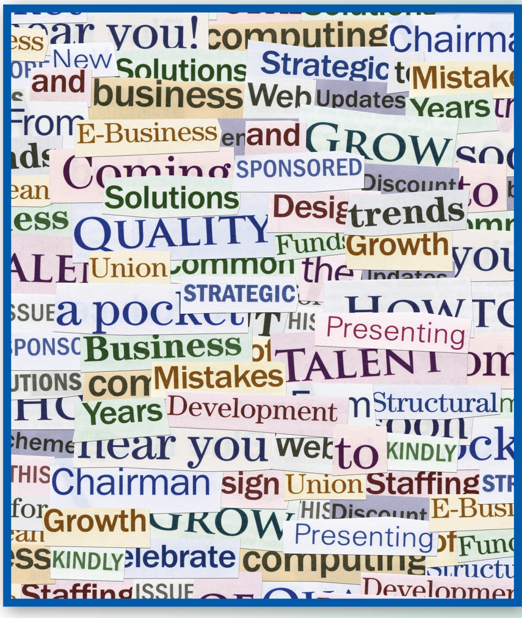
PART II

Organizing Data



iStockphoto/Thinkstock

Words alone don't tell a story. A writer organizes words into sentences and organizes the sentences into a story line. If the words are badly organized, the story isn't clear. Data also need organizing if they are to tell a clear story. Too many words obscure a subject rather than illuminate it. Vast amounts of data are even harder to digest—we often need a brief summary to highlight essential facts. How to organize, summarize, and present data are our topics in the second part of this book.

Organizing and summarizing a large body of facts opens the door to distortions, both unintentional and deliberate. This is no less (but also no more) the case when the facts take the form of numbers rather than words. We will point out some of the traps that data presentations can set for the unwary. Those who picture statistics as primarily a piece of the liar's art concentrate on the part of statistics that deals with summarizing and presenting data. We claim that misleading summaries and selective presentations go back to that after-the-apple conversation among Adam, Eve, and God. Don't blame statistics. Do remember the saying "Figures won't lie, but liars will figure," and beware.

Graphs, Good and Bad

10

Monkey Business Images/Shutterstock Images LLC

CASE STUDY The National Center for Health Statistics (NCHS) produces a report each year summarizing leading causes of death, changes in mortality over time, and other characteristics for mortality in the United States. In its March 2013 brief, "Death in the United States, 2011," Figure 10.1 was included to display the top five causes of death for four age categories. A clear graphical display can make it easy for policy and decision makers to decide how to allocate resources for prevention or intervention. For instance, a quick glance at the pie chart for ages 1–24 shows that a large portion, 13% of deaths in that age range, are attributed to suicide. A clear display of this staggering number makes a strong argument for allocation of resources.

Statistics deals with data, and we use tables and graphs to present data. Tables and graphs help us see what the data say. But not all tables and graphs do so accurately or clearly. In this chapter, you will learn some basic methods for displaying data and how to assess the quality of the graphics you see in the media. By the end of the chapter, you will be able to determine whether Figure 10.1 is a good or a bad graphic.

Data tables

Take a look at the *Statistical Abstract of the United States*, an annual volume packed with almost every variety of numerical information. Has the number of private elementary and secondary schools grown over time? What about minority enrollments in these schools? How many college degrees were given in each of the past several years, and how were these degrees divided among fields of study and by the age, race, and sex of the students? You can find all this and more in the education section of

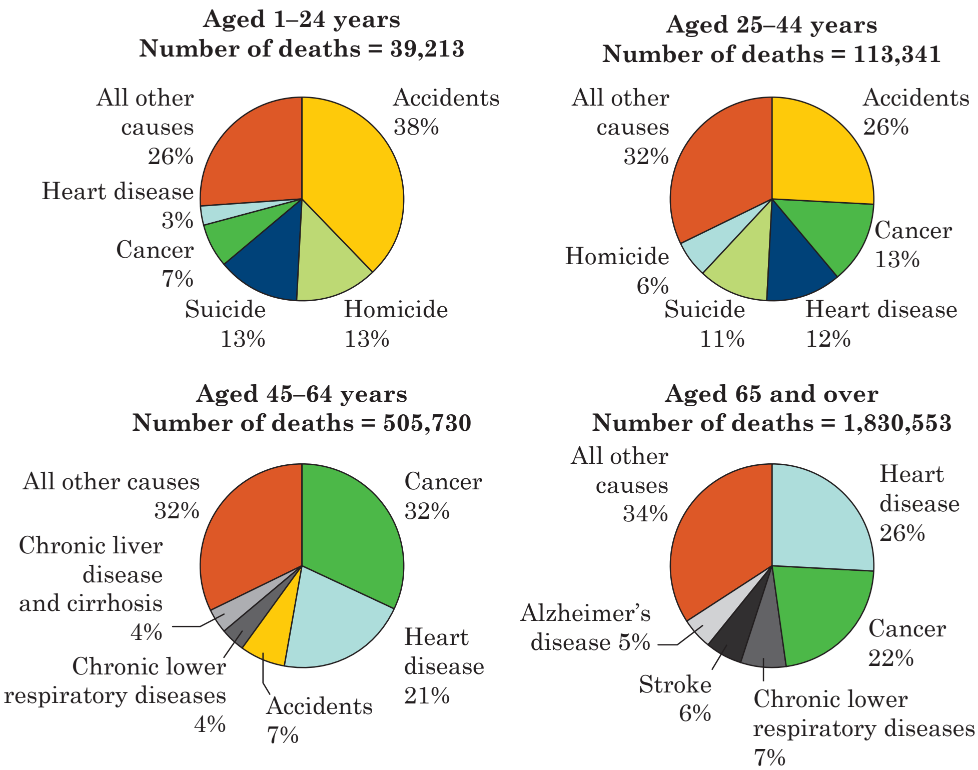


Figure 10.1 Top five causes of death compared across four age categories. (http://www.cdc.gov/nchs/data/databriefs/db115.htm)

the *Statistical Abstract*. The tables *summarize* data. We don't want to see information on every college degree individually, only the counts in categories of interest to us.

EXAMPLE 1 What makes a clear table?

How well educated are adults? Table 10.1 presents the data for people aged 25 years and over. This table illustrates some good practices for data tables. It is clearly *labeled* so that we can see the subject of the data at once. The main heading describes the general subject of the data and gives the date because these data will change over time. Labels within the table identify the variables and state the *units* in which they are measured. Notice, for example, that the counts are in thousands. The *source* of the data appears at the foot of the table. This Census Bureau publication in fact presents data from our old friend, the Current Population Survey.

TABLE 10.1 Education of people aged 25 years and over in 2014

Level of education	Number of persons (thousands)	Percent
Less than high school	24,458	11.7
High school graduate	62,240	29.7
Some college, no degree	34,919	16.7
Associate's degree	20,790	9.9
Bachelor's degree	42,256	20.2
Advanced degree	24,623	11.8
Total	209,287	100.0

Source: Census Bureau, *Educational Attainment in the United States: 2014.*

Table 10.1 starts with the *counts* of people aged 25 years and over who have attained each level of education. *Rates* (percentages or proportions) are often clearer than counts—it is more helpful to hear that 11.7% of this age group did not finish high school than to hear that there are 24,458,000 such people. The percentages also appear in Table 10.1. The last two columns of the table present the *distribution* of the variable "level of education" in two alternate forms. Each of these columns gives information about what values the variable takes and how often it takes each value.

Distribution of a variable

The **distribution** of a variable tells us what values it takes and how often it takes these values.

EXAMPLE 2 Roundoff errors

Did you check Table 10.1 for consistency? The total number of people should be

$$26{,}458 + 62{,}240 + 34{,}919 + 20{,}790 + 42{,}256 \\ + 24{,}623 = 209{,}286 \text{ (thousands)}$$

But the table gives the total as 209,287. What happened? Each entry is rounded to the nearest thousand. The rounded entries don't quite add to the total, which is rounded separately. Such **roundoff errors** will be with us from now on as we do more arithmetic.

It is not uncommon to see roundoff errors in tables. For example, when table entries are percentages or proportions, the total may sum to a value slightly different from 100% or 1, often to 99.9 or 100.1%.

Types of variables

When we think about graphs, it is helpful to distinguish between variables that place individuals into categories (such as gender, occupation, or education level) and those whose values have a meaningful numerical scale (such as height in centimeters or SAT scores).

Categorical and quantitative variables

A **categorical variable** places an individual into one of several groups or categories.

A **quantitative variable** takes numerical values for which arithmetic operations such as adding and averaging make sense.

