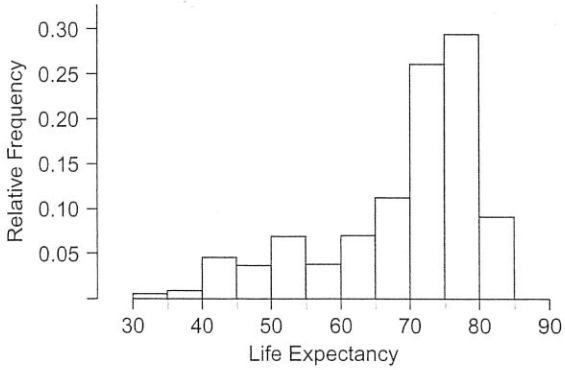
Relative Frequency of Life Expectancies

Example 2.3

Display 2.17 shows the relative frequency distribution of life expectancies for 223 countries around the world. How many countries have a life expectancy of at least 75 but less than 80 years? Give the proportion of countries that have a life expectancy of 70 years or more.

Elderly Tibetan couple.



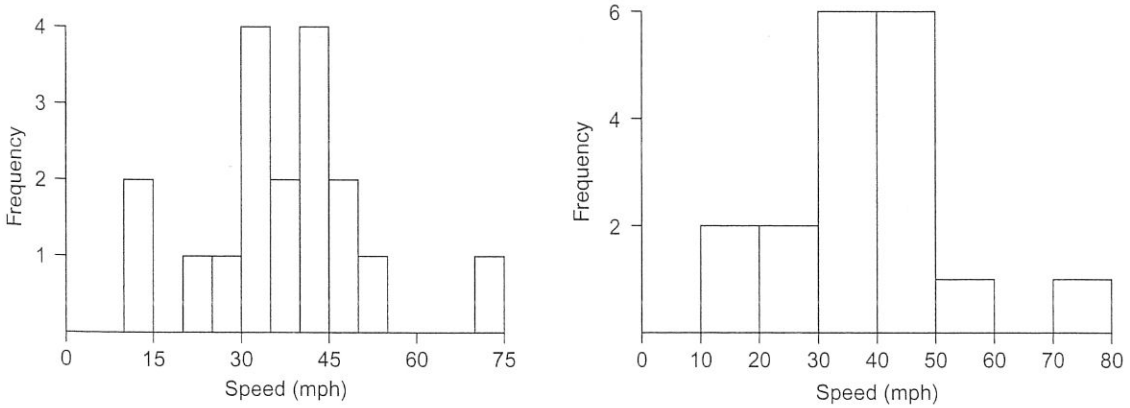
Display 2.17 Life expectancies of people by country. [Source: *www.cia.gov/library/publications/the-world-factbook/rankorder/2102rank.html.*]

Solution

The bar including 75 years and up to 80 years has a relative frequency of about 0.29, so the number of countries with a life expectancy of at least 75 years but less than 80 years is about 0.29 • 223, or approximately 65.

The proportion of countries with life expectancy of 70 years or greater is the sum of the heights of the three bars to the right of 70—about 0.26 + 0.29 + 0.08, or 0.63.

Changing the width of the bars in your histogram can sometimes change your impression of the shape of the distribution, especially when there are few values. The two histograms in Display 2.18 show the distribution of the speeds of a sample of mammals (from the table in Display 2.14 on page 31). The shapes appear somewhat different as the two peaks disappear when the histogram has wider bars. There is no "right" answer to the question of which bar width is best, just as there is no rule that tells a photographer when to use a zoom lens for a close-up.

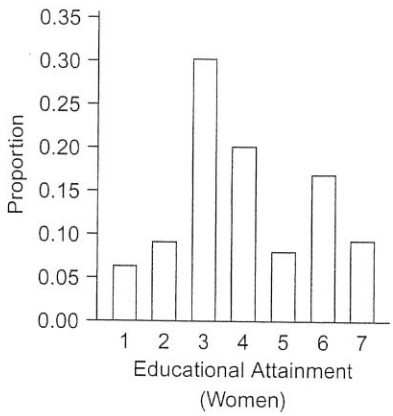


Display 2.18 Speeds of mammals using two different bar widths.

that there were about three times as many domesticated mammals as wild mammals in this sample. Note that the bars are separated so that there is no suggestion that the variable can take on a value of, say, 0.5.

