©iStockphoto.com/Rawpixel

2 SUMMARIZING DATA

Frequency Distributions in Tables and Graphs

LEARNING OBJECTIVES

After reading this chapter, you should be able to

LO 2.1	Define frequency and explain why it is valuable to summarize data.
LO 2.2	Construct a simple frequency distribution for grouped data.
LO 2.3	Construct and explain when it is appropriate to use a cumulative frequency, relative frequency, relative percentage, cumulative relative frequency, and cumulative percentage distribution.
LO 2.4	Identify percentile points and percentile ranks in a cumulative percentage distribution.
SPSS LO 2.5	Construct a frequency distribution for quantitative data using SPSS.
LO 2.6	Construct a frequency distribution for ungrouped or categorical data.
SPSS LO 2.7	Construct a frequency distribution for categorical data using SPSS.
LO 2.8	Identify how presenting images can meaningfully illustrate data.
LO 2.9	Construct and interpret graphs for distributions of continuous data.
LO 2.10	Summarize discrete or continuous data in a stem-and-leaf display.
LO 2.11	Construct and interpret graphs for distributions of discrete or categorical data.
SPSS LO 2.12	Construct a histogram, bar chart, pie chart, and stem-and-leaf display using SPSS.

Often, the types of data that interest you are simple counts. In other words, you usually ask questions that relate to how frequently something of interest occurs. An employer may ask, "When do most employees call off from work?" to decide about staffing; a college student may ask, "How often does an instructor receive positive ratings from their students?" to decide whether or not to enroll in a class; a frequent flyer may ask, "How frequently do popular airlines have flight delays?" to decide which airline to travel with. In each example, the decision being made is based upon how frequently something occurs.

Likewise, frequencies have a big place in science. For example, many aspects of behavior are naturally categorical, such as obesity and mental health, and scientists are concerned with how many people fall into categories that can range from "normal" to higher levels of severity or risk. Behaviors in educational or school settings often relate to rankings in which students are assigned to various levels of ability from high to low achieving. In workplace settings, employees are often rated on their job performance, and the number of employees meeting versus below standards may be counted to track the quality of employee performance across an organization. Likewise, general demographic data, such as sex and race, are often recorded in research studies and reported to disclose the types of participants observed. In this way, frequency data are often utilized in science, and it is therefore useful to have a set of procedures to simplify the presentation of how often events and behaviors occur.

Statistics provides a useful way to accomplish the summary of frequencies, typically by summarizing them in tables or graphs. In this chapter, we will introduce many ways in which researchers summarize, organize, and make sense of frequency data to appropriately construct and accurately interpret many of the tables and graphs used to summarize data in behavioral research.

2.1 WHY SUMMARIZE DATA?

LO 2.1	Define frequency and explain why it is valuable to summarize data sets.

Suppose you scored 90% on your first statistics exam. How could you determine how well you did in comparison to the rest of the class? First you would need to consider the scores of the other students. If there were 20 students in the class, the list of scores could look like Figure 2.1a. This list is not particularly helpful because it is not readily apparent how a score of 90% compares to the other grades. A more meaningful arrangement is to place the data in a summary table that shows the frequency of exam scores. A **frequency** is the number of times, or how often, a category, score, or range of scores occurs. In this example, it is the number of scores for each grade range. When you arrange the scores in this way, as shown in Figure 2.1b, it is now readily apparent that an exam score of 90% is very good. Only three students earned a score of 90% or higher, and most of the class earned lower scores.

FIGURE 2.1 ■ Summarizing the Frequency of Scores

Exam Scores	
90%	80%
59%	72%
64%	84%
77%	87%
88%	60%
78%	66%
94%	78%
96%	73%
65%	81%
79%	55%

(a)

Exam Scores (%)	Frequency
90–99	3
80–89	5
70–79	6
60–69	4
50–59	2

(b)

This figure shows (a) a list of 20 exam scores and (b) a summary table of the frequency of scores from that list.

This simple example illustrates why it is valuable to summarize data (recall from Chapter 1 that summarization is part of descriptive statistics), in that it can make the presentation and interpretation of a distribution of data clearer. For larger data sets, summarizing data can be even more valuable to clarify patterns or outcomes in the data that may otherwise go unnoticed or "hidden" if only the raw data set were presented. With the continued development of software and technologies today, we can construct just about any table or graph we need for data sets of any size. In this chapter, we start with a discussion of frequency distribution tables and conclude with graphs of distributions for frequency data. In all, this chapter will help you differentiate between the many types of tabular and graphical displays and appropriately construct and accurately interpret many of these tables and graphs used to summarize data in behavioral research.