Pie charts and bar graphs

The variable displayed in Table 10.1, "level of education," is a categorical variable. There are six possible values of the variable. To picture this distribution in a graph, we might use a **pie chart**. Figure 10.2 is a pie chart of the level

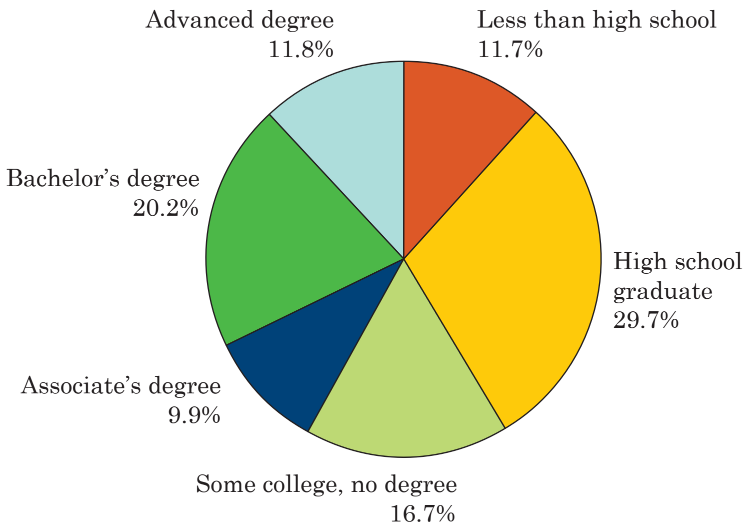


Figure 10.2 Pie chart of the distribution of level of education among persons aged 25 years and over in 2014.

of education of people aged 25 years and over. Pie charts show how a whole is divided into parts. To make a pie chart, first draw a circle. The circle represents the whole, in this case all people aged 25 years and over. Wedges within the circle represent the parts, with the angle spanned by each wedge in proportion to the size of that part. For example, 20.2% of those in this age group and over have a bachelor's degree but not an advanced degree. Because there are 360 degrees in a circle, the "bachelor's degree" wedge spans an angle of

$$0.202 \times 360 = 72.72 \text{ degrees}$$

Pie charts force us to see that the parts do make a whole. However, it is much easier for our eyes to compare the heights of the bars on a bar graph than it is to compare the size of the angles on a pie chart.

Figure 10.3 is a **bar graph** of the same data. There is a single bar for each education level. The height of each bar shows the percentage of people aged 25 years and over who have attained that level of education. Notice that each bar has the same width—this is always the case with a bar graph. Also, there is a space between the bars. It is obvious on either the pie chart or bar graph that the largest category is "High school graduate." However, smaller differences are much more subtle on a pie chart. It is very clear on the bar graph that "Bachelor's degree" is the second largest category. This is not as clear on the pie chart. Bar graphs are better for making comparisons of the sizes of categories. In addition, if there is a natural ordering of the variable, such as how much education a person has, this order can be displayed along the horizontal axis of the bar graph but cannot be displayed in an obvious way in a pie chart.

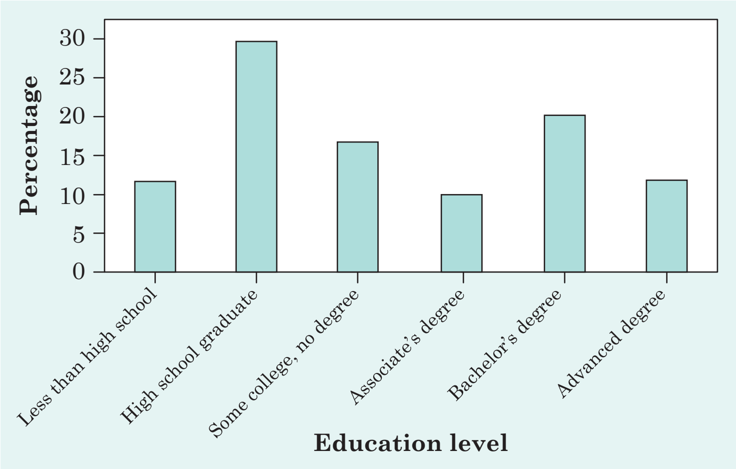


Figure 10.3 Bar graph of the distribution of level of education among persons aged 25 years and over in 2014.

Pie charts and bar graphs can show the distribution (either counts or percentages) of a categorical variable such as level of education. A pie chart usually displays the percentage for each category (rather than the count) and only works if you have all the categories (the percentages add to 100). A bar graph can display an entire distribution or can be used to display only a few categories. A bar graph can also compare the size of categories that are not parts of one whole. If you have one number to represent each category, you can use a bar graph.

EXAMPLE 3 High taxes?

Figure 10.4 compares the level of taxation in eight democratic nations. Each democratic nation is a category and each category is described by one value, so a bar graph is a good choice for a graphical display of these data. The heights of the bars show the percentages of each nation's gross domestic product (GDP, the total value of all goods and services produced) that is taken in taxes. Americans accustomed to complaining about high taxes may be surprised to see that the United States, at 25.4% of GDP, is at the bottom of the group. Notice that a pie chart is not possible for these data because we are displaying eight separate quantities, not the parts of a whole.

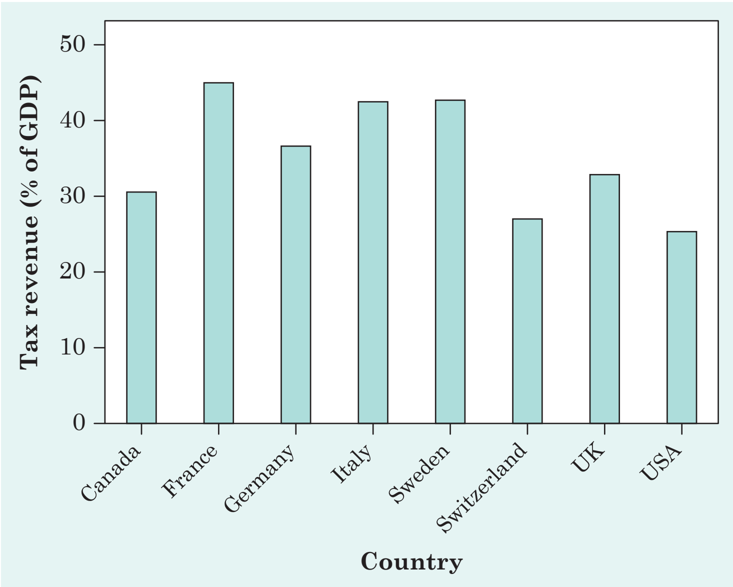


Figure 10.4 Total tax revenue as a percentage of GDP in eight countries that are part of the OECD, Example 3. (Data from the Organization for Economic Cooperation and Development, http://stats.oecd.org/index.aspx)

NOW IT'S YOUR TURN

10.1 Taxes. There are currently seven states in the United States that do not collect income tax. The accompanying table contains the combined state and average local sales tax in 2014 for these seven states as computed by the Tax Foundation.

State	State and average local sales tax
Alaska	1.69
Florida	6.62
Nevada	7.93
South Dakota	5.83
Texas	8.15
Washington	8.88
Wyoming	5.49

Is the variable "state" categorical or quantitative? Should you use a pie chart or a bar graph to display the state and average local sales tax for these states? Why?

EXAMPLE 4 Government tax revenue breakdown

Governments generate tax revenue through various means. In its 2014 report, the Tax Foundation analyzed data for 2011 from the Organisation for Economic Co-operation and Development (OECD). The United States' tax revenue comprises individual income tax (37.1%), social insurance tax (22.8%), consumption tax (18.3%), property tax (12.4%), and corporate income tax (9.4%). A bar graph is appropriate to display these data because we have one value to explain the size of each tax category. Notice, a pie graph is appropriate as well, because these are all the parts of a whole. What if we wanted to compare the distribution of tax revenue for the United States to the average for all the other OECD countries? Pie charts are not good for comparisons. Figure 10.5 is a bar graph for the distribution of government tax revenue for the United States with a second set of bars representing the average for all other OECD nations adjacent to the bars for the United States. This is called a side-by-side bar graph. A side-by-side bar graph is useful for making comparisons. It is now clear that the United States relies more heavily on individual income tax, while the other countries rely more heavily on consumption taxes.

