

**Table 2-9** Colleges of Undergraduates

College	Relative Frequency
Public 2-Year	36.8%
Public 4-Year	40.0%
Private 2-Year	1.6%
Private 4-Year	21.9%

**EXAMPLE 6**

**College Undergraduate Enrollments** Table 2-9 shows the distribution of undergraduate college student enrollments among the four categories of colleges (based on data from the U.S. National Center for Education Statistics). The sum of the relative frequencies is 100.3%, which is slightly different from 100% because of rounding errors.

**EXAMPLE 7**

**Education and Smoking: Frequency Distribution?** Table 2-10 is a type of table commonly depicted in media reports, but it is *not* a relative frequency distribution. (Table 2-10 is based on data from the Centers for Disease Control and Prevention.) The definition of a frequency distribution given earlier requires that the table shows how a data set is distributed among all of several categories, but Table 2-10 does not show how the population of smokers is distributed among the different education categories. Instead, Table 2-10 shows the percentage of smokers in each of the different categories. Also, the sum of the frequencies in Table 2-10 is 157%, which is clearly different from 100%, even after accounting for any rounding errors. Table 2-10 has value for conveying important information, but it is not a frequency distribution.

**Table 2-10** Education and Smoking

Education	Percentage Who Smoke
0-12 (no diploma)	26%
GED diploma	43%
High school graduate	25%
Some college	23%
Associate degree	21%
Bachelor's degree	12%
Graduate degree	7%

**Table for Exercise 3**

Downloaded Material	Percent
Music	32%
Games	25%
Software	14%
Movies	10%

**Table for Exercise 4**

Height (in.)	Frequency
35-39	6
40-44	31
45-49	67
50-54	21
55-59	0
60-64	0
65-69	6
70-74	10

## 2-2 Basic Skills and Concepts

### Statistical Literacy and Critical Thinking

- 1. Frequency Distribution** Table 2-7 on page 51 is a frequency distribution summarizing the weights of 72 different pennies. Is it possible to identify the original list of the 72 individual weights from Table 2-7? Why or why not?
- 2. Relative Frequency Distribution** After constructing a relative frequency distribution summarizing IQ scores of college students, what should be the sum of the relative frequencies?
- 3. Unauthorized Downloading** A Harris Interactive survey involved 1644 people between the ages of 8 years and 18 years. The accompanying table summarizes the results. Does this table describe a relative frequency distribution? Why or why not?
- 4. Analyzing a Frequency Distribution** The accompanying frequency distribution summarizes the heights of a sample of people at Vassar Road Elementary School. What can you conclude about the people included in the sample?

*In Exercises 5–8, identify the class width, class midpoints, and class boundaries for the given frequency distribution. The frequency distributions are based on data from Appendix B.*

5. Tar (mg) in Nonfiltered Cigarettes

Class	Frequency
10–13	1
14–17	0
18–21	15
22–25	7
26–29	2

6. Tar (mg) in Filtered Cigarettes

Class	Frequency
2–5	2
6–9	2
10–13	6
14–17	15

7. Weights (lb) of Discarded Metal

Class	Frequency
0.00–0.99	5
1.00–1.99	26
2.00–2.99	15
3.00–3.99	12
4.00–4.99	4

8. Weights (lb) of Discarded Plastic

Class	Frequency
0.00–0.99	14
1.00–1.99	20
2.00–2.99	21
3.00–3.99	4
4.00–4.99	2
5.00–5.99	1

**Critical Thinking.** *In Exercises 9–12, answer the given questions that relate to Exercises 5–8.*

**9. Identifying the Distribution** Using a strict interpretation of the relevant criteria on page 50, does the frequency distribution given in Exercise 5 appear to have a normal distribution? Does the distribution appear to be normal if the criteria are interpreted very loosely?

**10. Identifying the Distribution** Using a strict interpretation of the relevant criteria on page 50, does the frequency distribution given in Exercise 6 appear to have a normal distribution? Does the distribution appear to be normal if the criteria are interpreted very loosely?

