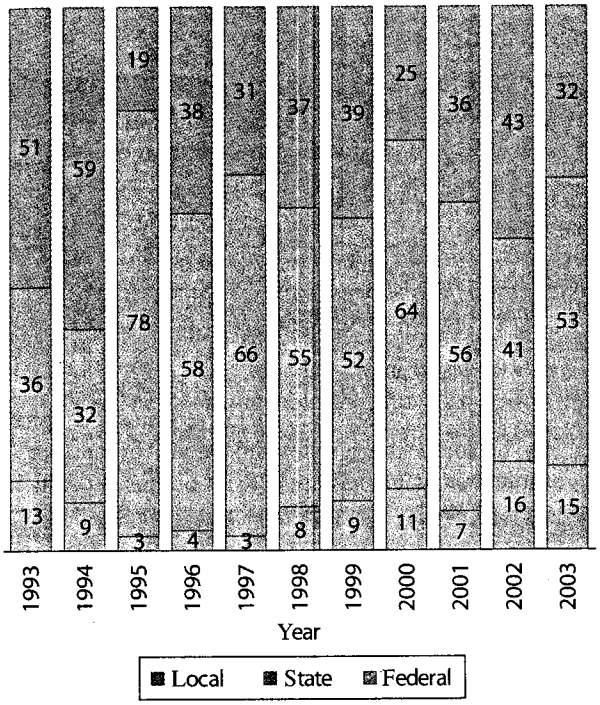


36. The labor-force status of children, 15 to 17 years old, for four family-income categories is given in the following proportional bar graph. What conclusion can you make about 15- to 17-year-old children who work in relation to the different family income levels?



37. According to the National Education Association, one indicator of support for improvement in education is an increase in new funding each year. The following proportional bar graph gives the percentage of federal, state, and local revenue that contributes to the *increase* in funding compared to the previous year.

Percentage of the Revenue Increase by Source, 1993–2003



a. Of the three sources that contribute to the revenue increase each year, which source contributes the least, on average? Which source contributes the most?

b. Which of the three sources is the most stable; that is, which source fluctuates the least?

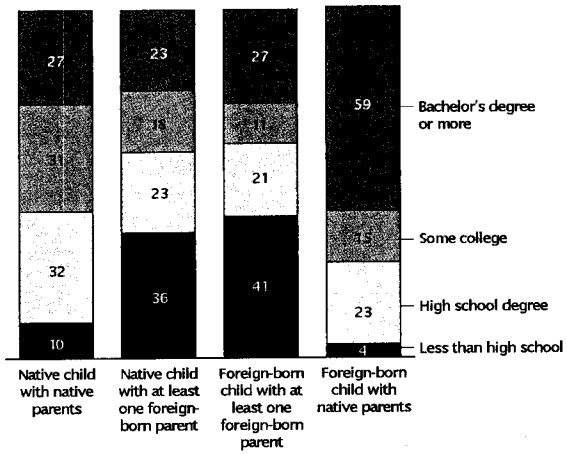
c. In which year was the percentage contributed by the local source the smallest?

d. Between which 2 years did the percentage contributed by the state change the most? Was it an increase or a decrease?

38. The foreign-born population of the United States has increased significantly since the 1970s. Approximately 12 million native children (born in the United States.) lived with at least one foreign-born parent in 2002. For children living with at least one parent in the United States, the following proportional bar graph summarizes the children's and parents' nativity (place of birth) classified according to the parent's education.

Children's and Parents' Nativity by Parent's Education for Children Living With At Least One Parent: March 2002[1]

(In percent)



[1] Children with any foreign-born parents are included with foreign-born parent, children with native parents only are with native parents.
Education is the mother's, if not available, the father's is used.
Source: U.S. Census Bureau, Annual Demographic Supplement to the March 2002 Current Population Survey.

Source: www.census.gov, Children's Living Arrangements and Characteristics: March 2002.

a. What is the most striking difference between foreign-born parents and native parents when education is considered?

b. In which category is a child most likely to be living with educated parents, that is, with parents with at least a high-school degree?

c. What conclusion can be made about education in the United States for native parents?

39. Several of the leading causes of death for people in the United States aged 65 and over for the years 1980 and 2001 are listed in the following table.

Cause of Death	1980	2001
Heart diseases	44.4	32.4
Pneumonia and influenza	3.4	3.0
Malignant neoplasms	19.3	21.7
Diabetes mellitus	1.9	3.0
Cerebrovascular diseases	11.0	8.0
Other	20.0	31.9

Source: www.cdc.gov.

a. Present these data using a proportional bar graph.
b. Present these data using pie charts.
c. Summarize the changes in the leading causes of death between 1980 and 2001.
d. Which type of graph, created in parts (a) and (b), best demonstrates the important features of the data? Explain.

40. The highest level of education attained by U.S. citizens by gender in the year 2000 is given in the next table.

a. Present these data using a proportional bar graph.
b. Present these data using pie charts.
c. Summarize the differences in the highest level of education attained for males and females in 2000.
d. Which type of graph, created in parts (a) and (b), best demonstrates the important features of the data? Explain.

Highest Level of Education	Male	Female
Less than high school	28.6	21.8
High-school graduate or equivalent	30.2	26.9
Some college or associate degree	34.9	42.0
Bachelor's degree or higher	6.3	9.3

Source: www.census.gov.

41. The following table contains tax revenue per capita for 11 Southeastern states and the tax revenue per capita that would have resulted if proposed tax increases had been in place in 2000.

a. Study the table and comment on the current tax revenue per capita versus the tax revenue per capita that would have resulted from the proposed tax increase.
b. Create a graph to emphasize what you noticed from part (a). Give a reason why you selected the type of graph you created.

State	Current Tax Revenue per Capita (in Dollars)	State	Proposed Tax Revenue per Capita (in Dollars)
Georgia	2841	Georgia	2841
North Carolina	2664	North Carolina	2664
Florida	2624	Florida	2624
Kentucky	2517	Kentucky	2517
Louisiana	2436	Louisiana	2436
West Virginia	2413	West Virginia	2413
South Carolina	2379	Alabama	2387
Arkansas	2230	South Carolina	2379
Mississippi	2214	Tennessee	2359*
Tennessee	2185	Arkansas	2230
Alabama	2117	Mississippi	2214

*Includes revenue from a tax increase passed by Tennessee in 2002
Source: Public Affairs Research Council of Alabama.

	Percent Distribution of People 55 Years and Over			
Ethnic Group	55 to 64	65 to 74	75 to 84	85 and Over
White	42.1	30.1	21.6	6.2
Black	46.7	33.0	16.1	4.3
American Indian and Alaska Native	51.7	30.0	14.6	3.6
Asian and Pacific Islander	51.1	30.0	16.1	2.9
Hispanic	50.6	30.6	14.7	4.1

Source: www.census.gov.

42. The preceding table contains the percentages of people aged 55 years and over by ethnic group.

a. Study the table and comment on the similarities and differences.

b. Create a graph to emphasize what you noticed from part (a). Give a reason why you selected the type of graph you created.

43. Dollar stores are attracting more shoppers these days. The following table contains the percentages of households that shopped in dollar stores by income level in 2000 and in 2002.

	Percentage of Households That Shop in Dollar Stores	
Household Income in Dollars	2000	2002
At least $70,000	37	45
$50,000 to $69,999	48	58
$40,000 to $49,999	54	64
$30,000 to $39,999	57	67
$20,000 to $29,999	62	71
Less than $20,000	67	74

Source: www.acnielson.com.

a. Create two different displays of these data. Which graphical representation seems to best illustrate the data, and why?

b. Summarize the pattern in the percentage of shoppers shopping in dollar stores.

44. Turkey exports are increasingly important to turkey producers in the United States. The following table shows United States' turkey exports, in millions of pounds, from the years 1991 through 2002.

Year	Whole Body	Parts and Cut-Up
1991	16.6	105.4
1992	20.3	181.5
1993	19.5	224.2
1994	24.1	256.3
1995	24.8	323.2
1996	34.7	403.3
1997	40.8	565.1
1998	36.6	409.6
1999	41.9	337.1
2000	33.3	412.0
2001	20.3	466.6
2002	15.2	423.3

Source: www.eatturkey.com.

a. Create two different displays of this data. Which graphical representation seems to best illustrate the data, and why?

b. Summarize the trends in turkey exports.

Extended Problems

45. Collect as many examples as you can find of comparison histograms, multiple-bar graphs, multiple-line graphs, multiple pie charts, and proportional bar graphs. Look in newspapers, magazines, and political fliers. From the graphs you find, select three different types and write a paragraph explaining what the graph illustrates. Evaluate the appropriateness of the form of the graph. If you think the data would have been more effectively represented by a different form of graph, create a new graph and explain why your graph is an improvement over the one you found.

46. Search the U.S. Census Bureau's website at www.census.gov to find four different sets of data.

a. Find data related to males and females such as educational levels or employment status over the past 10 to 50 years. Create a double-line graph to represent the data. Summarize in a paragraph the interesting features or changes in the data.

b. Find data such as population marital status for 1990 and 2000. Create comparison pie charts or a proportional bar graph. In a paragraph, summarize interesting features or changes in the data.

c. Find household income, housing, or other data for 1990 and 2000. Create a comparison histogram. Summarize in a paragraph the interesting features or changes in the data.

47. When using a computer program such as Microsoft Excel or Microsoft Word to create graphs, you will find many options, some of which may be new to you. The **Chart Wizard**  in Excel, displays all of the options. Consider the following three types of charts:

The **Area Chart,**  , emphasizes the magnitude of changes over time. An area chart shows the relationships of each part compared to the whole as well as changes over time.

The **Doughnut Chart,**  , is similar to the pie chart but it can contain more than one data series. Each ring of the doughnut chart represents one data series.

In a **Radar Chart,**  , each category's values radiate from the center. All points in the same category or series are connected by lines.

a. In problem 19, a double-line graph displays the percentages of men and women, from 1965 to 2000, who were smokers. The following table gives the actual data values used to create the double-line graph.

	1965	1970	1975	1980	1985	1990	1995	2000
Males	51.9	44.1	43.0	37.6	32.6	28.4	27.0	25.7
Females	33.9	31.5	32.0	29.3	27.9	22.8	22.6	21.0

Use Excel to create an area chart to display these data. Compare and contrast the area chart to the double-line graph from problem 19. Which graph do you feel is a better representation of the data? Why?

b. In problem 32, two pie charts displayed the employment status of mothers 15 to 44 years old in 2002. The following table gives the percentages used to create the two pie charts.

Employment Status	Mothers with an Infant	Mothers without an Infant
Unemployed	5%	4%
Employed part-time	16%	16%
Employed full-time	34%	51%
Not in the labor force	45%	28%

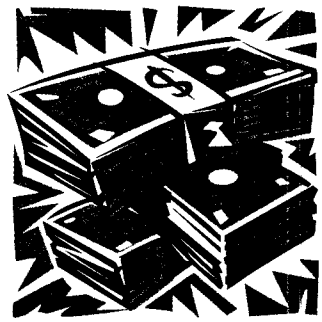
Use Excel to create a doughnut chart to display these data. Compare and contrast the doughnut chart to the pie charts from problem 32. Which representation do you feel is easier to interpret? Why?

c. Problem 24 presented the median age at first marriage for females and males from 1970 through 2000, and you were asked to create a double-line graph. Use Excel to create a radar graph of those data. Compare and contrast the radar chart to the double-line graph you created in problem 24. Which chart is easier to interpret? Why?

48. The Gross Domestic Product (GDP) is a measure of the total goods and services produced by the United States. Gather the data on the GDP for the most recent 10-year period available. This information may be found at the U.S. Department of Commerce Bureau of Economic Analysis's website at www.bea.doc.gov. Prepare two bar graphs for the data. One graph should show the actual value of the GDP, and the other should show the percentage change in GDP from one year to the next. How are the increases and decreases of the values in one graph related to those of the other?

8.3 Enhancement, Distraction, and Distortion

INITIAL PROBLEM



Suppose that you are on a debating team and know you will argue the topic, "Resolved, the most important economic issue facing the country today is the federal debt." You do not know which side of the issue you will have to argue. As part of your preparation for the debate, you decide to make two graphs, each of which illustrates the federal debt over time. The graphs will show the federal debt from two different perspectives. One graph will depict the debt in the most unfavorable light possible and the other will show the debt in a more optimistic manner. Make two such graphs using the federal debt data from 1965–2000 shown in the following table.

Year	Federal Debt (to the Nearest Billion Dollars)
1965	321
1970	389
1975	577
1980	930
1985	1946
1990	3233
1995	4974
2000	5674

A solution of this Initial Problem is on page 543.

Tidbit

The Dow Jones Industrial Average (DJIA) was created in 1884. The prices of 12 stocks originally constituted the average; prices were adjusted so the average began at 100. The number of stocks determining the DJIA increased to 20 in 1916 and finally to 30 in 1928. If one of the 30 stocks goes out of business or is deemed not to be worthy, it is replaced. In 1999, Intel and Microsoft, two tech stocks, replaced Chevron and Goodyear, two industrial stocks.

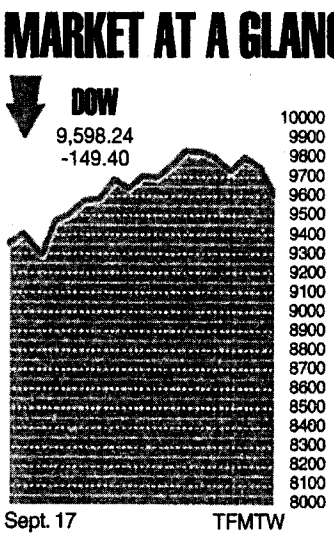


Figure 8.25

Source: Salem Statesman Journal, November 27, 2003. Adapted from Gannett News Service.

Accurately communicating your ideas, either verbally or visually, is not always an easy task. When we present quantitative information in a graphical form, we have to consider which type of graph to use, what trend or pattern in the data to emphasize, and how to construct the actual graphs. If some aspect of the graph is distorted, the graph may give a misleading impression or readers may draw the wrong conclusion. Some forms of distortion are considered common practice, and those who regularly use such graphs are accustomed to them. One example is the reporting of stock market information such as the Dow Jones Industrial Average. The graph in Figure 8.25 shows the performance of the Dow Jones Industrial Average over the 5-week period from September 17, 2003, to October 22, 2003.

Notice that the vertical scale shown on the right side of the graph does not start at 0. Instead, only the values above 8000 on the vertical axis are shown in the graph. Cutting off the bottom portion of the vertical scale emphasizes the vertical change over shorter periods of time. In this case, the readers expect and want to see change over short periods of time. Including the entire vertical scale from 0 to 10,000 would make it more difficult to see the fluctuations in the Dow over time.