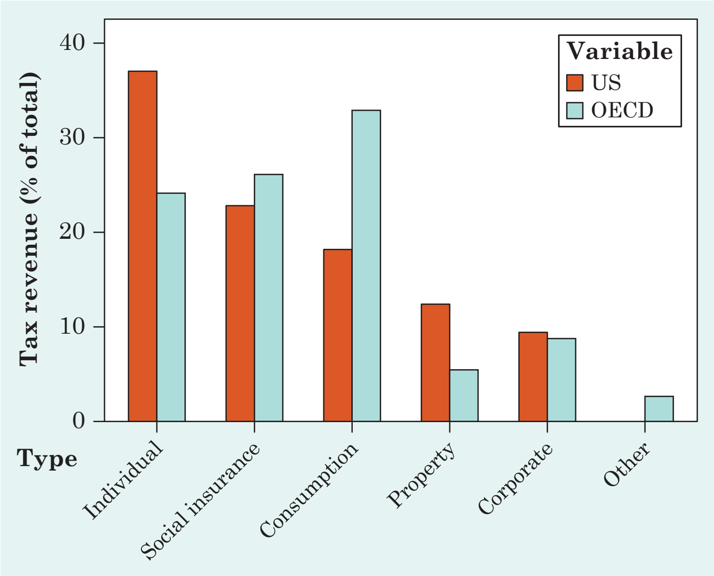


Figure 10.5 Side-by-side bar graph comparing the distribution for sources of government tax revenue in 2011 for the United States to the average of all other countries belonging to the Organisation for Economic Co-operation and Development. (http://taxfoundation.org/article/sources-government-revenue-oecd-2014)

Beware the pictogram

Bar graphs compare several quantities by means of the differing heights of bars that represent the quantities. Our eyes, however, react to the *area* of the bars as well as to their height. When all bars have the same width, the area (width $\times$ height) varies in proportion to the height and our eyes receive the right impression. When you draw a bar graph, make the bars equally wide. Artistically speaking, bar graphs are a bit dull. It is tempting to replace the bars with pictures for greater eye appeal.

EXAMPLE 5 A misleading graph

Figure 10.6 is a **pictogram**. It is a bar graph in which pictures replace the bars. The graph is aimed at consumers shopping for a vacuum. It claims that a Dyson vacuum has more than twice the suction of any other vacuum. The numbers above the vacuums show the calculated air watts for three vacuums. We see the Dyson vacuum (far right) has roughly four times the air watts (160 is roughly 4 times 43) of the vacuum on the far left. However, this graphic makes it appear as though there is a much larger difference. Why is this?

To magnify a picture, the artist must increase *both* height and width to avoid distortion of the image. By increasing both the height and width of the Dyson vacuum, it appears to be $4 \times 4 = 16$ times larger. Remember a bar graph should have bars of equal width—only the heights of the bars should vary. Replacing the bars on a bar graph with pictures is tempting, but it is difficult to keep the "bar" width equal without distorting the picture.

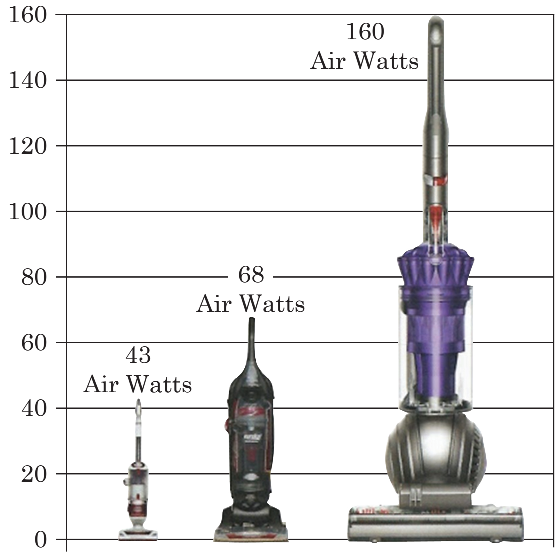


Figure 10.6 Advertisement found on Dyson's website for vacuum cleaners claiming a Dyson vacuum has more than twice the suction of any other vacuum, Example 5. (http://www.dyson.com/United States.aspx)

Change over time: Line graphs

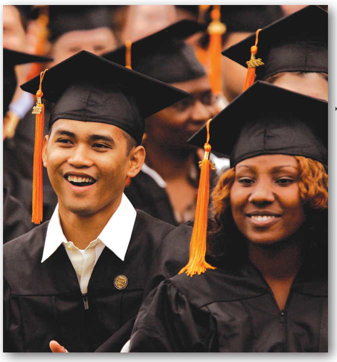
Many quantitative variables are measured at intervals over time. We might, for example, measure the height of a growing child or the price of a stock at the end of each month. In these examples, our main interest is change over time. To display how a quantitative variable changes over time, make a *line graph*.

Line graph

A **line graph** of a quantitative variable plots each observation against the time at which it was measured. Always put time on the horizontal scale of your plot and the variable you are measuring on the vertical scale. Connect the data points by lines to display the change over time.

When constructing a line graph, make sure to use equally spaced time intervals on the horizontal axis to avoid distortion. Line graphs can also be used to show how a quantitative variable changes over time, broken down according to some other categorical variable. Use a separate line for each category.

EXAMPLE 6 Unemployment by education level



AP Photo/Stephan Savoia

How has unemployment changed over time? Figure 10.7 is a line graph of the monthly national unemployment rate, by education level, for the United States from October 2008 through October 2014. For example, the unemployment rate for October 2009 was 15.2% for those with less than a high school diploma, 11% for those with a high school diploma, 8.7% for those with some college, and 4.7% for those with a bachelor's degree or higher.

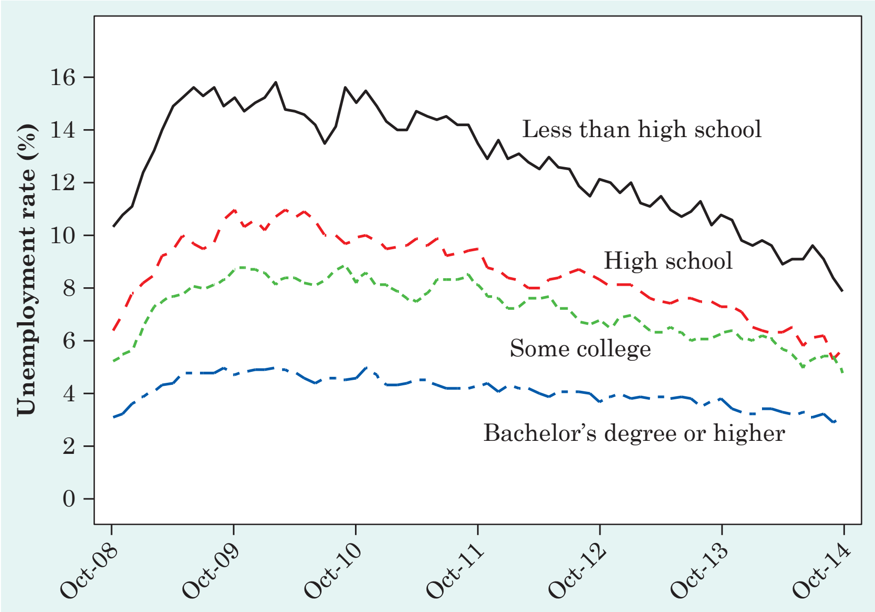


Figure 10.7 Monthly unemployment rate for four education categories from October 2008 through October 2014 reported by the Bureau of Labor Statistics. (http://www.bls.gov/opub/ted/2014/ted_20141112.htm)

It would be difficult to see patterns in a long table of numbers. Figure 10.7 makes the patterns clear. What should we look for?

- First, look for an **overall pattern**. For example, a **trend** is a long-term upward or downward movement over time. Unemployment was at its highest for all education groups in 2009 and 2010, due to the Great Recession. Since then, the overall trend is showing a decrease in the unemployment rate for all education levels (each line is generally decreasing).

- Next, look for striking **deviations** from the overall pattern. There is a noticeable increase from 2008 to the beginning of 2009. This was a side effect of the recession economy. Unemployment hovered around these record highs through 2010, when the rates finally started their decline to the current levels. There is a striking deviation in the unemployment rate for those with less than a high school degree around mid-2010. As the overall pattern of generally decreasing begins, there is a much more drastic dip in mid-2010 for those with less than a high school degree.

Statistics in Your World

The Vietnam effect Folklore says that during the Vietnam War, many men went to college to avoid being drafted. Statistician Howard Wainer looked for traces of this "Vietnam effect" by plotting data against time. He found that scores on the Armed Forces Qualifying Test (an IQ test given to recruits) dipped sharply during the war, then rebounded. Scores on the SAT exams, taken by students applying to college, also dropped at the beginning of the war. It appears that the men who chose college over the army lowered the average test scores in both places.