Bar charts, like histograms, can be scaled in terms of frequencies or relative frequencies. Display 2.24 shows the proportion (relative frequency) of the female labor force age 25 and older in the United States who fall into various educational categories. The educational categories have a natural ordering from least education to most and are coded 1 through 9:

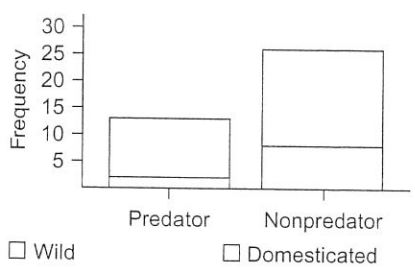
1. less than 9th grade
2. 9th to 12th grade, no diploma
3. high school graduate (includes equivalency)
4. some college, no degree
5. associate's degree
6. bachelor's degree
7. graduate or professional degree



Display 2.24 The female labor force age 25 and older by educational attainment. [Source: U.S. Census Bureau, *March 2007 Current Population Survey*, *www.census.gov*.]

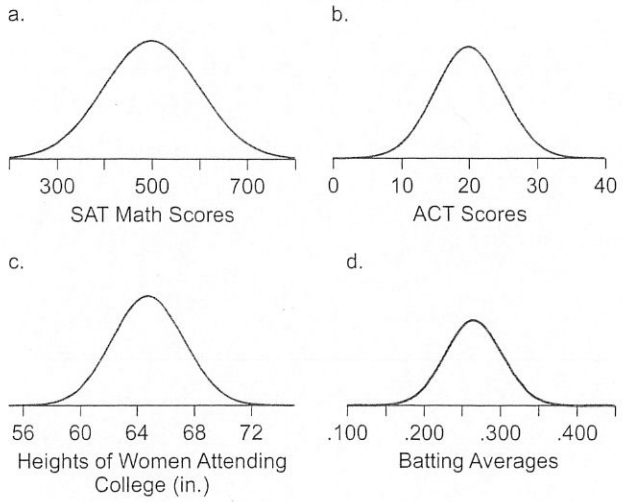
Because the ordering of categories in a bar chart is often arbitrary and the names of the categories need not be numbers, it makes little sense to talk about center and spread. But often it does make sense to talk about the **modal category**—the category with the highest frequency. More women in the labor force fall into the "high school graduate" category than into any other category.

Bar charts can be segmented to display two categorical variables on the same plot. In Display 2.25, the bars representing predators and nonpredators each are segmented into wild and domestic categories. The **segmented bar chart** makes it clear, for example, that there are more wild animals among the nonpredators than among the

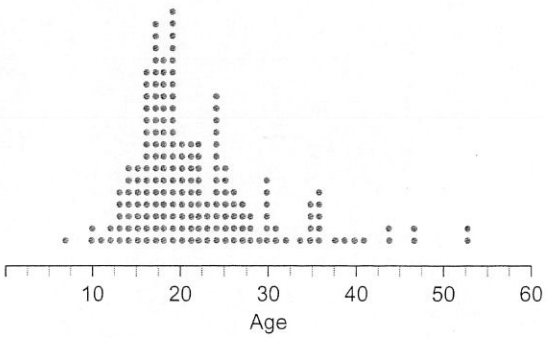


Display 2.25 A segmented bar chart of predator and nonpredator mammals.

"A typical SAT math score is roughly [mean], give or take [standard deviation] or so."



P4. The dot plot in Display 2.29 gives the ages of people who died at rock concerts over a 12-year period, most of them crushed by the crowd. Estimate the median and quartiles of the distribution. Then write a verbal summary of the distribution.