Part of the vertical scale is also commonly omitted in weather reports. The graph shown in Figure 8.26 appeared in a newspaper report of temperatures for a 24-hour period in Salem, Oregon, on November 28, 2003. Here the vertical scale begins at 20°F, rather than at 0°F. Newspaper readers are most interested in temperature variations and in the high and low temperature during the day. Shortening the vertical axis of the graph makes it easier for readers to focus on those features.

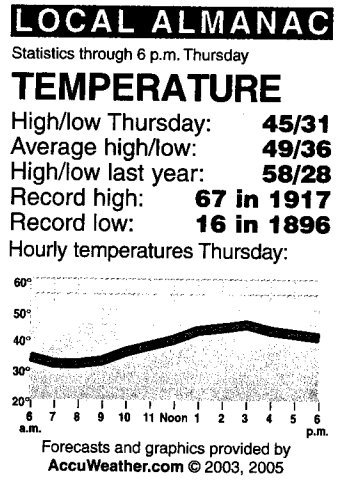


Figure 8.26

Source: Salem Statesman Journal, November 28, 2003. Permission courtesy of AccuWeather.com

As we will discover later in this section, shortening the vertical axis (starting at a value other than 0) is one way to create a misleading graph. Distortion in a graph may be unintentional and may not affect how the graph is interpreted. However, at other times distortion in a graph is deliberate, and its intent is to deceive or misdirect the reader. In this section, we will look at ways in which the elements of a graph can be manipulated to create different impressions of the data.

First, we consider variations on the basic kinds of graphs. In particular, we will consider ways in which a graph can subtly mislead a reader. We will learn how to determine when a graph is misleading and examine unbiased methods for presenting your viewpoint in a favorable light. Then we will consider graphs that have been enhanced by incorporating pictures. These graphs are more visually appealing and can reinforce a particular message, but they, too, can be misleading.

SCALING AND AXIS MANIPULATION

To emphasize the differences among the bars of a histogram or bar chart, you can display the chart with part of the vertical axis missing. We already mentioned the use of this technique in stock reports. If, however, your reader is not accustomed to seeing the vertical axis of a graph shortened or not labeled, your graph can be misleading. For example, Beary Sticks, a children's cereal with a high sugar content level (9 grams per serving), is advertised as wholesome because it contains less sugar than other children's cereals. The high-sugar-content cereals chosen for comparison have the following grams of sugar per serving: 15, 14, 13, and 11. The bar graph in Figure 8.27 appears on the Beary Sticks cereal box.

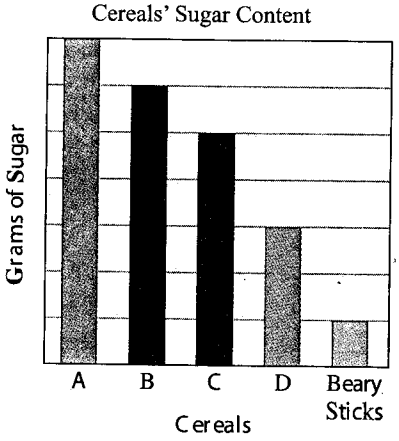


Figure 8.27

The graph implies that Beary Sticks cereal has far less sugar than the four other brands. Notice that the graph does not show the scale of the vertical axis, and the axis actually begins at 8 instead of 0, with each horizontal line on the vertical axis representing 2 grams of sugar. A less-misleading graph might look like the one in Figure 8.28. Notice in this case that the scale on the vertical axis is shown, the axis starts at 0 grams of sugar, and each darkened tick mark represents 2 grams of sugar. The graph still indicates that Beary Sticks cereal is lower in sugar than the other four brands, but the true difference in sugar content, as shown in this graph, is not quite so striking.

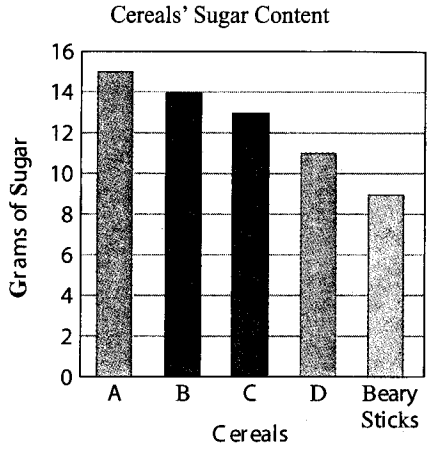


Figure 8.28

Notice also that the Beary Sticks Company chose not to compare the sugar content of their cereal with the sugar content of either Corn Flakes (at 2 grams per serving) or Shredded Wheat (at 0 grams per serving).

EXAMPLE 8.10 The prices of three brands of baked beans are as follows:

Brand X: 79¢ Brand Y: 89¢ Brand Z: 99¢.

Draw a bar graph of the price data so that Brand X looks like a much better buy than the other two brands.

SOLUTION To emphasize the difference in prices, we want to exaggerate the difference in the heights of the bars. Brand X can be made to look much cheaper than the other two brands by starting the price scale at 75¢, rather than at 0¢, as shown in Figure 8.29.

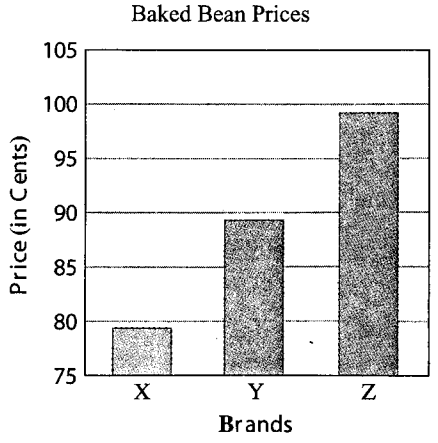


Figure 8.29

■

As you read articles in which data are presented visually, pay special attention to the vertical scale on bar graphs, line graphs, and histograms. Figure 8.30 shows several bar graphs on the performance of high school students on the SAT and ACT college-entrance exams for the years 1999 to 2003.

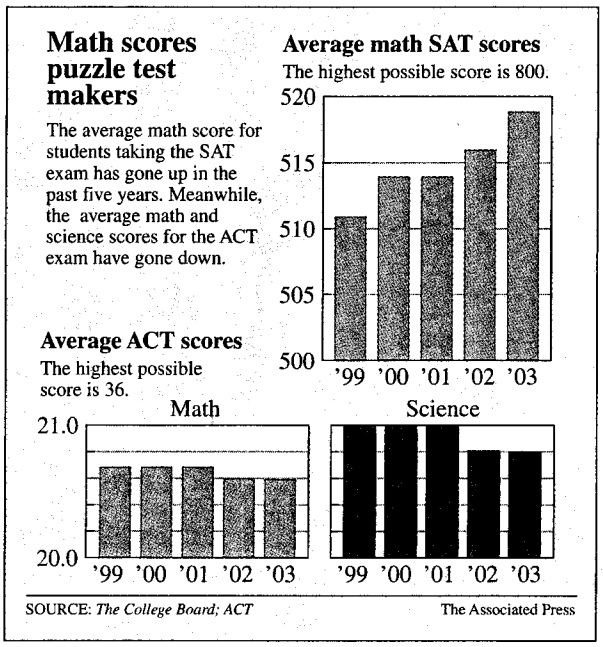


Figure 8.30

Source: AP/Wide World Photos

Tidbit

The Scholastic Assessment Test (SAT) is generally taken by high school juniors and seniors in preparation for admission to college. A new, revised version of the SAT was given for the first time in March 2005. Whereas the "old" SAT had two parts: Math and Verbal, the "new" SAT has three parts: Math, Critical Reading, and Writing, each of which is scored on the basis of 200–800 points. Thus, as of March 2005, a "perfect" SAT score is now 2400, compared to the old perfect score of 1600. While some parts of the old SAT test, such as verbal analogies, were eliminated, a new 25-minute written essay, was added to the 2005 version of the SAT.

None of the vertical scales of the graphs begins at 0. So the differences in students' performance on the exams from year to year appear to be greater than they are. If the math SAT bar graph were redrawn with the vertical scale starting at 0, the graph would look like the one in Figure 8.31. This new graph shows little difference in the heights of the bars but still shows that scores increased over the 5-year period.

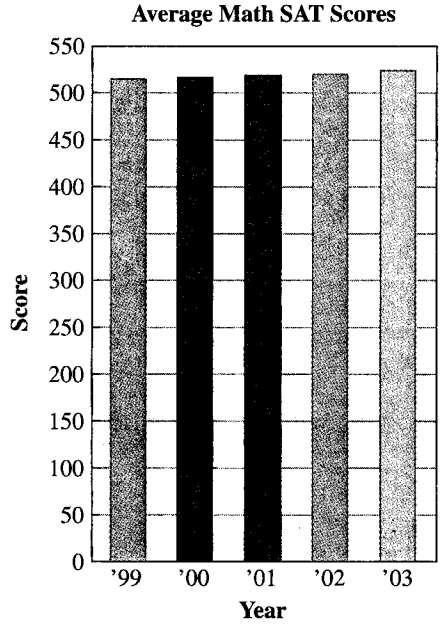


Figure 8.31

Reversing the axes or reversing the orientation of one of the axes is another technique that can create a misleading impression of the data. Figure 8.32 is a bar graph that shows a company's profits declining over time.

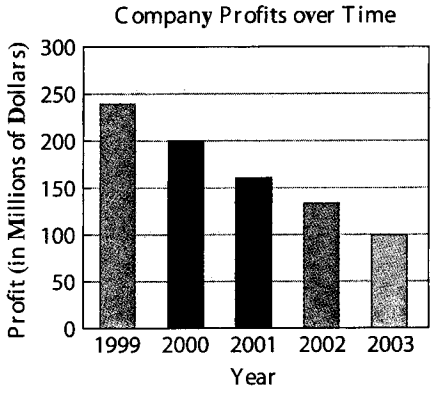


Figure 8.32

In Figure 8.33, the same data are displayed in a horizontal bar graph in which the years appear on the vertical axis rather than the horizontal axis. The years are also arranged in reverse order, with the most recent year listed at the bottom of the vertical scale.

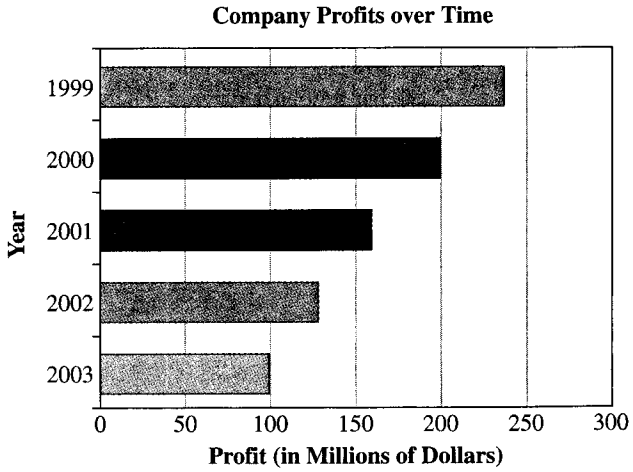


Figure 8.33

Table 8.13

Year	Crimes per 1000 People
1999	25
2000	30
2001	32
2002	34
2003	38

The chart in Figure 8.33 displays the same information as in Figure 8.32, but this new graph does not have the negative "feel" of a decreasing trend in profits. We are accustomed to reading graphs horizontally from left to right and vertically from bottom to top, so in Figure 8.33, the eye is naturally drawn to the top of the graph, where the longest bar and greatest profit are shown.

EXAMPLE 8.11 Some data from a crime-ridden community are given in Table 8.13. Present these data in a graph that might be used to give citizens the impression that things are getting better rather than worse.

SOLUTION The graph in Figure 8.34 could give the impression that the community is becoming safer. Three features of the graph contribute to this impression: (1) the vertical axis of the graph starts at 20 rather than at 0, (2) the horizontal axis shows the years in decreasing order, and (3) the graph is drawn narrow and tall so that the apparent declining trend in the crime rate is emphasized.

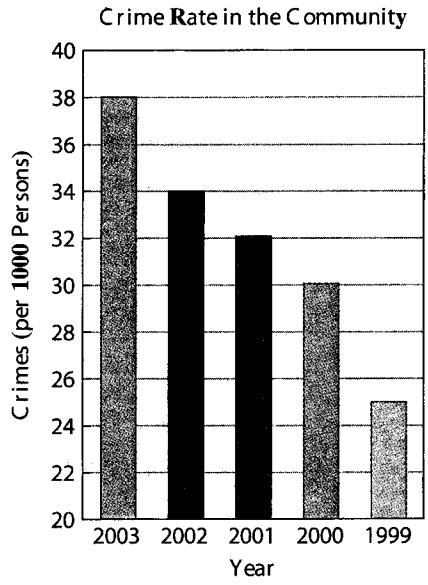


Figure 8.34

LINE GRAPHS AND CROPPING

We have seen that shortening the vertical axis or reversing the scale of the axes can make bar graphs deceiving. The same changes can also make line graphs misleading. Figure 8.35 shows the company's profits from Figure 8.32 displayed as a line graph.

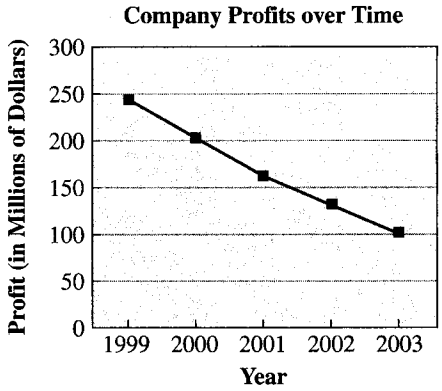


Figure 8.35

The decline in profits can be made to appear less dramatic by extending the scale of the vertical axis so that the maximum is 500 rather than 300, as shown in Figure 8.36. Notice that the change in scale makes the line graph look less steep.

The scale manipulation in Figure 8.36 is an example of **cropping**—that is, the choice of window used to present the data. Selecting a smaller or larger window can make a

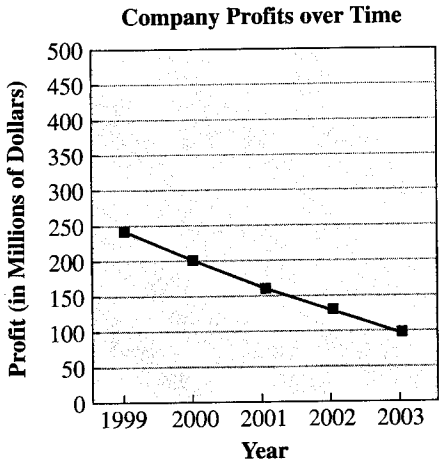


Figure 8.36

trend appear more or less impressive. For example consider a close-up television image of a violent street demonstration. If the camera pans and we see that there are only a few demonstrators, the event may not seem very impressive.