- Change over time often has a regular pattern of **seasonal variation** that repeats each year. Calculating the unemployment rate depends on the size of the workforce and the number of those in the workforce who are working. For example, the unemployment rate rises every year in January as holiday sales jobs end and outdoor work slows in the north due to winter weather. It would cause confusion (and perhaps political trouble) if the government's official unemployment rate jumped every January. These are regular and predictable changes in unemployment. As such, the Bureau of Labor Statistics makes seasonal adjustments to the monthly unemployment rate before reporting it to the general public.

Seasonal variation, seasonal adjustment

A pattern that repeats itself at known regular intervals of time is called **seasonal variation**. Many series of regular measurements over time are **seasonally adjusted**. That is, the expected seasonal variation is removed before the data are published.

The line graph of Figure 10.7 shows the unemployment rate for four different education levels. A line graph may have only one line or more than one, as was the case in Figure 10.7. Picturing the unemployment rate for all four groups over this time period gives an additional message. We see that unemployment is always lower with more education—a strong argument for a college education! Also, the increase in unemployment was much more drastic during the Great Recession for the lower education levels. It appears that the unemployment rate is much more stable for those with higher education levels and much more variable for those with lower education levels.

NOW IT'S YOUR TURN

10.2 Food stamp participation. The Supplemental Nutrition Assistance Program (SNAP), formerly know as the Food Stamp Program, is a federal program that provides money to be spent on food to eligible low-income individuals and families. The following table gives the percent of the population receiving benefits.

Date	Rate (%)	Date	Rate (%)
1970	2.12	1980	9.28
1972	5.29	1982	9.37
1974	6.01	1984	8.84
1976	8.51	1986	8.09
1978	7.19	1988	7.63

(Continued)

SNAP-ED/USDA

Date	Rate (%)	Date	Rate (%)
1990	8.04	2004	8.13
1992	9.96	2006	8.90
1994	10.55	2008	9.28
1996	9.63	2010	13.03
1998	7.32	2012	14.85
2000	6.09	2014	14.59
2002	6.64		

Make a line graph of these data and comment on any patterns you observe.

Watch those scales!

Because graphs speak so strongly, they can mislead the unwary. The careful reader of a line graph looks closely at the *scales* marked off on the axes.

EXAMPLE 7 Living together

The number of unmarried couples living together has increased in recent years, to the point that some people say that cohabitation is delaying or even replacing marriage. Figure 10.8 presents two line graphs of the number of unmarried-couple households in the United States. The data once again come from the Current Population Survey. The graph on the left suggests a steady but moderate increase. The right-hand graph says that cohabitation is thundering upward.

The secret is in the scales. You can transform the left-hand graph into the right-hand graph by stretching the vertical scale, squeezing the horizontal scale, and cutting off the vertical scale just above and below the values to be plotted. Now you know how to either exaggerate or play down a trend in a line graph.

Which graph is correct? Both are accurate graphs of the data, but both have scales chosen to create a specific effect. Because there is no one "right" scale for a line graph, correct graphs can give different impressions by their choices of scale. Watch those scales!

Another important issue concerning scales is the following. When examining the change in the price or value of an item over time, plotting the actual increase can be misleading. It is often better to plot the percentage increase from the previous period.

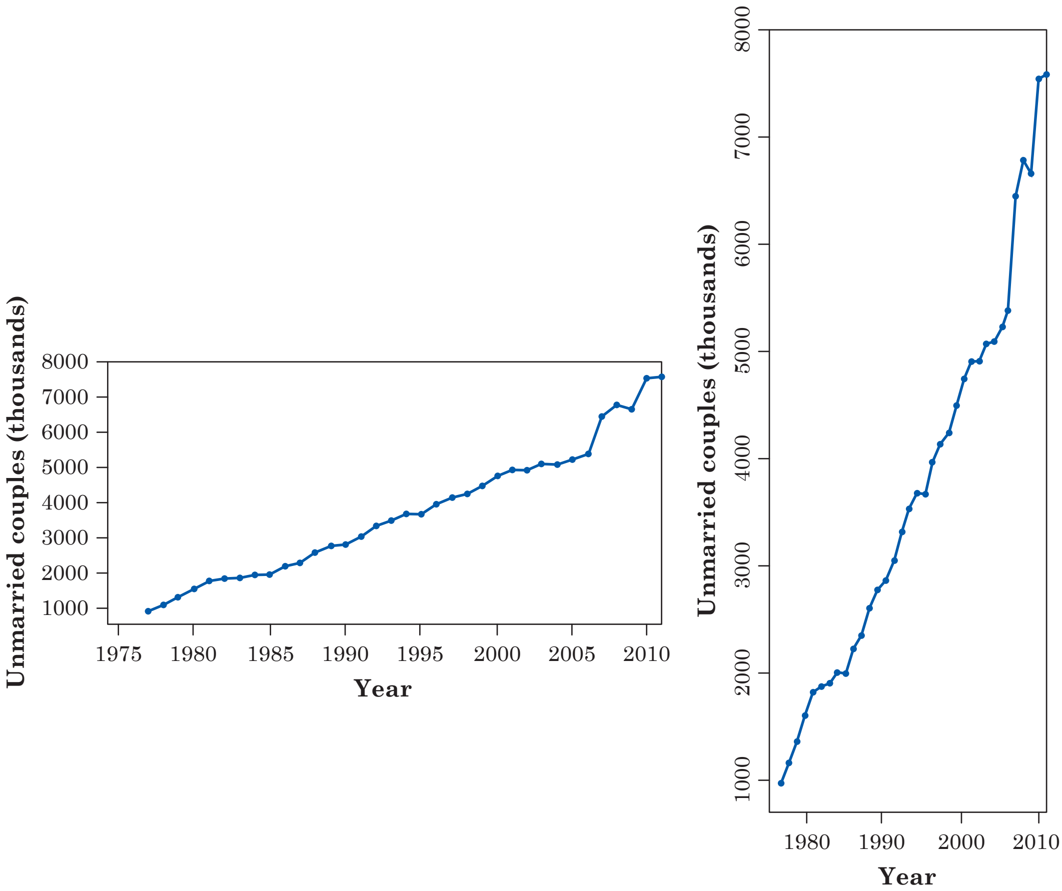


Figure 10.8 The effect of changing the scales in a line graph, Example 7. Both graphs plot the same data, but the right-hand graph makes the increase appear much more rapid.

EXAMPLE 8 Getting rich in hindsight

The end of the twentieth century saw a great bull market (a period when the value of stocks rises) in U.S. common stocks. How great? Pictures tell the tale more clearly than words.

Look first at Figure 10.9. This shows the percentage increase or decrease in stocks (measured by the Standard & Poor's 500 index) in each year from 1971 to 2011. Until 1982, stock prices bounce up and down. Sometimes they go down a lot—stocks lost 14.7% of their value in 1973 and another 26.5% in 1974. But starting in 1982, stocks go up in 17 of the next 18 years, often by a lot. From 2000 to 2009 stocks again bounce up and down, with a large loss of 37% in the recession that began in 2008.

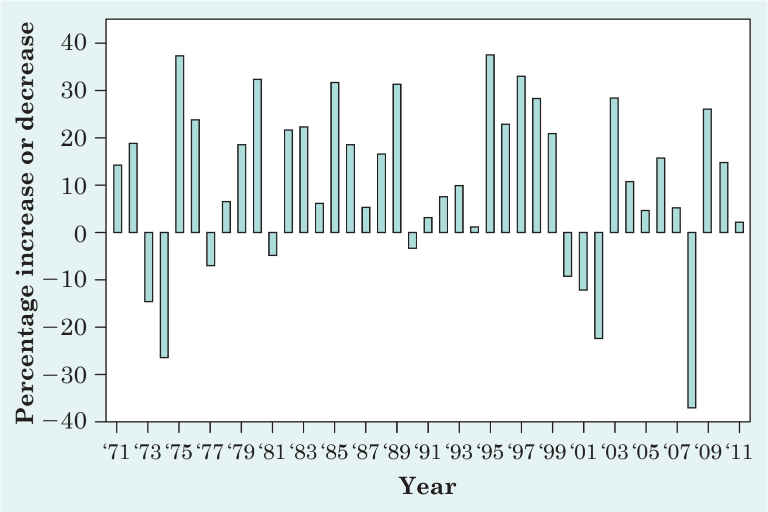


Figure 10.9 Percentage increase or decrease in the S&P 500 index of common stock prices, 1971 to 2011, Example 8.