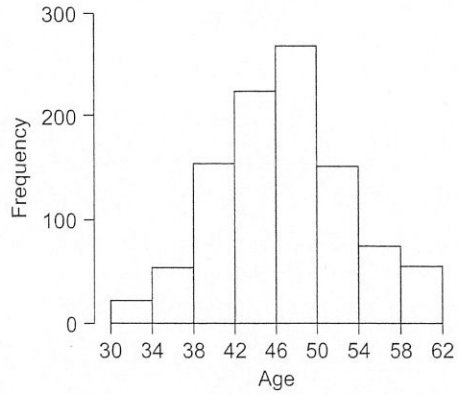


Display 2.29 Ages of people who died at rock concerts. [Source: Crowd Management Strategies, *www.crowdsafe.com/thewall.html.*]

Graphical Displays of Distributions

P5. Refer to the gestation periods of the mammals listed in Display 2.14 on page 31.
   a. Make a dot plot of these gestation periods.
   b. Write a sentence summarizing the shape, center, and spread of this distribution.
   c. What kinds of mammals have longer gestation periods?

P6. The histogram in Display 2.30 gives the ages of a sample of 1000 people.
   a. Describe the shape, center, and spread of this distribution.
   b. Convert the histogram into a relative frequency histogram.
   c. About what proportion of the people are age 50 or older?



Display 2.30 Ages of 1000 people.

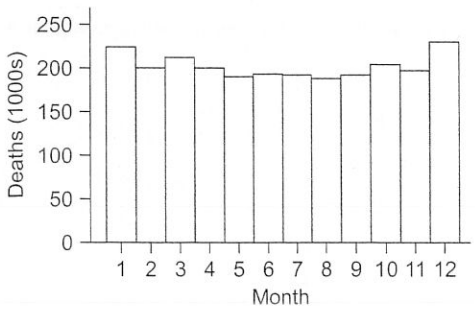
P7. Refer to the relative frequency histogram of life expectancy in countries around the world in Display 2.17 on page 33.
   a. Estimate the proportion of countries with a life expectancy of less than 50 years.
   b. Estimate the *number* of countries with a life expectancy of less than 50 years.
   c. Describe the shape, center, and spread of this distribution.

P8. Refer to the table in Display 2.14 on page 31.
   a. Make a back-to-back stemplot of the average longevities and maximum longevities.
   b. Describe how the distributions differ in terms of shape, center, and spread.
   c. Why do the differences occur?

P9. Using the technology available to you, make histograms of the average longevity and maximum longevity data in Display 2.14 on page 31, using bar widths of 4, 8, and 16 years. Comment on the main features of the shapes of these distributions. Which bar width appears to display these features best?

Bar Charts for Categorical Data

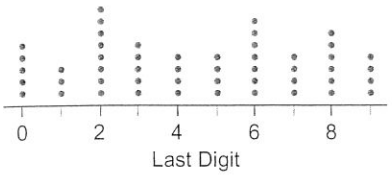
P10. The plot in Display 2.31 gives the number of deaths in the United States per month in 2007, with January coded as 1, February as 2, and so on. Does the number of deaths appear to be uniformly distributed over the months? Give a verbal summary of the way deaths are distributed over the months of the year.



Display 2.31 Deaths, in thousands, per month, 2007. [Source: *www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57_06.html.*]

P11. Suppose you collect this information for each student in your class: age, hair color, number of siblings, gender, and miles he or she lives from school. What are the cases? What are the variables? Classify each variable as quantitative or categorical.

P12. The plot in Display 2.32 shows the last digit of the Social Security numbers of the students in a statistics class. Describe this distribution.

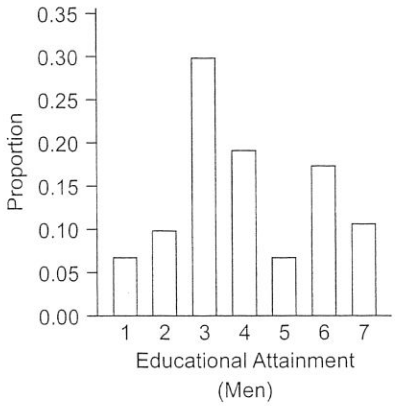


Display 2.32 Last digit of a sample of Social Security numbers.