EXAMPLE 8.12 Draw two line graphs of the crime data from Example 8.11 that give different impressions of the situation.

SOLUTION One way to present different views of the trend in the crime rate is to show different portions of the vertical axis. We will create one graph in which the vertical axis begins at 0 [Figure 8.37(a)] and one graph in which the vertical axis begins at 24 [Figure 8.37(b)].

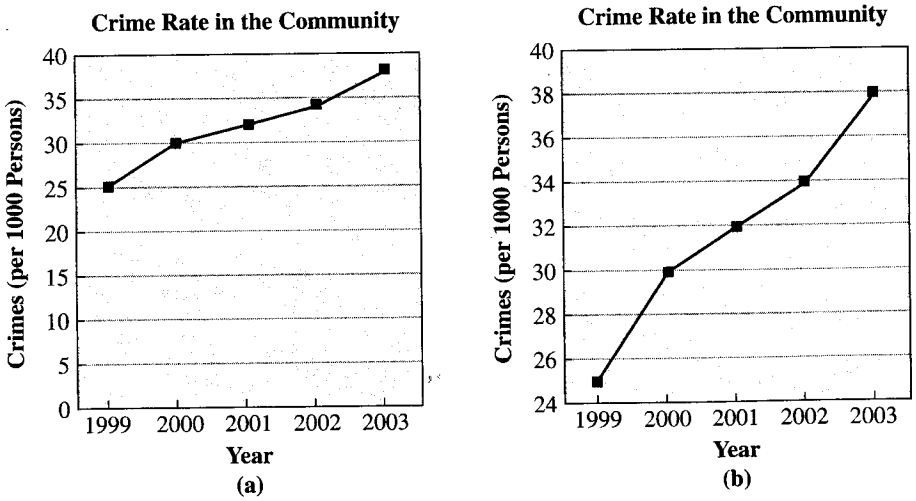


Figure 8.37

The graph in Figure 8.37(a) suggests that the rate of crime is growing slowly, whereas the cropped graph in Figure 8.37(b) gives the impression that the crime rate is rising more rapidly. ■

Another example of cropping in line graphs is shown in Figure 8.38. It shows the price of one share of a particular stock from April 25 through May 5.

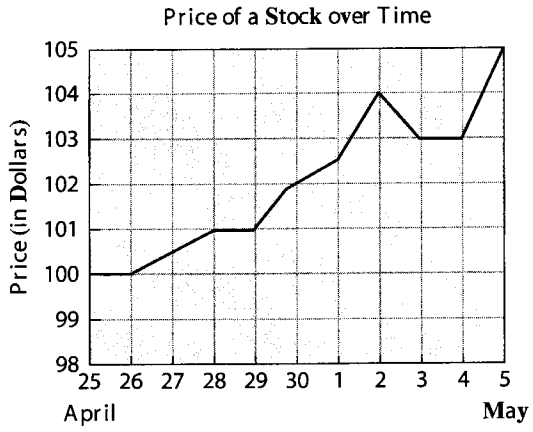


Figure 8.38

The stock appears to be a good buy because its price is increasing. Notice that the graph rises to the very upper edge of the vertical scale. This feature of the graph makes the upward trend appear more dramatic.

Figure 8.39 shows a different line graph of the price of the stock. The horizontal axis now shows the price of the stock over the previous 5 months, with the stock price plotted every 5 days. Notice that the vertical axis has also been changed. It now shows stock prices ranging from 0 to 140 dollars.

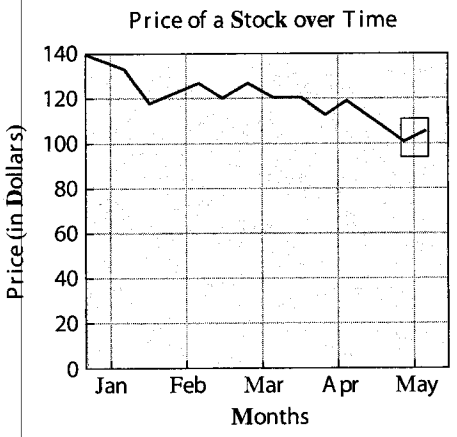


Figure 8.39

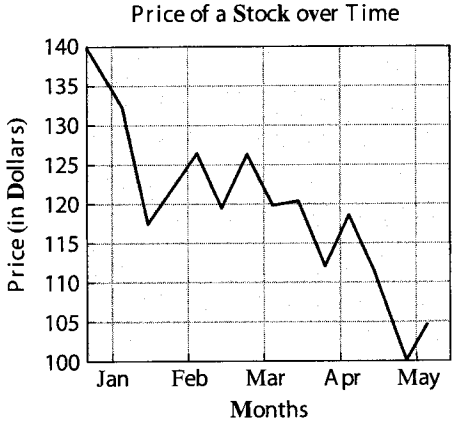


Figure 8.40

The data from Figure 8.38 are now contained in the small box of Figure 8.39, Thus, this latter graph gives a different perspective of the value of the stock. This different perspective is caused by the change in scales. The downward trend in Figure 8.39 would be even more apparent if we choose the vertical scale to be between 100 and 140. as shown in Figure 8.40.

Notice how changing the vertical axis gives a different impression of the trend in the stock price. In the graph shown in Figure 8.40, the dips in stock price appear much larger, and the increase in price during late April and early May (as seen earlier in Figure 8.38) seems less significant compared with the much greater decreases earlier in the year.

THREE-DIMENSIONAL EFFECTS

Three-dimensional effects, which newspapers and magazines often feature, may make a graph more attractive but can also obscure a true picture of the data. These types of graphs are difficult to draw by hand but are easily created using computer graphing software.

The data for the profits of a company, as shown in Figure 8.32, are displayed in a bar graph with three-dimensional effects in Figure 8.41.

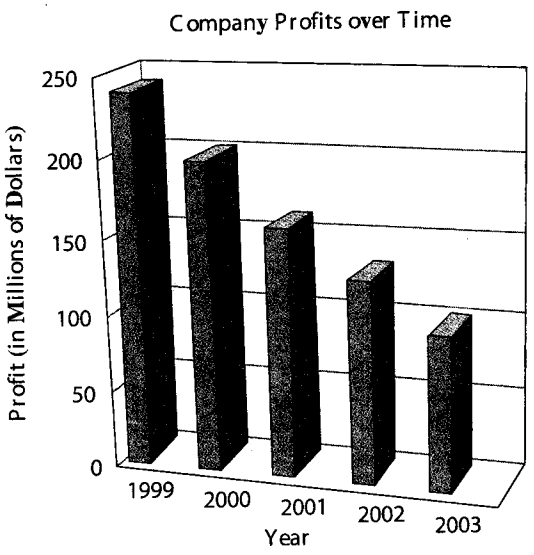


Figure 8.41

The perspective of a 3-D graph sometimes makes it difficult to see exact values. For example, the profits in 2003 were about $100,000, but a quick glance at the graph might lead a reader to estimate the profits as low as $80,000.

Line charts with three-dimensional effects may also obstruct a reader's view of critical information, as shown in the 3-D line graph in Figure 8.42.

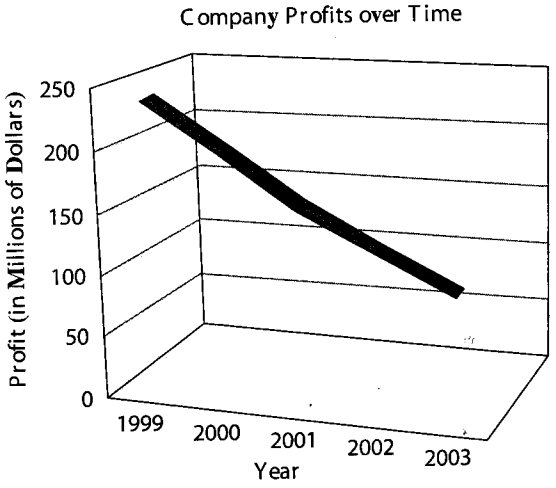


Figure 8.42

This is a graph of the same data shown in Figure 8.41. The downward trend in the profit is still apparent, but the exact dollar value of the profit for a given year is very difficult to read.

We can also manipulate pie charts to reinforce a particular message or even to mislead. Illustrators sometimes "explode" one sector of the pie, that is, move it slightly away from the center, as shown in Figure 8.43.

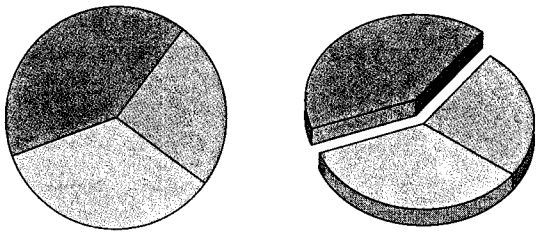


Figure 8.43

This version of the pie chart gives the exploded sector more emphasis. The exploded sector is actually about 50% larger than each of the others. Making the pie chart three-dimensional and exploding the sector makes that sector seem even larger.

PICTOGRAPHS

Tidbit

Archeologists believe that the earliest forms of pictographs are even older than numbers. Early accountants kept track of grain and agricultural supplies by making a symbol in clay that was a picture representing a unit of grain. Eventually, the accountants realized that it was more economical to make one picture for the grain and another symbol to tell how many units there were. So numbers were invented.

A **pictograph** is a type of graph in which pictures, symbols, or icons represent quantities. Newspapers, magazines, encyclopedias, and textbooks often use pictographs because of their visual appeal. One of the earliest forms of a pictograph was a horizontal or vertical bar chart in which icons rather than bars were used to represent numerical quantities, such as population, bushels or pounds of produce, barrels of oil, etc. Although pictographs can represent data in an interesting way, they can also be misleading.

The graph in Figure 8.44 is a pictograph displaying world population in 2000 and projecting world population for coming years.

Figure 8.44

The graph indicates that world population in 2000 was approximately 6 billion people and, according to this estimate, the population is expected to reach approximately 9 billion by the year 2045. Each person icon in the graph represents 1 billion people. In Figure 8.44, world populations were rounded to the nearest billion. However, fractions of 1 billion persons could have been represented by a portion of an icon. The graph implies that although the world population will continue to increase, the rate of increase will slow down.

The pictograph in Figure 8.45 represents data from a 2001 Coleman Happy Camper Report, which summarized campers' responses to a survey about what they most enjoyed about camping.

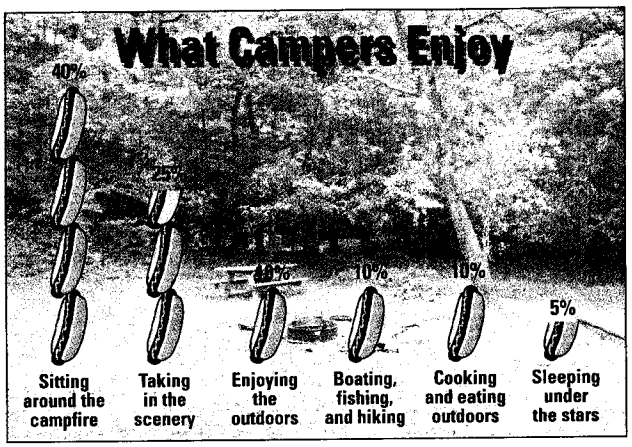


Figure 8.45

In this pictograph, hot dogs represent percentages of campers who rated various activities as their favorite. Notice that one hot dog icon represents 10%. The graph shows that four times as many campers said they most enjoyed sitting around the campfire as said their favorite activity was cooking and eating outdoors. Rather than rounding to the "nearest hot dog," this pictograph uses half a hot dog to indicate that 5% of those who responded to Coleman's survey said their favorite part of camping was sleeping under the stars.

Pictures can be used to embellish graphs in a variety of ways to make the graphs more interesting and to provide context. Sometimes the bars of a bar graph are stylized to fit the data being represented. For example, the graph in Figure 8.46 shows the kinds of gifts that shoppers intended to buy during the 2004 holiday season.

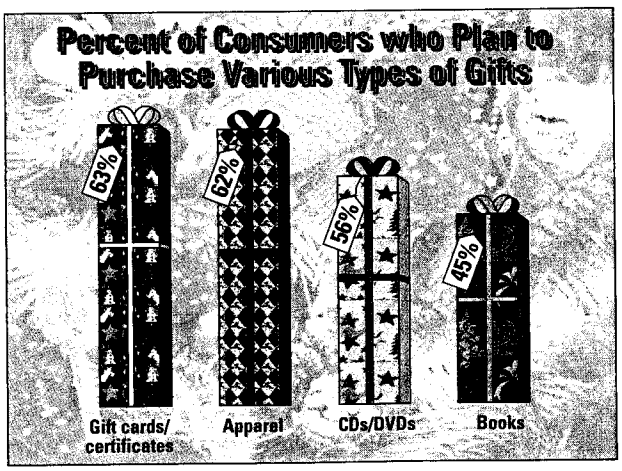


Figure 8.46

In this case, the bars of the graph are drawn as colorfully wrapped gifts. The graph makes it easy to see that gift cards and gift certificates were highest on shoppers' lists, with apparel a close second. Notice that the graph in Figure 8.46 has no scale, which

is a common omission for graphs of this type. If no scale is indicated, the relative lengths or heights of the bars are intended to show the relationship between the categories. For the graph to be accurate, the lengths or heights of the bars should be proportional to the amounts they represent. For instance, in Figure 8.46, because about 45% of shoppers plan to purchase books, compared to about 56% who plan to purchase CDs or DVDs, the bar that represents books is about 80% as tall as the bar representing CDs/DVDs. We can check the accuracy of the graph more carefully by actually measuring the heights of the bars and verifying that they are proportional to the percentages they represent.

The graph shown in Figure 8.46 is an example of a valid way to represent the data and illustrate differences in consumer's gift-giving intentions. Embellishing a graph with pictures can lead to confusion, however, and sometimes pictographs can be downright deceptive. A few examples of misleading graphs follow.

The graph in Figure 8.47 shows a breakdown of the 74.9 million U.S. students of all ages in 2003 according to the U.S. Census Bureau.