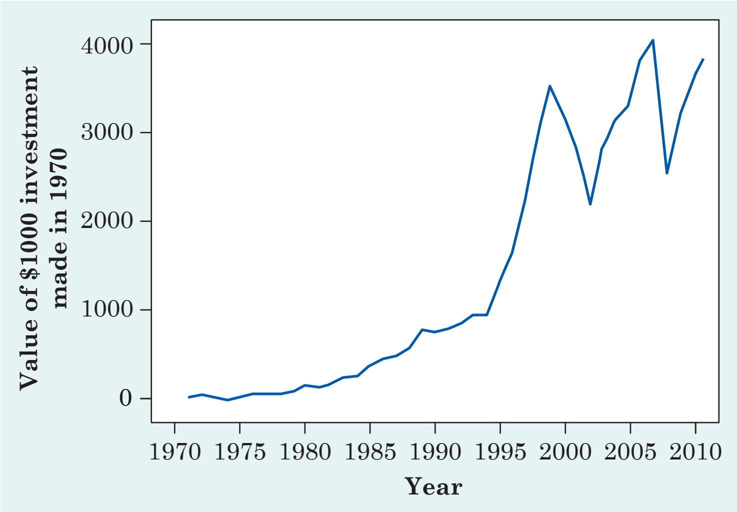


Figure 10.10 Value at the end of each year, 1971 to 2011, of $1000 invested in the S&P 500 index at the end of 1970, Example 8.

Figure 10.10 shows, in hindsight, how you could have become rich in the period from 1971 to 2011. If you had invested $1000 in stocks at the end of 1970, the graph shows how much money you would have had at the end of each of the following years. After 1974, your $1000 was down to $853, and at the end of 1981, it had grown to only $2145. That's an increase of only 7.2% a year. Your money would have grown faster in a bank during these years. Then the great bull market begins its work. By the end of 2011, it would have turned your $1000 into $36,108. Unfortunately, over the next 10 years as a whole, stocks lost value, and by the end of 2009, your $36,108 would have declined to $32,818.

Figures 10.9 and 10.10 are instructive. Figure 10.10 gives the impression that increases between 1970 and 1980 were negligible but that increases between 1995 and 1999 were dramatic. While it is true that the *actual value* of our investment increased much more between 1995 and 1999 than it did between 1970 and 1980, it would be incorrect to conclude that investments in general increased much more dramatically between 1995 and 1999 than in any of the years between 1970 and 1980.

Figure 10.9 tells a different, and more accurate, story. For example, the percentage increase in 1975 (approximately 37%) rivaled that in any of the years between 1995 and 1999. However, in 1975 the actual value of our investment was relatively small ($1170) and a 37% increase in such a small amount is nearly imperceptible on the scale used in Figure 10.10. By 1995, the actual value of our investment was about $14,000, and a 37% increase appears much more striking.

Making good graphs

Graphs are the most effective way to communicate using data. A good graph will reveal facts about the data that would be difficult or impossible to detect from a table. What is more, the immediate visual impression of a graph is much stronger than the impression made by data in numerical form. Here are some principles for making good graphs:

- **Make sure labels and legends** tell what variables are plotted, their units, and the source of the data.
- **Make the data stand out.** Be sure that the actual data—not labels, grids, or background art—catch the viewer's attention. You are drawing a graph, not a piece of creative art.
- **Pay attention to what the eye sees.** Avoid pictograms and be careful choosing scales. Avoid fancy "three-dimensional" effects that confuse the eye without adding information. Ask if a simple change in a graph would make the message clearer.

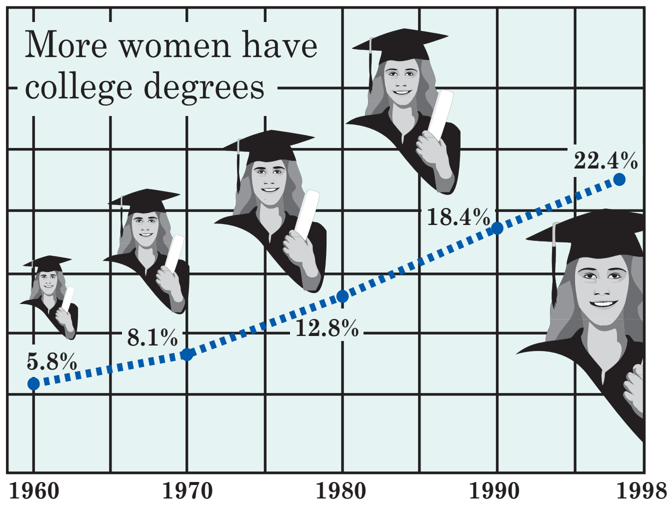


Figure 10.11 Chart junk: this graph is so cluttered with unnecessary ink that it is hard to see the data, Example 9.

EXAMPLE 9 The rise in college education

Figure 10.11 shows the rise in the percentage of women aged 25 years and over who have at least a bachelor's degree. There are only five data points, so a line graph should be simple. Figure 10.11 isn't simple. The artist couldn't resist a nice background sketch and also cluttered the graph with grid lines. We find it harder than it should be to see the data. Using pictures of the female graduate at each time point distorts the values in the same way as a pictogram. Grid lines on a graph serve no purpose—if your audience must know the actual numbers, give them a table along with the graph. A good graph uses no more ink than is needed to present the data clearly. In statistical graphics, decoration is always a distraction from the data, and sometimes an actual distortion of the data.

EXAMPLE 10 High taxes, reconsidered

Figure 10.4 is a respectable bar graph comparing taxes as a percentage of gross domestic product in eight countries. The countries are arranged alphabetically. Wouldn't it be clearer to arrange them in order of their tax burdens? Figure 10.12 does this. This simple change improves the graph by making it clearer where each country stands in the group of eight countries. Figure 10.12 also demonstrates the ability to display a bar graph horizontally.

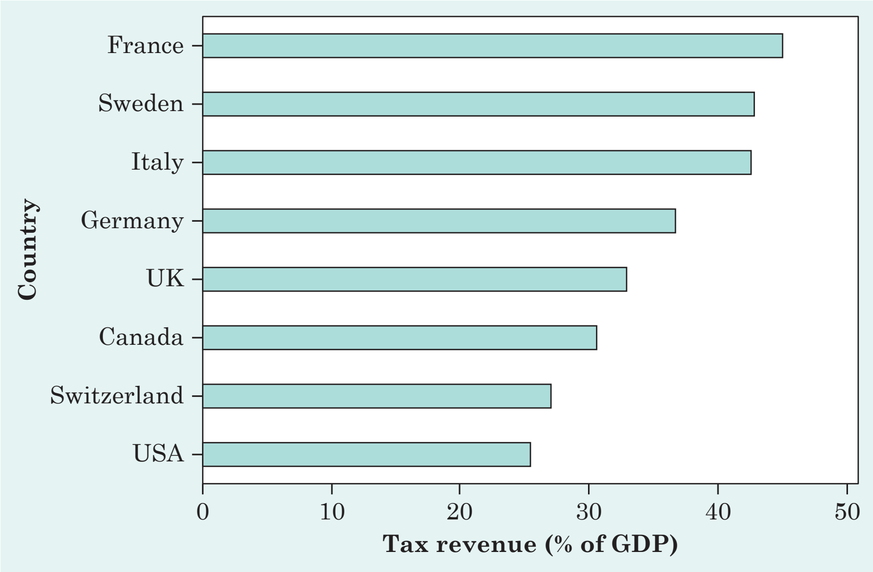


Figure 10.12 Total U.S. government tax revenue as a percentage of the GDP for eight countries who are part of the OECD, Example 10. Changing the order of the bars has improved the graph in Figure 10.4. (http://stats.oecd.org/index.aspx)

STATISTICS IN SUMMARY

Chapter Specifics

- To see what data say, start with graphs.
- The choice of graph depends on the type of data. Do you have a **categorical variable**, such as level of education or occupation, which puts individuals into categories? Or do you have a **quantitative variable** measured in meaningful numerical units?
- Check data presented in a table for **roundoff errors**.
- The **distribution** of a variable tells us what values it takes and how often it takes those values.
- To display the **distribution** of a categorical variable, use a **pie chart** or a **bar graph**. Pie charts always show the parts of some whole, but bar graphs can compare any set of numbers measured in the same units. Bar graphs are better for comparisons. Bar graphs can be displayed vertically or horizontally.
- To show how a quantitative variable changes over time, use a **line graph** that plots values of the variable (vertical scale) against time (horizontal scale). If you have values of the variable for different categories, use a

separate line for each category. Look for **trends** and **seasonal variation** in a line graph, and ask whether the data have been **seasonally adjusted**.

- Graphs can mislead the eye. Avoid **pictograms** that replace the bars of a bar graph by pictures whose height and width both change. Look at the scales of a line graph to see if they have been stretched or squeezed to create a particular impression. Avoid clutter that makes the data hard to see.