P13. Display 2.33, which gives the educational attainment of the male labor force, is the counterpart of Display 2.24 on page 37.

a. What are the cases, and what is the variable?

b. Describe the distribution you see here.



Display 2.33 The male labor force age 25 years and older by educational attainment. [Source: U.S. Census Bureau, March 2007 Current Population Survey, *www.census.gov*.]

c. How does the distribution of female education compare to the distribution of male education?

d. Why is it better to look at relative frequency bar charts rather than frequency bar charts to make this comparison?

Exercises

Exercises are mixed in difficulty—some are like the more routine practice problems while others require original thought and understanding of several concepts. Also unlike practice problems, exercises are not necessarily in the order that the concepts were introduced in the section. Each odd-numbered exercise is followed by a similar even-numbered exercise, should you want more practice. Answers to the odd-numbered exercises are given in the back of the book.

E1. Using your knowledge of the variables and what you think the shape of the distribution might be, match each variable in this list with the appropriate histogram in Display 2.34.

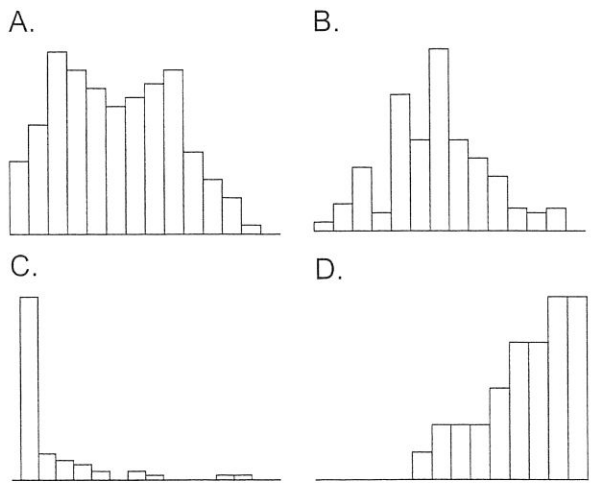
i. scores on a fairly easy examination in statistics

ii. heights of a group of mothers and their 12-year-old daughters

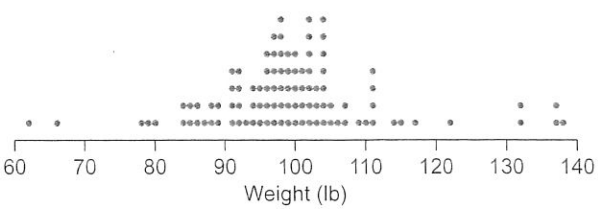
iii. numbers of medals won by medal-winning countries in the 2008 Summer Olympics

iv. weights of grown hens in a barnyard

E2. The distribution in Display 2.35 shows measurements of the strength in pounds of 22s yarn (22s refers to a standard unit for measuring yarn strength). What is the basic shape of this distribution? What feature makes it uncharacteristic of distributions with that shape?



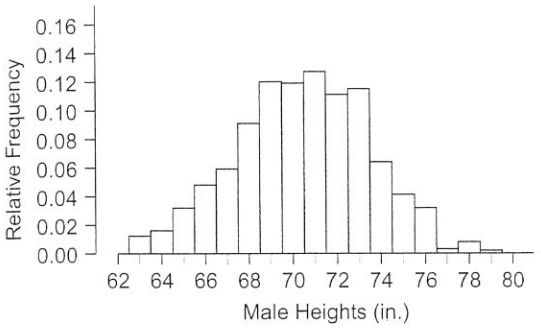
Display 2.34 Four histograms with different shapes.



Display 2.35 Strength of yarn. [Source: Data and Story Library at Carnegie-Mellon University, *lib.stat.cmu.edu*.]

lie between 12 and 34. Half of the values are above 18 and half are below. Sketch what the distribution might look like. [Source: Superfund Site Information, U.S. Environmental Protection Agency, 2009, *www.epa.gov/superfund/sites.*]

E9. Display 2.38 shows the distribution of the heights of U.S. males between the ages of 18 and 24. The heights are rounded to the nearest inch.