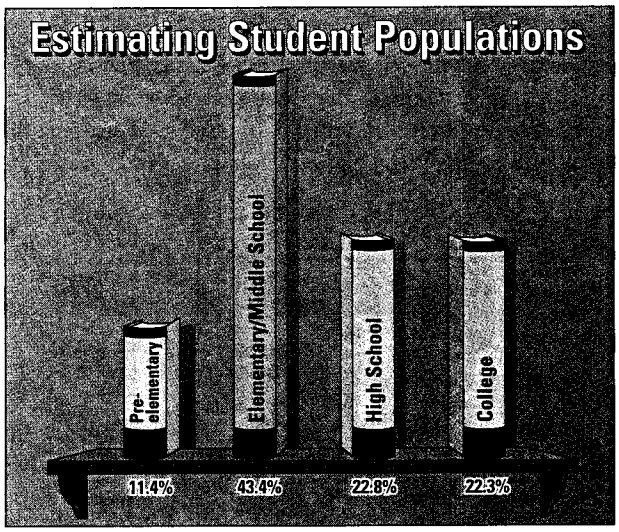


Figure 8.47

Given that in 2003, approximately 43.4% of students were in elementary or middle school and 11.4% were in pre-elementary school (pre-school and kindergarten), the length of the bar representing elementary/middle school students should be nearly four times as tall as the bar for the pre-elementary level. In the graph shown, however, the bar for elementary/middle school is only about 3 times as tall as the bar for pre-elementary school. (You can verify this by measuring the heights of the books in the graph and dividing the height of the bar for elementary/middle school by the height of the bar for pre-elementary school). Thus, the graph implies a smaller difference between the percentage of students in elementary/middle school and those in pre-elementary school.

Notice that if we consider only the *colored* portions of the book bindings, ignoring the brown portions, we find that the heights are, in fact, proportional. By adjusting only the colored portions of the book bindings, the creator of the graph caused it to be misleading.

The graph in Figure 8.48 shows the percentage of downhillers in various age categories who consider themselves snowboarders and indicates that the percentage decreases with the age of the downhiller.

Figure 8.48

Each bar in the graph is a snowboard, the length of which should be proportional to the percentage it represents. This graph is misleading for several reasons. First, the lengths of the bars are not proportional. The ratio of downhillers aged 25–34 who snowboard to downhillers aged 35–44 who snowboard is $\frac{49\%}{27\%} \approx 1.8$. If the graph is accurate, then the ratio of the lengths of the bars should also be about 1.8. However, the bar corresponding to ages 25–34 is about twice as long as the bar for ages 35–44. Thus, the graph makes it appear that about twice as many downhillers in the middle age category (25–34) call themselves snowboarders as downhillers aged 35–44, which slightly overstates the difference between the two groups.

Another feature of the graph in Figure 8.48 further exaggerates the difference between the percentages. Notice that the snowboards are angled. This design feature emphasizes the lengths of the top bars and makes the graph look top-heavy. It also makes the lowest bar appear even shorter than it really is.

In a pictograph, objects, either two-dimensional or three-dimensional, represent quantities. Consider the pictograph of milk cartons showing the increased sales in milk from 1997 to 2003 (Figure 8.49). (The numbers are for illustrative purposes; they do not reflect actual milk sales.)

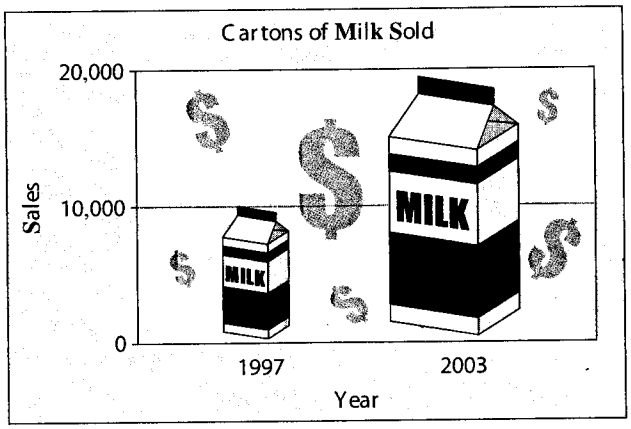


Figure 8.49

The amount of milk sold in 2003 was about twice the amount sold in 1997. At first glance, it might seem appropriate that the second carton is twice as tall as the other. However, looking at the pictures of the two cartons, we get the impression that the taller one has much more than twice the volume of the other. In addition to being twice as tall as the smaller carton, the larger carton is also twice as wide and twice as deep as the smaller one. Thus, the carton on the right represents a volume that is $2 \times 2 \times 2 = 8$ times as large as the volume of the carton on the left.

EXAMPLE 8.13 According to a 2002 Harris Interactive study, college students spend an average of about \$287 per month on things other than essentials such as tuition, books, room, and board. A significant part of that spending is for snacks and beverages. The pictograph in Figure 8.50 shows what college students spend on various types of drinks. How might this graph be misleading?

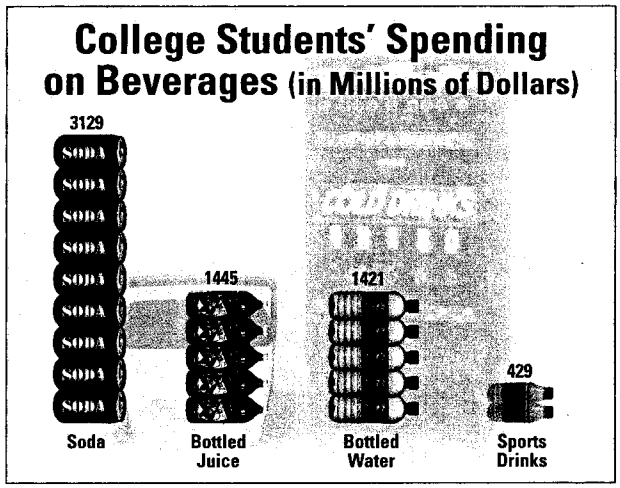


Figure 8.50

SOLUTION The dollar amounts that students spent on different types of beverages are represented by stacks of bottles. Consider the stack that represents money spent on sports drinks. Because this stack represents \$429 million, it would be reasonable to assume that each bottle represents $\frac{\$429}{2} = \$214.5$ million spent. If we focus on the number of bottles in each stack, the graph would indicate that college students spent an average of about $5(\$214.5) = \$1072.5$ million on bottled water and $9(\$214.5) = \$1930.5$ million on soda in 2002. However, these dollar amounts do not match the numbers atop the stacks.

On the other hand, if we ignore the *number* of bottles in each stack and focus instead on the *heights* of the stacks, we find the heights are roughly proportional to the dollar amounts atop the stacks. Seen this way, the graph gives a more accurate representation of the relative dollar amounts spent. This graph is misleading because a reader would not immediately know how to interpret the stacks of bottles, by height or by the number of bottles. That is, should this graph be viewed as a pictograph or a bar graph? ■

PIE CHARTS

We may create a **customized pie chart** to make it more appealing by embedding it in a picture or by adding context. The graph in Figure 8.51 provides information about the habits of 100 million coffee drinkers in the U.S.

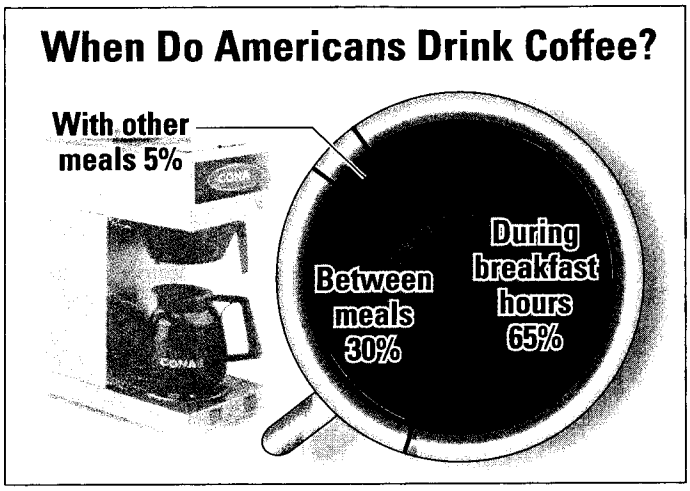


Figure 8.51

Using the top of a coffee cup as the circle for the pie chart makes the graph more eye-catching and helps to put the percentages in context. According to a 2004 survey conducted by the National Coffee Association, the vast majority of coffee (65%) is consumed in the morning, and very little coffee is drunk with lunch or dinner.

We can create distortions in a pie chart. Figure 8.52 displays a CD made into a pie chart that represents teenagers' opinions about downloading music from the Internet. The design of the graph gives the impression of far more than 54% of teenagers see nothing wrong with it.

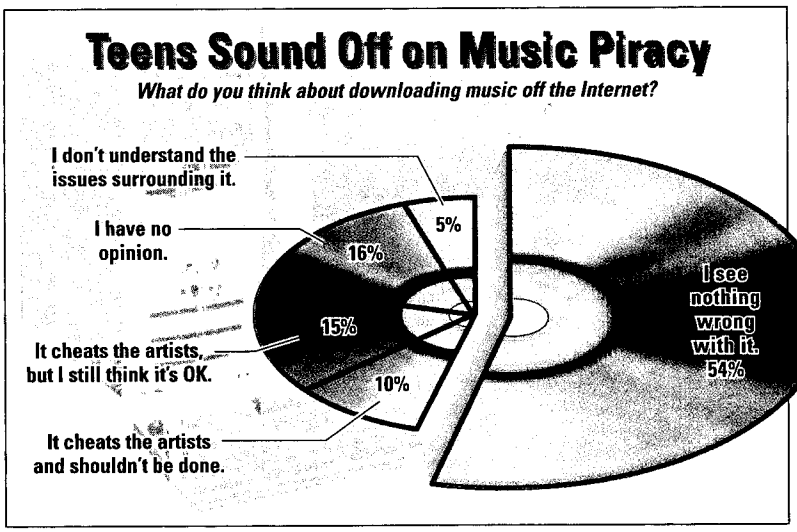
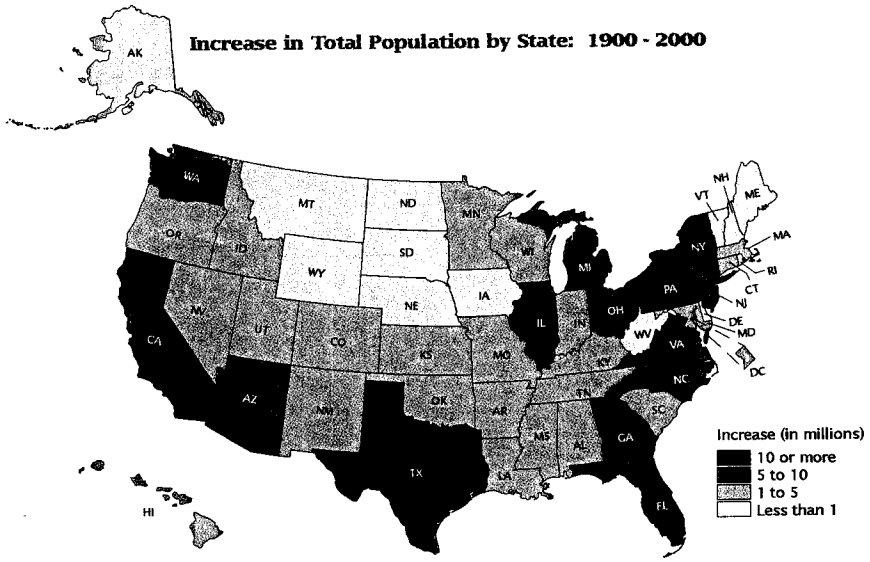


Figure 8.52

GRAPHICAL MAPS

Maps can summarize information about geographical areas and can show patterns related to national or world concerns. Figure 8.53 shows a **graphical map** indicating the increase in total U.S. population from 1900 to 2000 by state, according to the U.S. Census Bureau.

The areas of the shaded regions are not proportional to the population increases during the 100-year period. Instead, the various shades of blue indicate the population increases. By coincidence, some of the larger states like California and Texas did experience the greatest increases in population. Florida and New York, which are relatively small states, also experienced increases of more than 10 million persons.



Source: U.S. Census Bureau, decennial census of population, 1900 and 2000.

Figure 8.53

Source: www.census.gov/prod/2002pubs/censr-4.pdf.

Another graphical map you may see nearly every day is a national weather map. One example is shown in Figure 8.54, which gives a national weather map for Friday, November 28, 2003. At a glance, you can tell the expected weather in any part of the country for that day. A table giving such data would be less informative and more difficult to interpret than this pictorial representation.

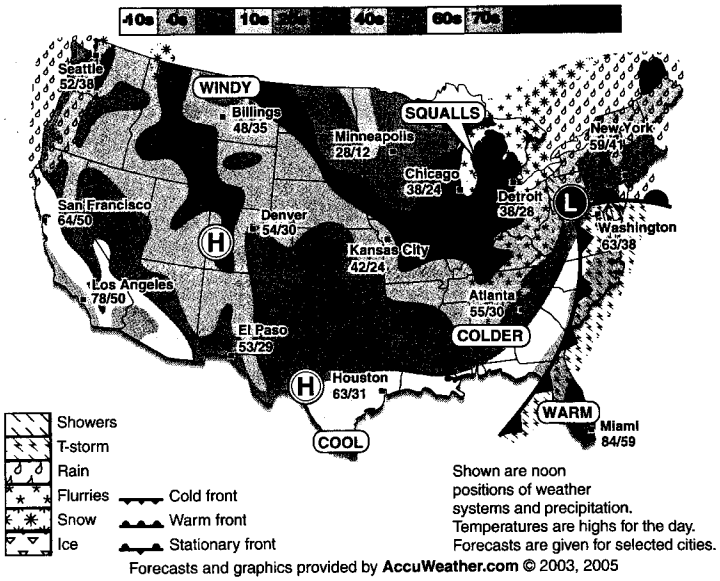
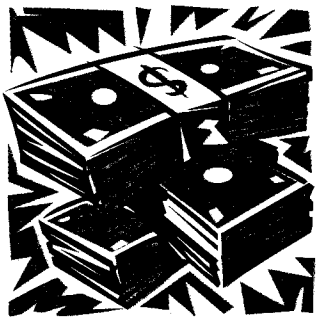


Figure 8.54

Source: Salem Statesman Journal, November 28, 2003. Permission courtesy of AccuWeather.com

SOLUTION OF THE INITIAL PROBLEM



Suppose that you are on a debating team and know you will argue the topic, "Resolved, the most important economic issue facing the country today is the federal debt." You do not know which side of the issue you will have to argue. As part of your preparation for the debate, you decide to make two graphs, each of which illustrates the federal debt over time. The graphs will show the federal debt from two different perspectives. One graph will depict the debt in the most unfavorable light possible and the other will show the debt in a more optimistic manner. Make two such graphs using the federal debt data from 1965 to 2000 shown in the following table.