LINK IT In reasoning from data to a conclusion, where the data come from is important. We studied this in Chapters 1 through 9. Once we have the data, and are satisfied that they were produced appropriately, we can begin to determine what the data tell us. Tables and graphs help us do this. In this chapter, we learned some basic methods for displaying data with tables and graphs. We learned what information these graphics provide. An important type of information is the distribution of the data—the values that occur and how often they occur. The concept of the distribution of data or the distribution of a variable is a fundamental way that statisticians think about data. We will encounter it again and again in future chapters.

Data that are produced badly can mislead us. Likewise, graphs that are produced badly can mislead us. In this chapter, we learned how to recognize bad graphics. Developing "graphic sense," the habit of asking if a graphic accurately and clearly displays our data, is as important as developing "number sense," discussed in Chapter 9.

CASE STUDY EVALUATED Look again at Figure 10.1, described in the Case Study that opened the chapter. Based on what you have learned, is Figure 10.1 the best graphical representation of the top five leading causes of death across four major age categories? What are the drawbacks to the current graphical display? Discuss which graphical display would be better and why. If you have access to statistical software, create the graphical display you think is better.

LaunchPad macmillan learning Online Resources

- The Snapshots Video *Visualizing and Summarizing Categorical Data* discusses categorical data and describes pie charts and bar graphs to summarize categorical data in the context of data from a NASA program.
- The StatClips Video *Summaries and Pictures for Categorical Data* discusses how to draw a pie chart and a bar graph using two examples.
- The StatClips Video *Summaries and Pictures for Categorical Data Example A* discusses pie charts and bar graphs in the context of data about the choice of field of study by first-year students.

- The StatClips Video *Summaries and Pictures for Categorical Data Example B* discusses bar graphs in the context of data from a survey about high-tech devices.
- The StatClips Video *Summaries and Pictures for Categorical Data Example C* discusses the use of pie charts and bar graphs in the context of data from the Arbitron ratings.

CHECK THE BASICS

For Exercise 10.1, see page ~~223~~; for Exercise 10.2, see page ~~227~~.

10.3 Categorical and quantitative variables. A survey was conducted and respondents were asked what color car they drive and how many miles they travel per day. The variable type for "car color" and "miles per day" are

(a) both categorical.

(b) both quantitative.

(c) car color—quantitative, miles per day—categorical.

(d) car color—categorical, miles per day—quantitative.

10.4 Line graph. Which of the following is not acceptable for a line graph?

(a) Having lines that cross

(b) Having equally spaced time intervals

(c) Having time intervals that are not equally spaced

(d) Including seasonal variation

10.5 Distributions. A recent report on the religious affiliation of Hispanics says 55% are Catholic, 22% Protestant, 18% unaffiliated, and 4% other. This adds up to 99%. You conclude

(a) the remaining 1% are not accounted for.

(b) there was a calculation or data entry error.

(c) this is due to roundoff error.

(d) there is a religion missing from the survey.

10.6 Which graph? You have the average SAT score of entering freshmen for five universities. The best graphical display for these data would be a

(a) pie chart.

(b) bar graph.

(c) line graph.

(d) side-by-side bar graph.

10.7 Which graph? You want to show how the price of cable television has changed over the years. You should use a

(a) pie chart.

(b) bar graph.

(c) line graph.

(d) side-by-side bar graph.

10.8 Bar graph. For a bar graph to be accurate,

(a) it should be vertical.

(b) the bars should have equal height.

(c) the bars should touch each other.

(d) the bars should have equal width.

CHAPTER 10 EXERCISES

10.9 Lottery sales. States sell lots of lottery tickets. Table 10.2 shows where the money comes from in the state of Illinois. Make a bar graph that shows the distribution of lottery sales by type of game. Is it also proper to make a pie chart of these data? Explain.

10.10 Consistency? Table 10.2 shows how Illinois state lottery sales are divided among different types of games. What is the sum of the amounts spent on the 11 types of games? Why is this sum not exactly equal to the total given in the table?

10.11 Marital status. In the U.S. Census Bureau document *America's Families and Living Arrangements: 2011*, we find these data on the marital status of American women aged 15 years and older as of 2011:

Marital status	Count (thousands)
Never married	34,963
Married	65,000
Widowed	11,306
Divorced	13,762

(a) How many women were not married in 2011?

(b) Make a bar graph to show the distribution of marital status.

(c) Would it also be correct to use a pie chart? Explain.

TABLE 10.2 Illinois state lottery sales by type of game, fiscal year 2010

Game	Sales (dollars)
Pick Three	301,417,049
Pick Four	191,038,518
Lotto	111,158,528
Little Lotto	119,634,946
Mega Millions	221,809,484
Megaplier5	848,077
Pick N Play	1,549,252
Raffle	19,999,460
Powerball	43,269,461
Power Play	8,469,680
Instants	1,190,109,917
Total	2,209,304,371

Source: Illinois Lottery Fiscal Year 2010 Financial Release.

10.12 We pay high interest. Figure 10.13 shows a graph taken from an advertisement for an investment that promises to pay a higher interest rate than bank accounts and other competing investments. Is this graph a correct comparison of the four interest rates? Explain your answer.

10.13 Attitudes on same-sex marriage. Attitudes on same-sex marriage have changed over time, but they also differ according to age. Figure 10.14 shows change in the attitudes on same-sex marriage for four generational cohorts. Comment on the overall trend in change in attitude. Explain how attitude on same-sex marriage differs according to generation.

10.14 Murder weapons. The 2010 *Statistical Abstract of the United States* reports FBI data on murders for 2007. In that year, 49.6% of all

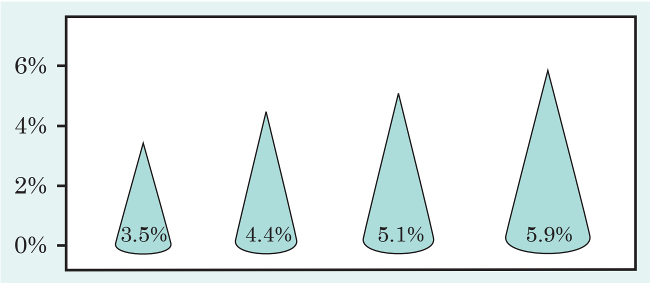


Figure 10.13 Comparing interest rates, Exercise 10.12.

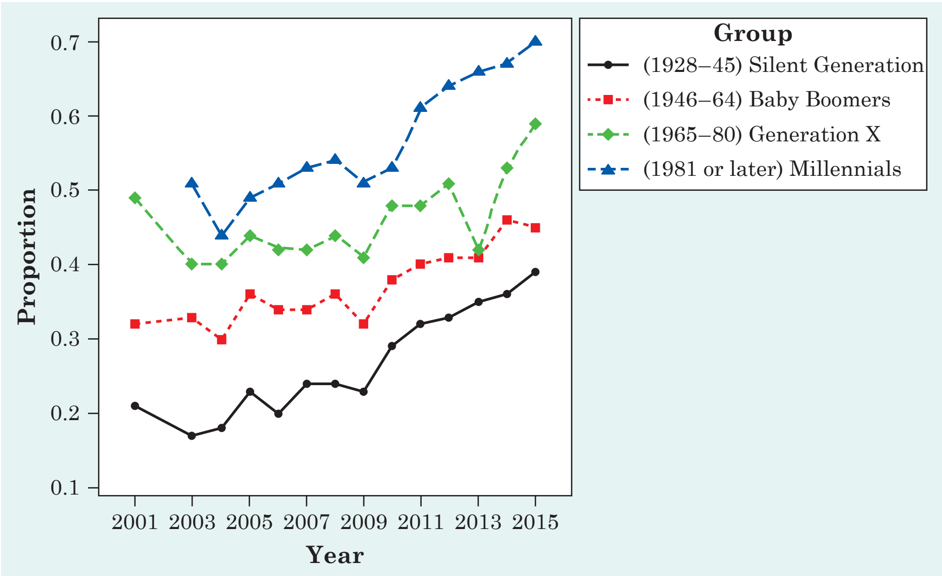


Figure 10.14 Changing attitudes on same-sex marriage by generational cohort, Exercise 10.13. (Data from Pew Research Center)

murders were committed with handguns, 18.4% with other firearms, 12.1% with knives or other cutting implements, 5.8% with a part of the body (usually the hands, fists, or feet), and 4.4% with blunt objects. Make a graph to display these data. Do you need an "other methods" category? Why?

10.15 The cost of imported oranges. Figure 10.15 is a line graph of the average cost of imported oranges each month from July 1995 to April 2012. These data are the price in U.S. dollars per metric ton.