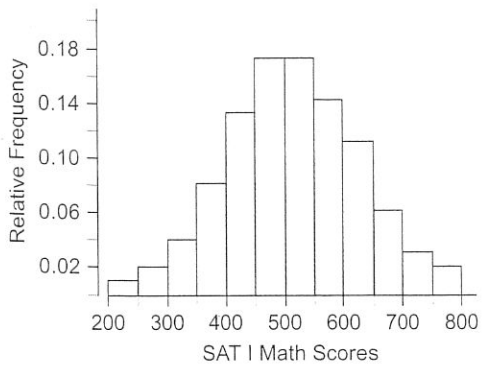


Display 2.38 Heights of males, ages 18 to 24. [Source: U.S. Census Bureau, *Statistical Abstract of the United States, 2009*, Table 201.]

a. Draw a smooth curve to approximate the histogram.
b. Without doing any computing, estimate the mean and standard deviation.
c. Estimate the proportion of men ages 18 to 24 who are 74 in. tall or less.
d. Estimate the proportion of heights that fall below 68 in.
e. Why should you say that the distribution of heights is "approximately" normal rather than simply saying that it is normally distributed?

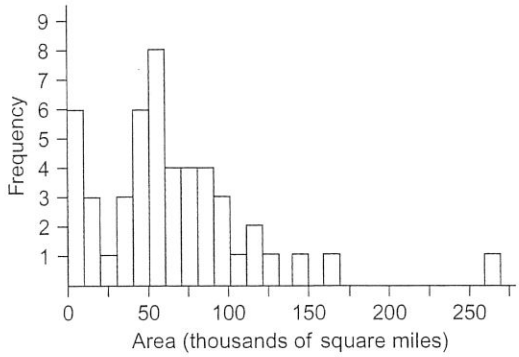
E10. The histogram in Display 2.39 shows the distribution of SAT I math scores.

a. Without doing any computing, estimate the mean and standard deviation.
b. Roughly what percentage of the SAT I math scores would you estimate are within one standard deviation of the mean?
c. For SAT I critical reading scores, the shape was similar, but the mean was 10 points lower and the standard deviation was 2 points smaller. Draw a smooth curve to show the distribution of SAT I critical reading scores.



Display 2.39 Relative frequency histogram of SAT I math scores, 2004–2005. [Source: College Board Online, *www.collegeboard.org.*]

E11. The table in Display 2.41 provides the area, the population, and population density of the U.S. states. The histogram in Display 2.40 shows the areas of the states. It does not include Alaska because Alaska is so large compared to the other states that it doesn't fit on the plot.



Display 2.40 Area of the U.S. states, excluding Alaska.

a. The distribution has two peaks. What simple geographic factor could help explain this?
b. Split the states into two groups according to that factor and use the technology available to you to make a plot of areas for each group. Do two peaks appear in each?

State	Area (sq mi)	Population (thousands)	Density (people/ sq mi)
Alabama	52,419	4,662	88.9
Alaska	663,267	686	1.0
Arizona	113,998	6,500	57.0
Arkansas	53,179	2,855	53.7
California	163,696	36,757	224.5
Colorado	104,094	4,939	47.4
Connecticut	5,543	3,501	631.6
Delaware	2,489	873	350.7
Florida	65,755	18,328	278.7
Georgia	59,425	9,686	163.0
Hawaii	10,931	1,288	117.8
Idaho	83,570	1,524	18.2
Illinois	57,914	12,902	222.8
Indiana	36,418	6,377	175.1
Iowa	56,272	3,003	53.4
Kansas	82,277	2,802	34.1
Kentucky	40,409	4,269	105.6
Louisiana	51,840	4,411	85.1
Maine	35,385	1,316	37.2
Maryland	12,407	5,634	454.1