**11. Comparing Relative Frequencies** Construct one table (similar to Table 2-8 on page 51) that includes relative frequencies based on the frequency distributions from Exercises 5 and 6, then compare the amounts of tar in nonfiltered and filtered cigarettes. Do the cigarette filters appear to be effective?

**12. Comparing Relative Frequencies** Construct one table (similar to Table 2-8 on page 51) that includes relative frequencies based on the frequency distributions from Exercises 7 and 8, then compare the weights of discarded metal and plastic. Do those weights appear to be about the same or are they substantially different?

*In Exercises 13 and 14, construct the cumulative frequency distribution that corresponds to the frequency distribution in the exercise indicated.*

13. Exercise 5

14. Exercise 6

*In Exercises 15 and 16, use the given qualitative data to construct the relative frequency distribution.*

**15. Titanic Survivors** The 2223 people aboard the *Titanic* include 361 male survivors, 1395 males who died, 345 female survivors, and 122 females who died.

**16. Smoking Treatments** In a study, researchers treated 570 people who smoke with either nicotine gum or a nicotine patch. Among those treated with nicotine gum, 191 continued to smoke and the other 59 stopped smoking. Among those treated with a nicotine patch, 263 continued to smoke and the other 57 stopped smoking (based on data from the Centers for Disease Control and Prevention).



**28. Weights of Quarters** Refer to Data Set 20 in Appendix B and use the weights (grams) of the post-1964 quarters. Construct a frequency distribution. Begin with a lower class limit of 5.5000 g, and use a class width of 0.0500 g. Compare the frequency distribution to the result from Exercise 27.

**29. Blood Groups** Listed below are blood groups of O, A, B, and AB of randomly selected blood donors (based on data from the Greater New York Blood Program). Construct a table summarizing the frequency distribution of these blood groups.

O A B O O O O O A B O O O O B O B O A A A O A A B A B  
A B A A A A O A O O A A O O A O O O O A A A A A B

**30. Train Derailments** An analysis of 50 train derailment incidents identified the main causes listed below, where T denotes bad track, E denotes faulty equipment, H denotes human error, and O denotes other causes (based on data from the Federal Railroad Administration). Construct a table summarizing the frequency distribution of these causes of train derailments.

T T T E E H H H H H O O H H H E E T T T E T H O T  
T T T T T T H T T H E E T T E E T T T H T T O O O

## 2-2 Beyond the Basics

**31. Interpreting Effects of Outliers** Refer to Data Set 21 in Appendix B for the axial loads of aluminum cans that are 0.0111 in. thick. The load of 504 lb is an *outlier* because it is very far away from all of the other values. Construct a frequency distribution that includes the value of 504 lb, then construct another frequency distribution with the value of 504 lb excluded. In both cases, start the first class at 200 lb and use a class width of 20 lb. State a generalization about the effect of an outlier on a frequency distribution.

**32. Number of Classes** According to Sturges's guideline, the ideal number of classes for a frequency distribution can be approximated by  $1 + (\log n)/(\log 2)$ , where  $n$  is the number of data values. Use this guideline to complete the table in the margin.

**Table for Exercise 32**

Number of Data Values	Ideal Number of Classes
16–22	5
23–45	6
?	7
?	8
?	9
?	10
?	11
?	12

## 2-3 Histograms

**Key Concept** In Section 2-2 we introduced the frequency distribution as a tool for summarizing a large data set and determining the distribution of the data. In this section we discuss a visual tool called a *histogram*, and its significance in representing and analyzing data. Because many statistics computer programs and calculators can automatically generate histograms, it is not so important to master the mechanical procedures for constructing them. Instead we focus on the information we can obtain from a histogram. Namely, we use a histogram to analyze the shape of the distribution of the data.

### DEFINITION

A **histogram** is a graph consisting of bars of equal width drawn adjacent to each other (without gaps). The horizontal scale represents classes of quantitative data values and the vertical scale represents frequencies. The heights of the bars correspond to the frequency values.

A histogram is basically a graphic version of a frequency distribution. For example, Figure 2-3 on page 56 shows the histogram corresponding to the frequency distribution in Table 2-2 on page 47.