Year	Federal Debt (to the Nearest Billion Dollars)
1965	321
1970	389
1975	577
1980	930
1985	1946
1990	3233
1995	4974
2000	5674

Source: www.publicdebt.treas.gov/opd/opd.htm.

SOLUTION To make the national debt appear as serious as possible, we could plot the amount of the debt over the years, emphasizing its upward trend. To do this, we might create a line graph, using horizontal and vertical scales that result in a graph that appears as a tall thin rectangle. We might even make the top of the curve go over the top of the scale, as shown in Figure 8.55.

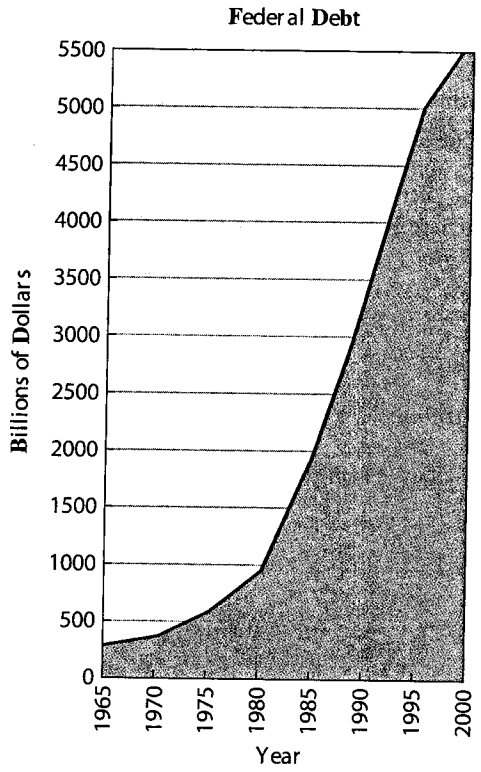


Figure 8.55

To make the problem of the national debt appear less serious, we might plot a related quantity, such as the annual percentage rate of increase instead of the actual debt amounts. The annual *rates* of increase do not change nearly as much as the actual amount of the federal debt, and they may even go down when debt goes up. One possible graph of the variation in the annual rates of increase is shown in Figure 8.56.

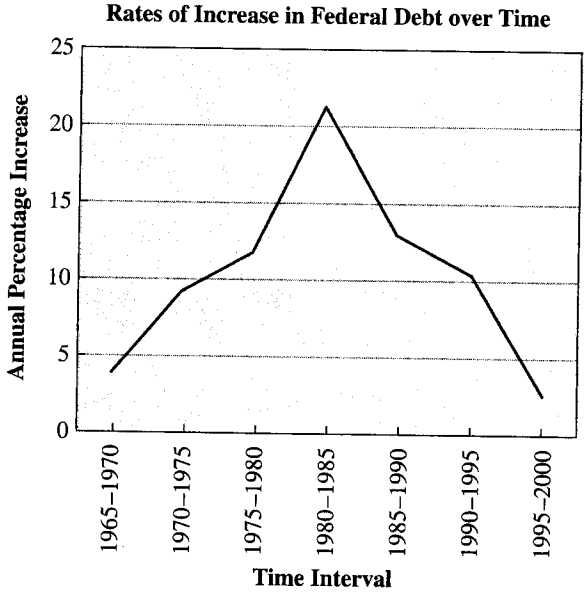
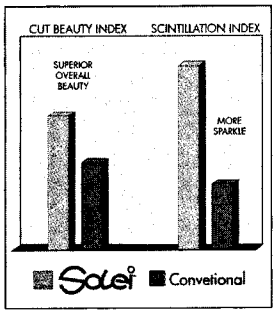


Figure 8.56

Notice how this graph presents a much more positive view of the federal debt than the graph in Figure 8.55.

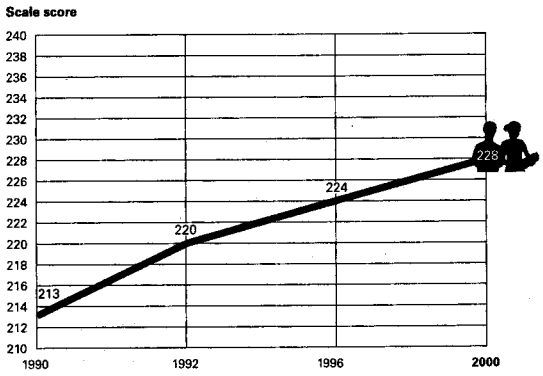
PROBLEM SET 8.3

1. A jewelry catalog advertising Solei diamonds displayed the following bar graph to demonstrate that Solei diamonds have superior overall beauty and more sparkle. Explain what is misleading about the graph.



Source: Permission courtesy of VisionCut™, New York, NY.

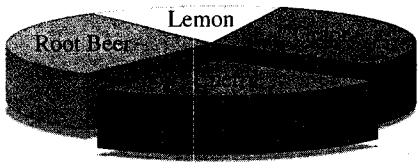
2. The following line graph is from the 2002 Brown Center Report published by the Brookings Institute. It shows fourth-grade math scores from 1990 through 2000. Explain what is misleading about the graph.



Source: From 2002 Brown Center Report on American Education by Tom Loveless. The Brookings Institution Press, Washington, D.C.

3. Suppose a beverage company conducts a taste test and includes the following pie chart in its advertising.

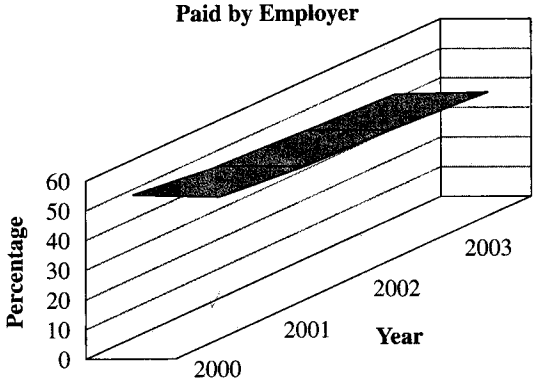
Soda Preferences



   a. Which type of soda do you think the company produces?
   b. In the actual taste test, 15% of consumers preferred lemon, 25% preferred cherry, and equal numbers preferred grape and root beer. The sectors in the pie chart are drawn using the correct degree measures. Explain what makes this pie chart misleading.

4. Suppose that during contract negotiations, management presents the workers with the following graph showing the percentage of health-care premiums paid by the employer.

Percentage of Health-Care Premium Paid by Employer

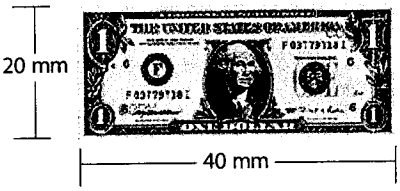


   a. Why do you think management presented the graph in this way?
   b. Estimate the percentages for each of the 4 years given in the graph and describe the trend. Explain what makes this graph misleading.

5. To see whether bar heights in a bar graph are proportional to the numerical values they represent, create ratios of bar heights, compare them to ratios of the corresponding quantities, and see if they are equal. For each case below, determine whether the bar heights and the numeric values they represent are proportional.
   a. Bar 1: height, 1 cm; value, $700
      Bar 2: height, 8.5 cm; value, $5950
   b. Bar 1: height, 1.3 inches; value, 52%
      Bar 2: height, 0.4 inch; value, 15%
   c. Bar 1: height, 3 mm; value, 1250 people
      Bar 2: height, 27 mm; value, 11,250 people
      Bar 3: height, 35 mm; value, 14,950 people

6. To see whether bar heights in a bar graph are proportional to the numerical values they represent, create ratios of bar heights, compare them to ratios of the corresponding quantities, and see if they are equal. For each case below, determine whether the bar heights and the numeric values they represent are proportional.
   a. Bar 1: height, 5 cm; value, $8000
      Bar 2: height, 5.2 cm; value, $8320
   b. Bar 1: height, 0.6 inch; value, 22%
      Bar 2: height, 2.2 inches; value, 91%
   c. Bar 1: height, 18 mm; value, 9000 people
      Bar 2: height, 15 mm; value, 7500 people
      Bar 3: height, 2.5 mm; value, 1250 people

7. In a pictograph, pictures, symbols, or icons represent quantities. Suppose the area of a dollar bill with length 40 mm and width 20 mm will represent the total cost of an item in the year 2001.



= Total Cost in 2001

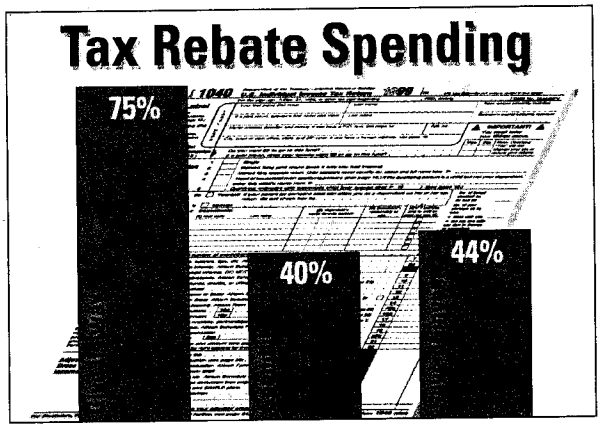
   a. In the year 2002, suppose that the total cost for the item tripled. Represent the total cost of the item in 2002 in a way that is not misleading without changing the size of the dollar bill.
   b. Suppose that in the year 2001, an item cost $5000, and in the year 2002, the same item cost $4000. Represent the total cost of the item in 2002 in a way that is not misleading by changing only the length of the dollar bill. What is the length of the dollar bill that must be used for the year 2002?
   c. Suppose that in the year 2002, the total cost for the item doubled. Represent the total cost of the item in 2002 in a way that is not misleading by changing both the length and the width of the dollar bill. In the original dollar bill, the length was twice the width. Maintain that relationship when creating the dollar bill for the year 2002. Find the length and width of the dollar bill, rounded to the nearest tenth of a millimeter, to use for the year 2002.

8. In a pictograph, pictures, symbols, or icons represent quantities. Suppose the area of a coin with diameter 20 mm will represent the total cost of an item in the year 2001.

= Total Cost in 2001

   a. In the year 2002, suppose that the total cost for the item quadrupled. Represent the total cost of the item in 2002 in a way that is not misleading without changing the size of the coin.
   b. Suppose that in the year 2001, an item cost $0.50, and in the year 2002, the same item cost $0.75. Represent the total cost of the item in 2002 in a way that is not misleading without changing the size of the coin.
   c. Suppose that in the year 2002, the total cost for the item doubled. Represent this in a way that is not misleading by changing the diameter of the coin. In the original coin, the diameter was 20 mm. Find the diameter of the coin, rounded to the nearest tenth of a millimeter, to be used for the year 2002.

9. In 2001, most Americans received a rebate check from the federal government. The following graph shows the percentage of Americans who said they would use their rebate money in a certain way.



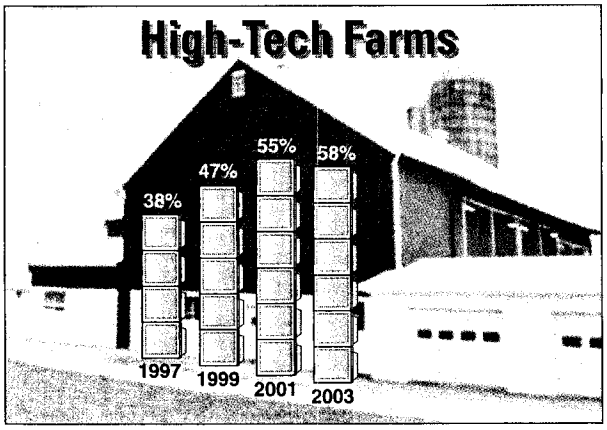
   a. Measure each bar height to the nearest tenth of a centimeter, and determine if the bar heights are proportional to the percentages they represent.
   b. If the bar that represents the 40% of Americans who would spend their rebate money was drawn 5 cm tall, find the heights that should be used for the other two bars so that bar heights are proportional to the percentages they represent. Round heights to the nearest tenth of a centimeter.

10. WALTHAM, a leading authority on pet care and nutrition, conducted a survey and found that pet owners humanize pets in several ways. The following graph summarizes some of the responses of pet owners from Los Angeles, San Francisco, Washington D.C., and Atlanta.

a. Measure each bar (dog collar) to the nearest tenth of a centimeter and determine if the bar lengths are proportional to the percentages they represent.

b. If the bar that represents the 28% of respondents who think of pets 5-10 times a day was drawn 3 cm long, find the lengths that should be used for the other two bars so that bar lengths are proportional to the percentages they represent. Round lengths to the nearest tenth of a centimeter.

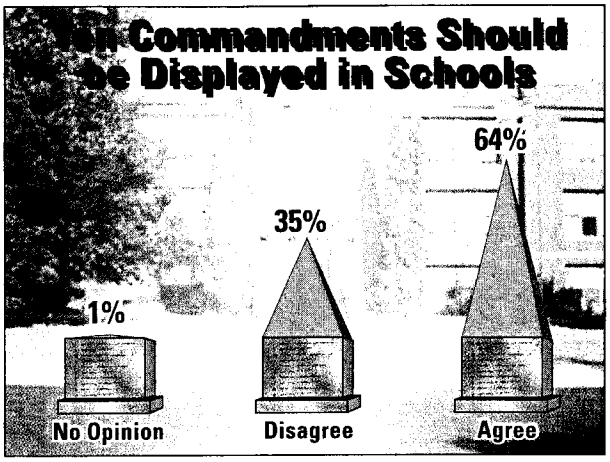
11. Farmers increasingly have access to computers in their businesses. The following pictograph gives the percentage of farms that had access to computers from 1997 to 2003.



a. Each computer monitor represents how many percentage points in the 1997 bar? Answer the same question for each of the other three bars. What do you notice?

b. Summarize what is misleading about this graph.