(a) The graph shows strong seasonal variation. How is this visible in the graph? Why would you expect the price of oranges to show seasonal variation?

(b) What is the overall trend in orange prices during this period, after we take account of the seasonal variation?

10.16 College freshmen. A survey of college freshmen in 2007 asked what field they planned to study. The results: 12.8%, arts and humanities; 17.7%, business; 9.2%, education; 19.3%, engineering, biological sciences, or physical sciences; 14.5%, professional; and 11.1%, social science.

(a) What percentage of college freshmen plan to study fields other than those listed?

(b) Make a graph that compares the percentages of college freshmen planning to study various fields.

10.17 Decreasing trend? A common statistic considered by administration at universities is credit hour production. This is calculated by taking the credit hours for a course taught times the student enrollment. For example, a faculty member teaching two 3-credit courses with 50 students each would have produced $6 \times 50 = 300$ credit hours. Suppose you are the chair for the department of statistics and you give Figure 10.16 to your dean. The dean asks you to explain the disturbing decreasing trend. How should you respond?

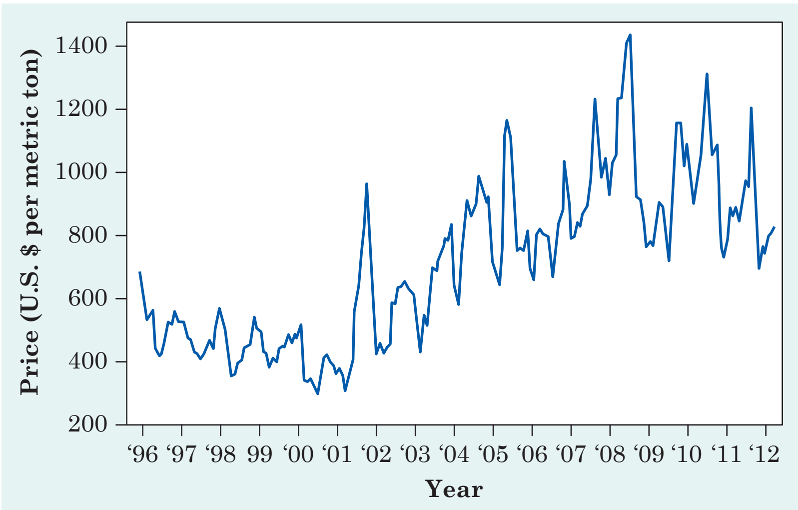


Figure 10.15 The price of oranges, July 1995 to April 2012, Exercise 10.15.

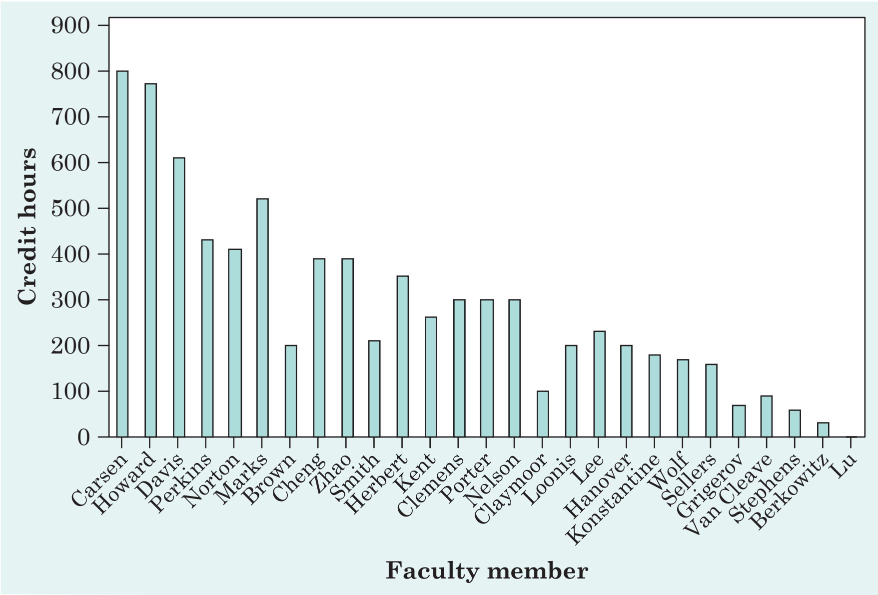


Figure 10.16 Bar graph depicting credit hours produced for each faculty member in a department, Exercise 10.17.

10.18 Civil disorders. The years around 1970 brought unrest to many U.S. cities. Here are government data on the number of civil disturbances in each three-month period during the years 1968 to 1972.

(a) Make a line graph of these data.

(b) The data show both a longer-term trend and seasonal variation within years. Describe the nature of both patterns. Can you suggest an explanation for the seasonal variation in civil disorders?

Period	Count	Period	Count
1968, Jan.–Mar.	6	1970, July–Sept.	20
Apr.–June	46	Oct.–Dec.	6
July–Sept.	25	1971, Jan.–Mar.	12
Oct.–Dec.	3	Apr.–June	21
1969, Jan.–Mar.	5	July–Sept.	5
Apr.–June	27	Oct.–Dec.	1
July–Sept.	19	1972, Jan.–Mar.	3
Oct.–Dec.	6	Apr.–June	8
1970, Jan.–Mar.	26	July–Sept.	5
Apr.–June	24	Oct.–Dec.	5

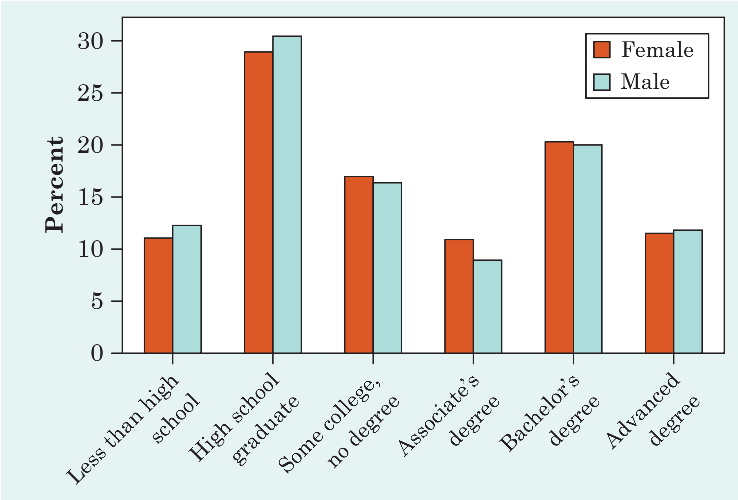


Figure 10.17 A side-by-side bar graph of educational attainment by sex, for those aged 25 and older, Exercise 10.19. (Data collected in the 2014 Current Population Survey)

10.19 Educational attainment by sex. Figure 10.17 is a side-by-side bar graph of educational attainment, by sex, for those aged 25 and older. These data were collected in the 2014 Current Population Survey. Compare educational attainment for males and females. Comment on any patterns you see.

10.20 A bad graph? Figure 10.18 shows a graph that appeared in the *Lexington (Ky.) Herald-Leader* on October 5, 1975. Discuss the correctness of this graph.

10.21 Seasonal variation. You examine the average temperature in Chicago each month for many years. Do you expect a line graph of the data to show seasonal variation? Describe the kind of seasonal variation you expect to see.

10.22 Sales are up. The sales at your new gift shop in December are double the November value. Should you conclude that your shop is growing more popular and will soon make you rich? Explain your answer.

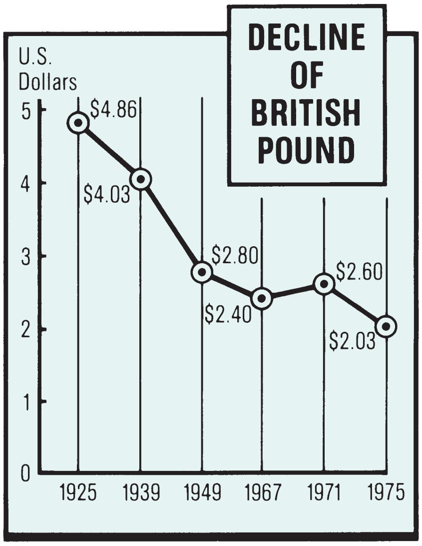


Figure 10.18 A newspaper's graph of the value of the British pound, Exercise 10.20. (Data from *Lexington (Ky.) Herald Leader*, Associated Press, 1975)

10.23 Counting employed people. A news article says:

The report that employment plunged in June, with nonfarm payrolls declining by 117,000, helped to persuade the Federal Reserve to cut interest rates yet again.... In reality, however, there were 457,000 more people employed in June than in May.