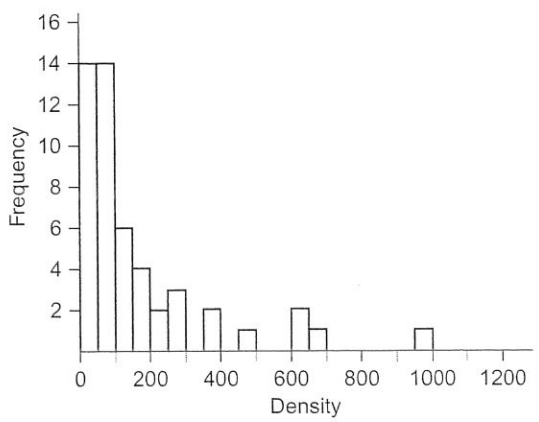
Display 2.41 *(Continued on next page)*

State	Area (sq mi)	Population (thousands)	Density (people/ sq mi)
Massachusetts	10,555	6,498	615.6
Michigan	96,716	10,003	103.4
Minnesota	86,939	5,220	60.0
Mississippi	48,430	2,939	60.7
Missouri	69,704	5,912	84.8
Montana	147,042	967	6.6
Nebraska	77,354	1,783	23.0
Nevada	110,561	2,600	23.5
New Hampshire	9,350	1,316	140.7
New Jersey	8,721	8,683	995.6
New Mexico	121,590	1,984	16.3
New York	54,556	19,490	357.2
North Carolina	53,819	9,222	171.4
North Dakota	70,700	641	9.1
Ohio	44,825	11,486	256.2
Oklahoma	69,898	3,642	52.1
Oregon	96,381	3,790	38.5
Pennsylvania	46,055	12,448	270.3
Rhode Island	1,545	1,051	680.3
South Carolina	32,020	4,480	139.9
South Dakota	77,117	804	10.4
Tennessee	42,143	6,215	147.5
Texas	268,581	24,327	90.6
Utah	84,899	2,736	32.2
Vermont	9,614	621	64.6
Virginia	42,774	7,769	181.6
Washington	71,300	6,549	91.9
West Virginia	24,230	1,814	74.9
Wisconsin	65,498	5,628	85.9
Wyoming	97,814	533	5.4

Display 2.41 Area and population of U.S. states. [Source: U.S. Census, *State and Metropolitan Handbook, 2009.*]

E12. Refer to the table in Display 2.41 of E11. The histogram in Display 2.42 shows the population densities of all 50 states.

a. Show how the population density for Vermont was computed.



Display 2.42 Population density (people per square mile) of the U.S. states.

b. Which states are outliers?

c. Is Alaska an outlier for this variable? How can you tell from the table? From the plot?

E13. How do countries compare with respect to the value of the goods they produce? Display 2.43 shows gross domestic product (GDP) per capita, a measure of the total value of all goods and services produced divided by the number of people in a country, and the average number of people per room in housing units, a measure of crowdedness, for a selection of countries in Asia, Europe, and North America.

A dot plot of the per capita GDP by country is shown in Display 2.44.

a. How would you describe this distribution?

b. Which two countries have the highest per capita GDP? Do they appear to be outliers?

Country	Per Capita GDP (U.S. $)	Average Number of People per Room
Austria	44,652	0.7
Azerbaijan	3,691	2.1
Belgium	43,470	0.6
Bulgaria	5,178	1.0
Canada	43,368	0.5
China	2,604	1.1
Croatia	11,256	1.2
Cyprus	27,465	0.6
Czech Republic	16,881	1.0
Finland	46,371	0.8
France	40,090	0.7
Germany	40,162	0.5
Hungary	13,777	0.8
India	976	2.7
Iraq	2,404	1.5
Israel	23,383	1.2
Japan	34,225	0.8
Korea, Republic of	19,841	1.1
Kuwait	38,574	1.7
Netherlands	46,669	0.7
Norway	82,465	0.6
Pakistan	996	3.0
Poland	11,008	1.0
Portugal	20,990	0.7
Romania	7,523	1.3
Serbia-Montenegro	5,383	1.2
Slovakia	13,702	1.2
Sri Lanka	1,676	2.2
Sweden	49,873	0.5
Switzerland	56,579	0.6
Syria	1,883	2.0
Turkey	6,511	1.3
United Kingdom	46,549	0.5
United States	46,047	0.5

Display 2.43 Per capita GDP and crowdedness for a selection of countries. [Source: United Nations, *unstats.un.org.*]