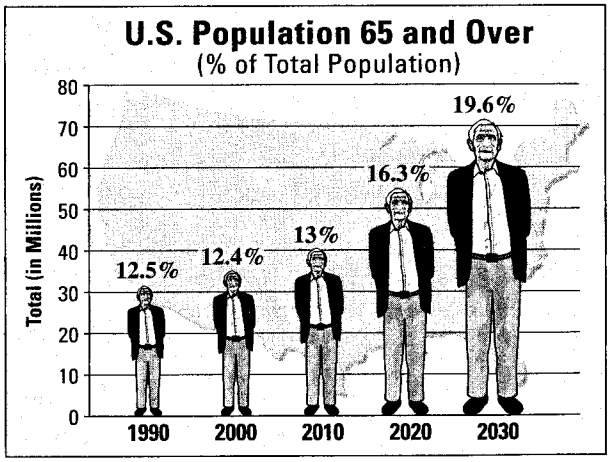
12. In a May 2005 free-speech survey, the majority of Americans polled supported displaying the Ten Commandments in schools as shown in the following graph.



a. What part of each monument is used to represent the results of the survey, and are the heights proportional to the percentages they represent?

b. Summarize what is misleading about this graph.

13. The U.S. population 65 years old and over, in millions, is given in the following graph for the years 1990 and 2000 and is predicted for the years 2010, 2020, and 2030.



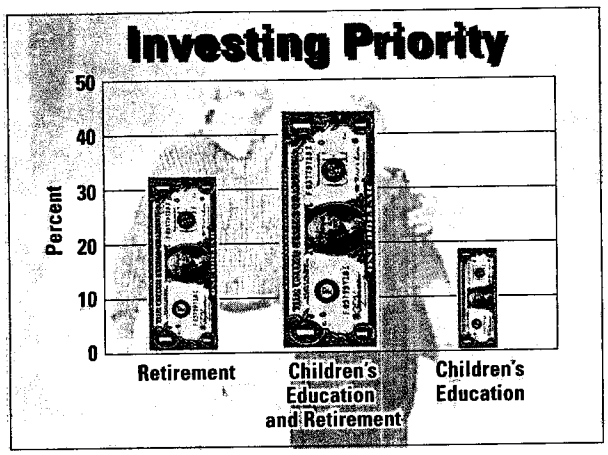
a. What does the height of each figure represent?

b. Approximately how many millions of people in the U.S. were 65 years old or over in 1990? Approximately how many millions are predicted to be 65 years old or over in 2030? The predicted 2030 value is how many times taller than the 1990 value?

c. The area of the 2030 figure is how many times larger than the area of the 1990 figure? Is the area of a figure meaningful in this graph?

d. Summarize what is misleading about this graph.

14. In a July 2005 Allstate survey, people with children were asked if investing in retirement is a higher priority, investing in children's education and retirement is of equal importance, or investing in children's education is a higher priority. The following graph summarizes the results.



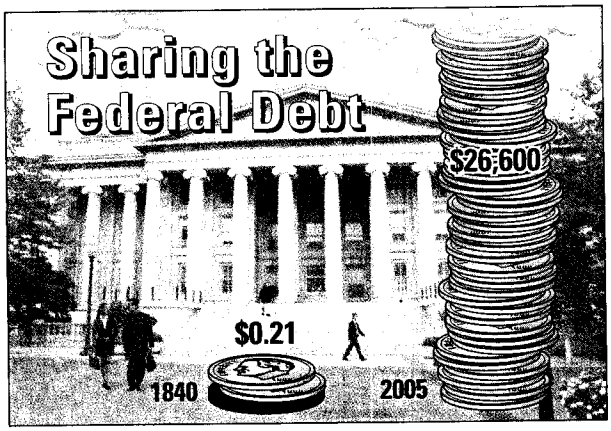
a. What does the height of each dollar represent?

b. Approximately what percent of respondents felt investing in their children's education was a higher priority? Approximately what percent of respondents felt investing in their children's education and retirement was of equal importance?

c. The area of the dollar that represents those who felt investing in their children's education and retirement was of equal importance is approximately how many times greater than the area of the dollar that represents those who felt investing in their children's education was a higher priority?

d. Summarize what is misleading about this graph.

15. In 1840, the cost of the federal debt was $0.21 per person. The cost grew to $26,600 per person in 2005, as shown in the following graph.

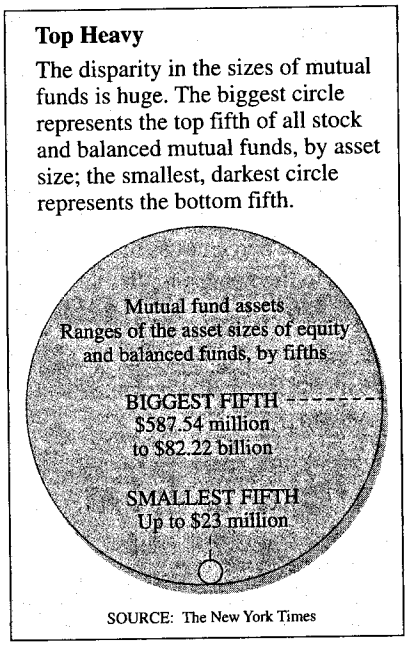


a. There are two dimes and a penny in the stack of coins that accurately represents the federal debt per person in 1840. If stack that represents the federal debt per person in 2005 contains only dimes, then how many dimes should be in the stack?

b. Suppose the stack of coins that represents the federal debt per person in 1840 was drawn 0.5 cm tall. How tall should the stack of coins that represents the federal debt per person in 2005 be if stack heights are proportional to the amounts they represent? Round to the nearest centimeter.

c. Summarize what is misleading about this graph.

16. Small, independent mutual funds attract investors even though there is no guarantee that investors will make more money with small funds rather than large funds. The following graph illustrates the relationship between the mutual funds with large asset size and those with small asset size.



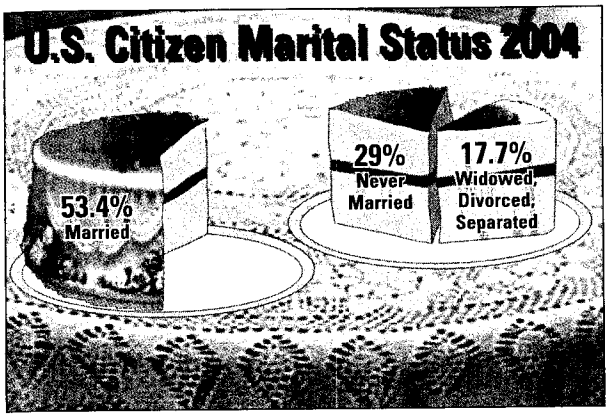
Source: Adapted from The New York Times

a. If the area of the largest circle represents the top fifth of all stock and balanced mutual funds with assets of $82.22 billion, and the circle used has a radius of 2 inches, what must the radius be for the circle that represents the smallest fifth, with assets of $23 million, assuming that area is proportional to assets?

b. If the area of the smallest circle is used to represent assets of $23 million dollars and measures 2 mm in radius, what must the radius be for the largest circle if it represents assets of $82.22 billion, assuming that area is proportional to assets?

c. Summarize what is misleading about this graph.

17. The U.S. Census Bureau classifies citizens who are at least 15 years old according to marital status. The following graph gives the percentages in each of three categories in 2004.



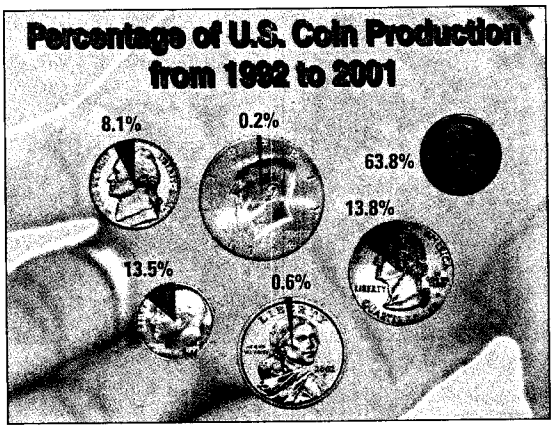
a. If you would create a single pie chart using the percentages given in the graph, what degree measure would be associated with each sector? Round to the nearest degree.

b. Using a protractor, measure the angles of each sector in the graph and compare the angles to the degree measures you found in part (a). How do they compare?

c. Summarize what is misleading about this graph.

d. Create a single pie chart using the degree measures you found in part (a).

18. There are six coins currently in circulation in the United States: penny, nickel, dime, quarter, half dollar, and dollar. The percentage of the U.S. coin production for each of these circulating coins is given in the following graph.



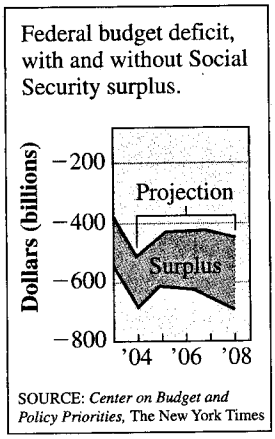
a. If you would create a single pie chart using the percentages given in the graph, what degree measure would be associated with each sector? Round to the nearest degree.

b. Using a protractor, measure the angles of each sector in the graph and compare the angles to the degree measures you found in part (a). How do they compare?

c. Summarize what is misleading about this graph.

d. Create a single pie chart using the degree measures you found in part (a).

19. The following double-line graph shows the actual federal budget deficit, with or without the Social Security surplus, along with projections for the next few years. The federal budget deficit without the Social Security surplus is represented by the lower line. The federal budget deficit with the Social Security surplus is represented by the upper line.



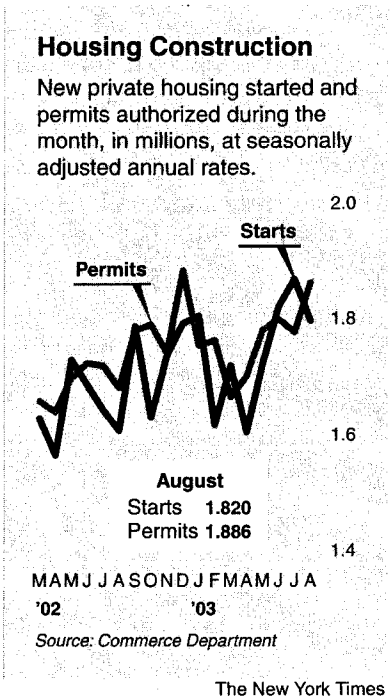
a. At first glance, what overall impression does the graph give?

b. How is the vertical scale labeled? Explain how this graph is misleading.

c. Estimate the value of the federal budget deficit with and without the Social Security surplus for the years from 2003 to 2008, and use them to create a new double-line graph, but begin the vertical scale at zero. Plot deficit values above the horizontal axis, using positive values along the vertical axis.

d. For the graph constructed in part (c), describe the trend in the federal budget deficit with or without the Social Security surplus.

20. The following double-line graph is designed to show the trends in private housing permits issued and actual housing starts from May 2002 to October 2003.



http://www.nytimages.com/portal/wieck_preview_page_116740.

a. At first glance, what overall impression does the graph give?

b. Explain how this graph is misleading.

c. Estimate permit values and housing-start values from the graph, and use them to create a new double-line graph, but begin the vertical scale at zero and expand the horizontal scale.

d. For the graph constructed in part (c), describe the overall impression the graph gives in private housing permits and new housing starts.

21. The dropout rates for grades 9 through 12 for Georgia and Nevada are given in the following table. The dropout rate for a school is found by dividing the number of dropouts by the total number of students enrolled in the school at the beginning of the school year, writing the result as a percentage.

a. Create a double-line graph to represent the data, and choose a vertical scale that downplays the differences in dropout rates for the two states.

b. Create a double-line graph to represent the data, and choose a vertical scale that emphasizes the differences in dropout rates for the two states.

State	'95–'96	'96–'97	'97–'98	'98–'99	'99–'00	'00–'01
Georgia	8.5	8.2	7.3	7.4	7.2	7.2
Nevada	9.6	10.2	10.1	7.9	6.2	5.2

Source: http://www.necs.ed.gov.

22. The following table gives the results of a survey in which people of different ethnic groups were asked whether they suffered from some common conditions in the past six months.

Ailment	Hispanic	Non-Hispanic
Cough, cold, flu	59%	24%
Heartburn	49%	14%
Frequent headaches	33%	13%
Back, neck, joint pain	45%	29%

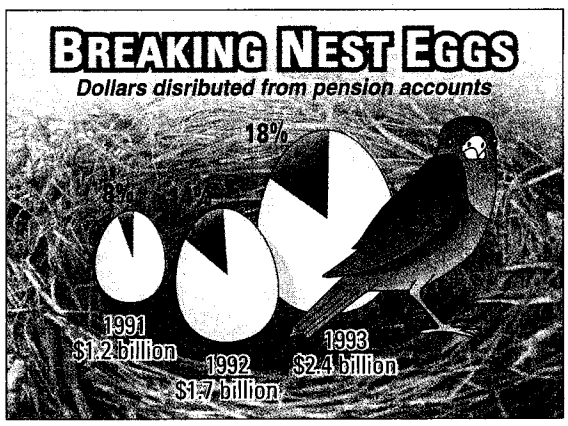
Source: http://www.acnielsen.com.

a. Create a double-bar graph to represent the data, and choose a vertical scale that downplays the differences in percentages.

b. Create a double-bar graph to represent the data, and choose a vertical scale that emphasizes the differences in percentages.

Problems 23 and 24

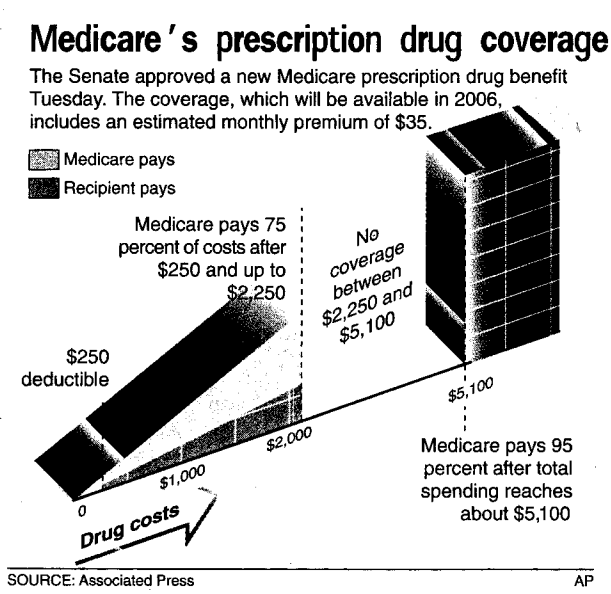
In the following pictograph, the ovals that represent the "nest eggs" have heights that are proportional to the total amounts in the pension accounts. For example, notice that the 1991 egg represents $1.2 billion and is half as tall as the 1993 egg, which represents $2.4 billion. The percentages given represent the percentages of the total pension funds that are distributed in that year.



23. Comment about how the use of an egg is misleading in the pictograph provided. Create a set of three pie charts based on the data from the pictograph. Make all the circles the same size. How does making the circles the same size affect a reader's impression of the amounts involved?