What explains the difference between the fact that employment went up by 457,000 and the official report that employment went down by 117,000?

10.24 The sunspot cycle. Some plots against time show **cycles** of up-and-down movements. Figure 10.19 is a line graph of the average number of sunspots on the sun's visible face for each month from 1900 to 2011. What is the approximate length of the sunspot cycle? That is, how many years are there between the successive valleys in the graph? Is there any overall trend in the number of sunspots?

10.25 Trucks versus cars. Do consumers prefer trucks, SUVs, and minivans to passenger cars? Here are data on sales and leases of new cars and trucks in the United States. (The definition of "truck" includes SUVs and minivans.) Plot two line graphs on the same axes to compare the change in car and truck sales over time. Describe the trend that you see.

Year:	1981	1983	1985
Cars (1000s):	8,536	9,182	11,042
Trucks (1000s):	2,260	3,129	4,682

Year:	1987	1989	1991
Cars (1000s):	10,277	9,772	8,589
Trucks (1000s):	4,912	4,941	4,136

Year:	1993	1995	1997	1999	2001
Cars (1000s):	8,518	8,636	8,273	8,697	8,422
Trucks (1000s):	5,654	6,469	7,217	8,704	9,046

Year:	2003	2005	2007	2009
Cars (1000s):	7,615	7,720	7,618	5,456
Trucks (1000s):	9,356	9,725	8,842	5,145

10.26 Who sells cars? Figure 10.20 is a pie chart of the percentage of passenger car sales in 1997 by various manufacturers. The artist has tried to make the graph attractive by using the wheel of a car for the "pie." Is the graph still a correct display of the data? Explain your answer.

10.27 Who sells cars? Make a bar graph of the data in Exercise 10.26. What advantage does your new graph have over the pie chart in Figure 10.20?

10.28 The Border Patrol. Here are the numbers of deportable aliens caught by the U.S. Border Patrol for 1971 through 2009. Display these data in a graph. What are the most important facts that the data show?

Year:	1971	1973	1975	1977	1979
Count (1000s):	420	656	767	1042	1076

Year:	1981	1983	1985	1987	1989
Count (1000s):	976	1251	1349	1190	954

Year:	1991	1993	1995	1997	1999
Count (1000s):	1198	1327	1395	1536	1714

Year:	2001	2003	2005	2007	2009
Count (1000s):	1266	932	1189	877	556

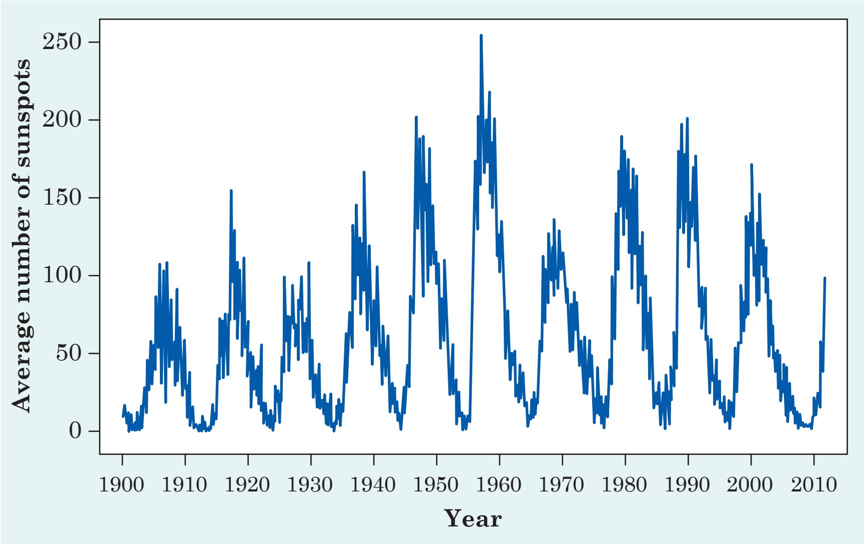


Figure 10.19 The sunspot cycle, Exercise 10.24. This is a line graph of the average number of sunspots per month for the years 1900–2011. (Data from the National Oceanic and Atmospheric Administration)

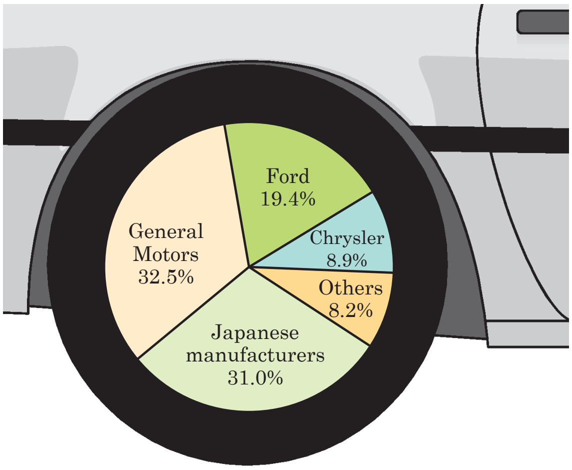


Figure 10.20 Passenger car sales by several manufacturers in 1997, Exercise 10.26.

10.29 Bad habits. According to the National Household Survey on Drug Use and Health, when asked in 2012, 41% of those aged 18 to 24 years used cigarettes in the past year, 9% used smokeless tobacco, 36.3% used illicit drugs, and 10.4% used pain relievers or sedatives. Explain why it is *not* correct to display these data in a pie chart.

10.30 Accidental deaths. In 2011 there were 130,557 deaths from unintentional injury in the United States. Among these were 38,851 deaths from poisoning, 33,804 from motor vehicle accidents, 30,208 from falls, 6,601 from suffocation, and 3,391 from drowning.

(a) Find the percentage of accidental deaths from each of these causes, rounded to the nearest percent. What percentage of accidental deaths were due to other causes?
(b) Make a well-labeled graph of the distribution of causes of accidental deaths.

10.31 Yields of money market funds. Many people invest in money market funds. These are mutual funds that attempt to maintain a constant price of $1 per share while paying monthly interest. Table 10.3 gives the average annual interest rates (in percent) paid by all taxable money market funds from 1973 (the first full year in which such funds were available) to 2008.

(a) Make a line graph of the interest paid by money market funds for these years.

(b) Interest rates, like many economic variables, show **cycles**, clear but repeating up-and-down movements. In which years did the interest rate cycle reach temporary peaks?

(c) A plot against time may show a consistent **trend** underneath cycles. When did interest rates reach their overall peak during these years? Describe the general trend downward since that year.

10.32 The Boston Marathon. Women were allowed to enter the Boston Marathon in 1972. The time (in minutes, rounded to the nearest minute)

TABLE 10.3 Average annual interest rates (in percent) paid by money market funds, 1973–2008

Year	Rate	Year	Rate	Year	Rate	Year	Rate
1973	7.60	1982	12.23	1991	5.71	2000	5.89
1974	10.79	1983	8.58	1992	3.36	2001	3.67
1975	6.39	1984	10.04	1993	2.70	2002	1.29
1976	5.11	1985	7.71	1994	3.75	2003	0.64
1977	4.92	1986	6.26	1995	5.48	2004	0.82
1978	7.25	1987	6.12	1996	4.95	2005	2.66
1979	10.92	1988	7.11	1997	5.10	2006	4.51
1980	12.68	1989	8.87	1998	5.04	2007	4.70
1981	16.82	1990	7.82	1999	4.64	2008	2.05

Source: Albert J. Fredman, "A closer look at money market funds," *American Association of Individual Investors Journal*, February 1997, pp. 22–27; and the 2010 *Statistical Abstract of the United States*.

TABLE 10.4 Women's winning times (minutes) in the Boston Marathon, 1972–2015

Year	Time	Year	Time	Year	Time	Year	Time
1972	190	1983	143	1994	142	2005	145
1973	186	1984	149	1995	145	2006	144
1974	167	1985	154	1996	147	2007	149
1975	162	1986	145	1997	146	2008	145
1976	167	1987	146	1998	143	2009	152
1977	168	1988	145	1999	143	2010	146
1978	165	1989	144	2000	146	2011	143
1979	155	1990	145	2001	144	2012	152
1980	154	1991	144	2002	141	2013	146
1981	147	1992	144	2003	145	2014	139
1982	150	1993	145	2004	144	2015	145

Source: See the website http://en.wikipedia.org/wiki/List_of_winners_of_the_Boston_Marathon/.

for each winning woman from 1972 to 2015 appears in Table 10.4.

(a) Make a graph of the winning times.

(b) Give a brief description of the pattern of Boston Marathon winning times over these years. Have times stopped improving in recent years?

EXPLORING THE WEB

Follow the QR code to access exercises.