24. Comment about how the use of the tree and leaves is misleading in the pictograph provided. Create a set of three pie charts based on the data from the pictograph. Make the area of each circle proportional to the amount in the pension fund; that is, the area of the circle for 1993 should be twice the area of the circle for 1991. How does this representation affect a reader's impression of the amounts involved?

25. The U.S. Senate approved a new Medicare prescription-drug benefit on November 25, 2003. Consider the following misleading three-dimensional graph intended to explain the new system. It compares the percentage of drug costs covered by Medicare to the percentage that must be paid by the senior citizen as drug costs increase.



AP/Wide World Photos

a. What quantity is represented by the horizontal axis of the graph?

b. What do the darker shaded region and lighter shaded region represent? Consider the region between drug costs of $250 and $2000 and comment on the slant of the graph and the misleading features.

c. What does the thickness of the graph represent?

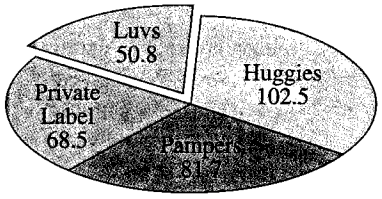
d. Notice that the portion of the graph between $0 and $250 is labeled as the deductible. A deductible is not paid by Medicare. What color should the portion of the graph between $0 and $250 be?

e. If there is "no coverage" between $2250 and $5100, as indicated, then who must pay and what color should that portion of the graph be?

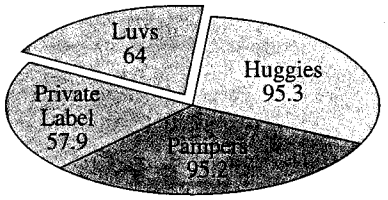
f. Create a proportional bar graph (see Section 8.2) using the information in this graph. Compare the proportional bar graph to the three-dimensional graph and comment on the differences in ease of interpretation.

26. Using perspective with pie charts can result in a graph that is deceptive. Consider the following examples, which compare sales of various brands of disposable diapers for two different years.

Diaper Sales
(Millions of Units)



52 weeks ending Dec 11, 1993



52 weeks ending Jun 13, 1992

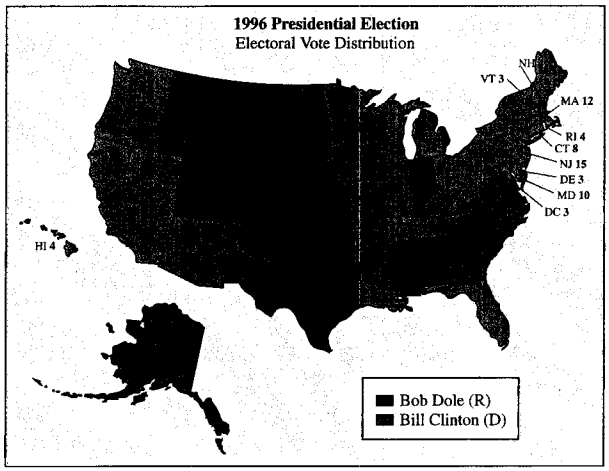
a. For each company, find the change in sales, in millions of units, from 1992 to 1993.

b. Notice how the sector for Luvs brand diapers is colored and raised up out of the pie charts. Suppose these pie charts were used in an investors' guide or in a company report for Luvs. Explain why the pie charts were presented in this manner and why they are misleading.

c. Construct a double-bar graph that emphasizes the differences in sales from 1992 to 1993.

d. How might investors in Luvs react to the double-bar graph from part (c)?

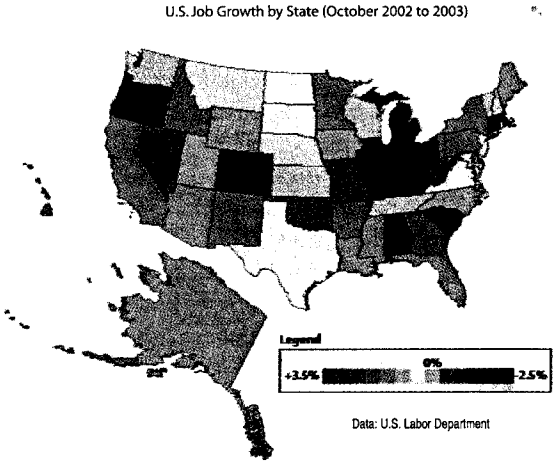
27. The following graphical map displays the electoral vote distribution of the 1996 Presidential election.



a. Based on land area alone, roughly estimate what fraction of the United States appears to have supported each candidate? What geographic regions of the country supported each candidate?

b. Bill Clinton received 379 electoral votes while Bob Dole received 159 electoral votes. Is area proportional to the1996 vote totals?

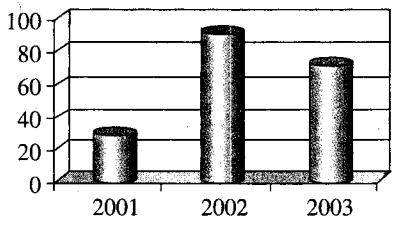
28. The following geographic map displays job growth for the United States for the year beginning October 2002.



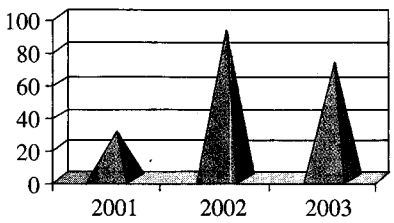
a. Based on land area alone, roughly estimate what fraction of the United States appears to have experienced zero or negative job growth?

b. What area(s) of the United States experienced the most dramatic job loss? What area(s) experienced the most dramatic job growth?

Three-Dimensional Graph Using Cylinders



Three-Dimensional Graph Using Pyramids



29. Three-dimensional graphs, such as the preceding examples, are much easier to create using a computer program such as Microsoft Excel than to do so by hand. Use the following guidelines to explore the three-dimensional graph options in Microsoft Excel.

a. Using the rebate percentages from problem 9, create a three-dimensional graph in Microsoft Excel. Discuss any aspects of this graph that might mislead a reader.

b. Using the pet percentages from problem 10, create a three-dimensional graph in Microsoft Excel. Discuss any aspects of this graph that might mislead a reader.

c. Using the percentages for computer use in farming from problem 11, create a three-dimensional graph in Microsoft Excel. Discuss any aspects of this graph that might mislead a reader.

d. Using the percentages for displaying the Ten Commandments in schools from problem 12, create a three-dimensional graph in Microsoft Excel. Discuss any aspects of this graph that might mislead a reader.

30. *USA Today* regularly includes graphical representations called "Snapshots" to emphasize statistics or survey results about money, life, sports, or news. These snapshots are colorful, eye-catching pictographs or bar charts designed to convey a message quickly. Go to the *USA Today* website at http://www.usatoday.com/news/snapshot.htm and look at the collection of snapshots. From them, select two examples of misleading graphs that each use different techniques discussed in this chapter. For each graph you select, summarize the features that are misleading.

31. Graphing in three dimensions is a common, attractive way to display data. However, as we have discussed, a three-dimensional bar graph can be misleading. A new method, called the "diamond graph" method, is designed to avoid the misleading aspects of the three-dimensional bar graph. This new type of graph was created by Dr. Alvaro Muñoz, who is a professor of epidemiology at the Johns Hopkins Bloomberg School of Public Health. What shape does the diamond graph use rather than parallel, three-dimensional bars? Why was this shape chosen? Find or create an example of this type of graph. Research the diamond graph by using search keywords "new diamond graph method" on the Internet. Summarize your findings in a report.

32. In 1861, a French engineer, Charles Minard, created a graphical presentation of Napoleon's Russian campaign of 1812. This display is considered by some to be the greatest statistical graphic ever created before the advent of computer graphics. Research the graph by Minard. What techniques were used to make the graph? What made the graph so effective? Was there anything misleading about the graph? One source of information on this and other graphs is Edward Tufte's book, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press). On the Internet, search keywords, "Charles Minard." Write a report of your findings about Minard's graph.

33. Compare the graphs used in different media publications such as the *New York Times,* the *Wall Street Journal, Time* magazine, or other diverse sources. From at least three different sources, find examples of misleading bar graphs, line graphs, three-dimensional graphs, pie charts, or pictographs. Which types of graphs seem to be used most commonly? How does the target audience of a publication influence the choice of the graphics used? What is the most common misleading feature of graphs used in the media? For each graph you selected, write a paragraph describing the misleading feature of the graph and explaining why a misleading graph might have been printed.

34. Research recent lawsuits over misleading advertising claims. Some of the companies that have been involved in misleading advertising lawsuits over the past few years are Aleve pain reliever, Claritin, Kaiser Permanente, Nike, Pepsi, Grey Goose Vodka, McDonald's, Burger King, and Kentucky Fried Chicken. Research claims against these companies or research lawsuits filed against other companies. What was considered misleading about the advertisement? Were any graphs involved? How was the lawsuit resolved? Write a report to summarize your findings.

CHAPTER 8 REVIEW

Key Ideas and Questions

The following questions review the main ideas of this chapter. Write your answers to the questions and then refer to the pages listed to make certain that you have mastered these ideas.

1. How is a dot plot created? pg. 478 What is an advantage of a stem-and-leaf plot over other types of graphs? pg. 479
2. How is a histogram similar to a stem-and-leaf plot? pg. 480 What is one drawback to using a histogram? pg. 480 How are a histogram and a relative frequency histogram related? pgs. 480–481 What factors should be considered when deciding on the intervals to use in a histogram? pgs. 481–482
3. How does a bar graph differ from a histogram? pg. 483 What do the heights or lengths of the bars in a bar graph represent? pg. 483
4. Which types of graphs show time trends? pgs. 485–486
5. For what type of data might you use a pie chart? pg. 487
6. What kinds of data would be used in the following types of graphs: stem-and-leaf plot, histogram, bar graph, line graph, pie chart? pg. 487
7. How is a double stem-and-leaf plot created? pgs. 503–504 What types of information can double-bar graphs conveniently illustrate? pg. 504 How can pie charts be used to show trends over time? pgs. 507–508
8. How is a proportional bar graph created? pg. 509
9. How can manipulating the axis in a graph alter the impression given by the graph? pgs. 528–531 How can cropping in a line graph alter the impression given by the graph? pgs. 532–534 How can three-dimensional effects obscure the true picture of the data? pgs. 534–536
10. What is a pictograph? pg. 536 How can pictographs be created to misrepresent the data? pgs. 538–540 How can pie charts be used to mislead? pg. 541
11. What is a graphical map? pgs. 542–543

Vocabulary

Following is a list of key vocabulary for this chapter. Mentally review each of these terms, write down the meaning of each one in your own words, and use it in a sentence. Then refer to the page number following each term to review any material that you are unsure of before solving the Chapter 8 Review Problems.

SECTION 8.1

Data Sets 477
Designed Experiment 477
Observational Study 477
Survey 477
Population 477
Exploratory Data Analysis 478
Dot Plot 478
Stem-and-Leaf Plot 479
Stem 479
Leaf 479
Cluster 479
Gap 479
Outlier 479
Histogram 480
Measurement Class 480
Bin 480
Frequency 480
Frequency Table 480
Relative Frequency Table 480
Relative Frequency 481
Relative Frequency Histogram 481
Bar Graphs/Bar Charts 483
Line Graphs 486
Pie Chart/Circle Graph 487

SECTION 8.2

Double Stem-and-Leaf Plots 504
Comparison Histogram 504
Multiple-Bar Graph/ Comparison Bar Graph 505
Double-Bar Graph 505
Multiple-Line Graph/ Comparison Line Graph 506
Double-Line Graphs 506
Multiple Pie Charts/ Comparison Pie Charts 507
Proportional Bar Graph 509

SECTION 8.3

Scaling and Axis Manipulation 528
Cropping 532
Three-Dimensional Effects 534
Pictographs 536
Customized Pie Charts 541
Graphical Map 542

1. The following stem-and-leaf plot gives the percentages of people who were at least 65 years old in the year 2000 in each of the 50 United States.

17	6
16	
15	3 6
14	0 3 4 5 7 9
13	0 0 0 1 2 2 3 3 3 4 5 5 6 8
12	0 0 1 1 1 1 3 4 4 5 7 8 9
11	0 2 2 3 3 6 7 7
10	6
9	6 7 9
8	5
7	
6	
5	7

Source: U.S. Census Bureau.

   a. Identify any clusters, gaps, and outliers in the data. Which state would you guess might have had the highest percentage of residents who were at least 65 years old?
   b. In what percentage of U.S. states was more than 10% of the population at least 65 years old? In what percentage of states was less than 14% of the population at least 65 years old?

2. Suppose that a survey reveals the starting salaries of a group of Master's of Business Administration (MBA) graduates who were hired for comparable positions in the communication industry. The starting salaries, rounded to the nearest $100, are as follows:

$42,500	$40,100	$41,100	$38,900	$43,200
$40,900	$42,200	$41,700	$43,800	$38,800
$40,800	$46,500	$40,600	$39,900	$39,100
$44,500	$40,100	$42,700	$41,200	$38,900

   a. Make a dot plot of these salaries.
   b. Make a stem-and-leaf plot of these salaries.
   c. Describe the salaries in terms of any clusters, gaps, and outliers.

3. Suppose that the study in problem 2 considers a different group of college graduates who have bachelor's degrees. Their starting salaries, after rounding to the nearest $100, were:

$39,300	$37,000	$37,900	$35,500	$38,300
$37,600	$38,100	$36,000	$41,400	$36,500
$38,100	$40,300	$37,300	$35,800	$35,200
$37,800	$39,500	$36,500	$36,700	$44,100

   a. Make a double stem-and-leaf plot comparing these salaries with the salaries of the MBA graduates in problem 2.
   b. Describe the salaries for the college graduates with bachelor's degrees in terms of any clusters, gaps, and outliers.
   c. What conclusions would you draw from this double stem-and-leaf plot?

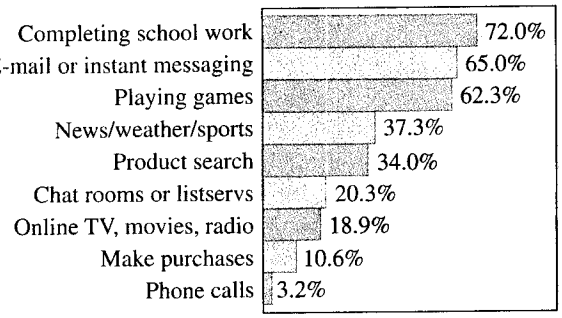
4. Consider the income data from problem 2.
   a. Make a histogram using bins of length 1000.
   b. Make a relative frequency histogram of this data. How do the relative histogram and the histogram from part (a) compare?

5. The following bar graph displays the total number of turkeys produced in the United States from 1975 through 2000.

   a. Which 5-year period showed the greatest change in turkey production?
   b. Create a line graph of the turkey-production data, using a vertical scale that emphasizes the changes in turkey production over time.

6. Consider the following bar graph.

Percentage of Youths Using Internet by Activity



Source: Adapted from AP/Wide World Photos

a. What do the percentages at the end of each bar represent?

b. Summarize the information from this bar chart in a paragraph.

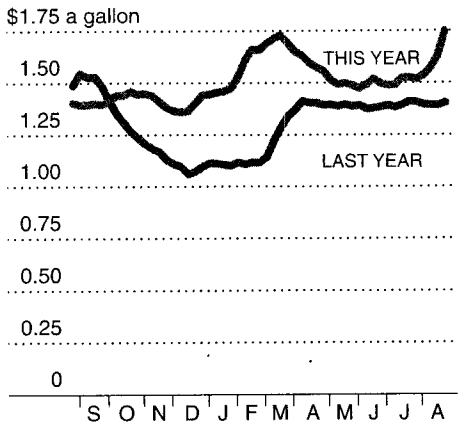
c. Explain why a pie chart is an inappropriate choice of graph to use to display these data.

7. Over the past few years, gasoline prices have fluctuated considerably. The following double-line graph shows the average price for regular gasoline from September 2002 through August 2003, labeled as "this year" on the graph, and the average price for regular gasoline from September 2001 through August 2002, labeled as "last year" on the graph.

Taking Its Toll

Gasoline prices fell sharply in the spring but have risen quickly over the last several weeks. They are now higher than they were a year ago.

Average price for regular gasoline in the United States, weekly



Source:
Energy Information Administration

The New York Times

a. During which month(s) from September 2002 through August 2003 were gas prices steadily rising? During which month(s) were gas prices fairly steady?

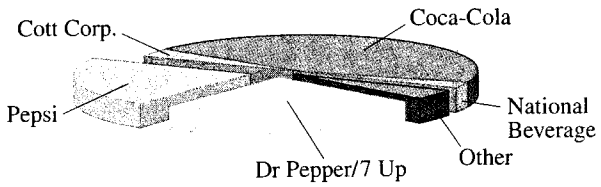
b. During which month(s) from September 2001 through August 2002 were gas prices steadily rising? For which month(s) were gas prices fairly steady?

c. During which month from September 2002 through August 2003 did the greatest increase in the price of gas occur? Answer the same question for the time period from September 2001 through August 2002.

d. Redraw the line graph to include only the months of May, June, July, and August for 2002 and 2003. Use a vertical scale that exaggerates the increase in the price of gas in 2003.

8. In 2001, five beverage companies held 96.9% of the market share. The following three-dimensional pie chart displays the percentage of the market held by each of the five companies.

Market Share 2001



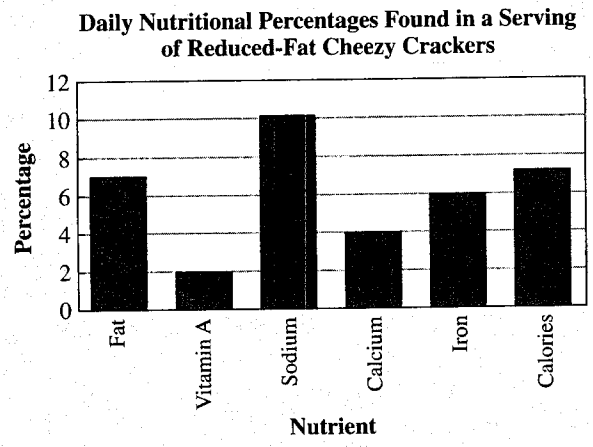
a. What is your first impression of the market share held by Pepsi and Dr Pepper/7 UP? Explain why this pie chart is misleading.

b. The actual market share percentages are given in the following table. Create a pie chart for these market share percentages.

Company	2001 Market Share Percentage
Coca-Cola	43.7
Pepsi	31.6
Dr Pepper/7 UP	15.6
Cott Corp.	3.8
National Beverage	2.2
Other	3.1

c. How does the pie chart in part (b) compare to the three-dimensional pie chart?

9. The following bar graph gives the percentages of the suggested daily requirements for fat, calories, and several vitamins and minerals found in a serving of Reduced-Fat Cheezy Snack Crackers. Percentages of daily values are based on a 2000-calorie diet.



a. How many calories are in one serving of Reduced-Fat Cheezy Snack Crackers?

b. The following label is from a box of another type of crackers, regular Cheezy Baked Snack Crackers. Use the nutritional information given on the label together with the bar graph for the Reduced-Fat Cheezy Snack Crackers to create a comparison bar graph for the two kinds of crackers. Use the same six categories as in the bar graph given. Describe the similarities between the two products and the most striking differences.

Cheezy Crackers

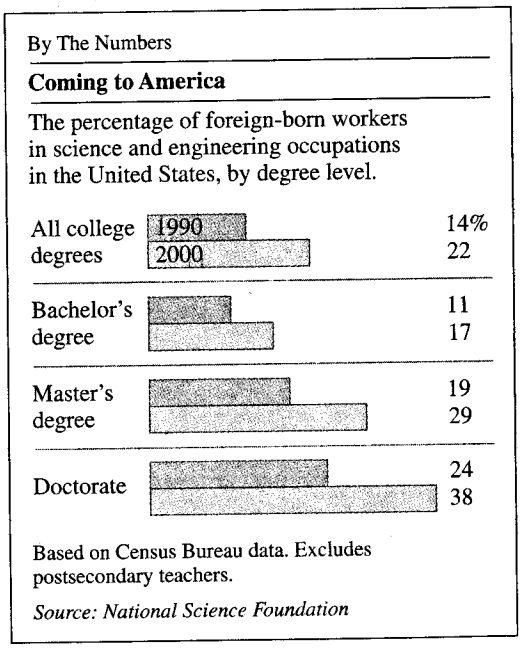
Amount Per Serving	
Calories 160	
	% Daily Value
Total Fat 8g	**12%**
Saturated Fat 2g	
Trans Fat 0g	
Cholesterol 0mg	**0%**
Sodium 250mg	**10%**
Total Carbohydrate 18g	
Dietary Fiber less than 1g	
Sugars less than 1g	
Protein 4g	
Vitamin A 2%	Vitamin C 0%
Calcium 4%	Iron 4%

c. To qualify as a "reduced-fat" product, the product must contain 25% less fat than the product to which it is being compared. Does a serving of Reduced-Fat Cheezy Snack Crackers qualify as a reduced-fat food? Explain. Create a comparison bar graph that emphasizes the fat difference between the two products.

10. Suppose the proportion of teenage consumer spending in relation to other age groups has been increasing over the past few years. Which type of graph would be most appropriate to use to display this trend? Justify your choice.

11. As part of a report you need to show numbers of refrigerators and dishwashers sold from the years 1950 through 2000. Which type of graph would most clearly communicate this information? Justify your choice.

12. The National Science Foundation found that 38 percent of all scientists and engineers with doctorates are immigrants who earned their degrees abroad. Some experts feel that this trend is alarming.



a. Create a comparison bar graph with vertical bars that emphasizes the differences in percentages between 1990 and 2000.

b. For the percentages of foreign-born workers with doctorate degrees, create a comparison line graph with years on the horizontal axis to emphasize the differences in percentages between 1900 and 2000.

c. Compare the given bar graph, the bar graph you created in part (a), and the comparison line graph created in part (b). If you wanted to emphasize the "alarming" trend, which graph would you select and why?

Age	1950	1960	1970	1980	1990	2000
Under 14	103.7	103.4	103.9	104.6	104.9	104.9
25 to 44	96.4	95.7	95.5	97.4	98.9	100.2
Over 65	89.6	82.8	72.1	67.6	67.2	70.8

Source: U.S. Census Bureau.

13. The preceding table contains the number of males per 100 females in the total U.S. resident population from 1950 through 2000. For example, the table shows that in 1950 there were 103.7 males under age 14 for every 100 females under age 14.

a. Construct a multiple-line graph that shows how these ratios changed from 1950 to 2000 for each age category.

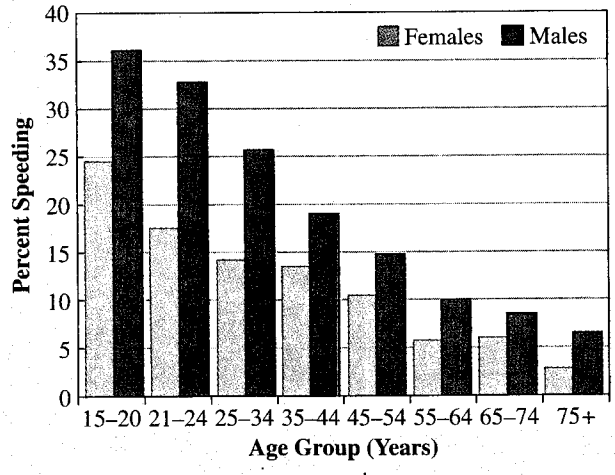
b. During which year, and for which age group, were there the fewest number of males per 100 females? During which year, and for which age group, were there about the same number of males and females?

c. Describe the trend in the number of males per 100 females for each of the three age groups over the past 50 years.

d. Based on the data, what would you conclude about life spans for males versus females in the United States?

14. The following bar chart gives the percentages of drivers in 2001 who were involved in fatal accidents and who were speeding. The graph shows data for both genders and for various age groups.

Fatal Crashes Involving Speeding Drivers by Age and Sex, 2001



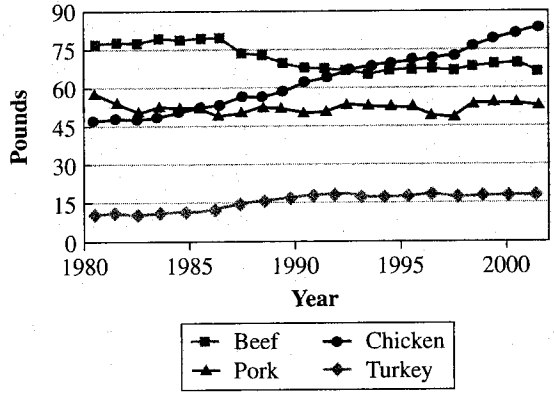
a. Approximately what percentage of male drivers aged 21 to 24 involved in fatal accidents were speeding? Between which two age groups is the greatest drop in the percentage of speeding male drivers involved in fatal crashes?

b. For what age group were approximately 10% of all female drivers involved in fatal accidents speeding? For which age group was the percentage of speeding drivers in fatal crashes for males and females most nearly equal?

c. Describe the trend over time in the percentages of males and females who were involved in fatal crashes and speeding. Why do you suppose car insurance costs are higher for males under the age of 25?

15. The largest segment of agriculture in the United States is the meat and poultry industry. The following multiple-line graph gives the average annual per-capita consumption of beef, pork, chicken, and turkey (in pounds).

Average Annual Per Capita Consumption



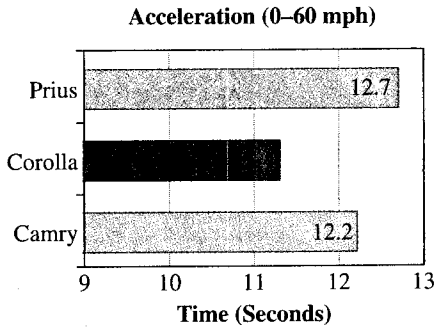
a. What time period is represented in the graph?

b. For what time period was the average annual consumption of beef more than 75 pounds per person?

c. In approximately what year did the average annual per-capita consumption of chicken equal that of beef? When did it equal the consumption of pork?

d. Estimate the total number of pounds of meat consumed per capita in 1980, 1985, 1990, 1995, and 2000. Create a line graph for the total number of pounds of meat consumed per capita over time, and discuss the trend revealed by the graph.

16. Motor Trend magazine named the Toyota Prius as the 2004 Car of the Year. The Toyota Prius is a gas/electric hybrid vehicle. The following bar graph compares the seconds required to accelerate from 0 to 60 miles per hour for three different Toyota vehicles.



a. What impression does the bar graph give? Why do you think the bar for the Corolla was placed next to the bar for the Prius?

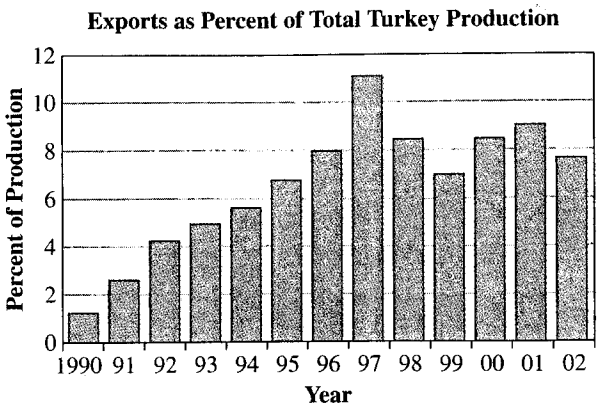
b. Redraw the bar graph so that the scale for the time axis begins at 0 seconds. Compare the bar graph you created with the given bar graph. How does changing the horizontal scale change the impression the graph gives about the comparative acceleration capabilities of the cars?

17. Turkey consumption continues to increase in the United States. The following table contains the number of turkeys raised in 2002 and 2003 in the top five turkey-producing states. Create a double-bar graph to represent the data.

State	2002	2003
North Carolina	45,500	45,900
Minnesota	44,000	45,500
Missouri	25,500	27,500
Arkansas	29,500	24,000
Virginia	20,000	23,000

Source: http://www.eatturkey.com.

18. Turkey exports expressed as a percentage of the total turkey production are displayed in the following bar graph for the years 1990 through 2002.



a. Describe the trend in the turkey exports since 1990.

b. What percentage of turkey production was exported in 1993? In which year(s) was more than 8% of the turkey production exported?

c. In which year was the largest percentage of turkey production exported?

d. In which year(s) did the percentage of turkey production exported decrease compared to the prior year?

19. Refer to the turkey export percentages from problem 18. In 2002, the percentage of production of turkey that was exported decreased compared to 2001. Create a misleading bar chart with bars for the years 2001 and 2002 that exaggerates the decline in exports.

20. Refer to the turkey export percentages from problem 18. Create a line graph showing the percentage of turkey production that is *not* exported each year. Describe the trend shown by the line graph.