LEARNING CHECK 2.1

1. Which of the following explains why is it valuable to summarize data?
   a. It can increase the complexity of very large data sets.
   b. It can make the presentation and interpretation of a distribution of data clearer.
2. A study shows that out of 50 employees, 70% reached their sales goals following a training seminar. Thus, the frequency of those reaching their sales goal is
   a. 35 employees
   b. 40 employees
   c. 50 employees
   d. 70 employees
3. Summarizing the frequency of scores in a distribution can be more valuable for ____ data sets. [Fill in the blank]
   a. smaller
   b. relative
   c. larger

Answers: 1. b; 2. a; 3. c.

2.2 SIMPLE FREQUENCY DISTRIBUTIONS FOR GROUPED DATA

LO 2.2 Construct a simple frequency distribution for grouped data.

A **frequency distribution** is a summary of data in terms of how often a category, score, or range of scores occurs. Frequency distributions are useful when researchers measure counts of behavior. Indeed, to find the frequency of values, we must quite literally count the number of times that scores occur, which can help the reader "see" themes or patterns in the data. Example 2.1 applies the steps for distributing the frequency of scores in a data set.

EXAMPLE 2.1

To demonstrate how to construct a simple frequency distribution, let us look at data for the time that children 8 years or younger spend using electronic devices, such as a smartphone or tablet (Pew Research Center, 2018). The values listed in Table 2.1 represent the average time (in minutes, rounded to the nearest whole number) that 50 randomly selected healthy children spent using their electronic devices per day.

Table 2.1 is not really informative as presented. It is only a listing of 50 numbers. To make sense of this list, researchers will consider what makes these data interesting—that is, they will determine how to make the results of this study more meaningful to someone wanting to learn something about children's use of electronic devices.

One way to make these data more meaningful is to summarize how often scores occur in this list using a simple frequency distribution. A **simple frequency distribution** is a summary display for (1) the frequency of each individual score or category (ungrouped data) in a distribution or (2) the frequency of scores falling within defined groups or intervals (grouped data) in a distribution. **Grouped data** are a set of scores distributed into intervals, where the frequency of each score can fall into any given interval. Consider that for many data sets we often collect hundreds or thousands of scores. With such large data sets with many different values recorded, it is generally clearer to summarize the frequency of scores in groups or intervals—each **interval** is a discrete range of values within which the frequency of a subset of scores is contained.

© iStock.com/JackF

TABLE 2.1 ■ The Average Time (in Minutes) That 50 Healthy Children Under 8 Years of Age Spent Using Electronic Devices Per Day

30	70	7	47	13
60	0	91	33	44
40	9	67	55	65
12	140	77	49	77
110	98	21	22	44
33	44	18	10	33
30	20	110	109	54
17	40	102	33	17
55	16	90	12	175
44	33	33	7	82

Table 2.1 lists 50 scores, which is large enough that we will need to summarize this list of scores into groups or intervals. To construct a simple frequency distribution for grouped data, follow three steps:

Step 1: Find the real range.

Step 2: Find the interval width.

Step 3: Construct the frequency distribution.

We will follow these three steps to construct a frequency distribution for the data given in Table 2.1.

Step 1: Find the real range. The **real range** is one more than the difference between the largest and smallest number in a data set. In Table 2.1, the smallest value is 0, and the largest value is 175; therefore, 175 – 0 = 175. The *real range* is 175 + 1 = 176.

Step 2: Find the interval width. The **interval width** is the range of values contained in each interval of a frequency distribution with grouped data. To find this, we divide the real range by the number of intervals chosen. The recommended number of intervals is between 5 and 20. Anything less provides too little summary; anything more is often too confusing. Regardless, *you* choose the number of intervals. The computation for the interval width can be stated as follows:

$$\text{Interval Width} = \frac{\text{Real Range}}{\text{Number of Intervals}}$$

If we decide to split the data in Table 2.1 into 10 intervals, then the computation is $\frac{176}{10}$; hence, the interval width is 17.6. Rounding is necessary when the value for the interval width is not the same degree of accuracy as the original list. For example, the data listed in Table 2.1 are rounded to the ones place (i.e., the nearest whole number). Thus, the interval width should also be a whole number. If it is not, then the interval width should be *rounded up* to the nearest whole number. For this example, we round 17.6 up to an interval width of 18 (the nearest whole number).

Step 3: Construct the frequency distribution. To construct the frequency distribution, we distribute the same number of intervals that we chose in Step 2. In this case, we chose 10 intervals. Table 2.2 shows that each interval has a width of 18. Notice that the first interval contains 18 times in seconds (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17), so it does have a width of 18. In a frequency distribution, the **interval boundaries** mark the cutoffs for a given interval—these are the upper and lower limits for each interval in a frequency distribution with grouped data. The **lower boundary** is the smallest value in each interval, and the **upper boundary** is the largest value in each interval in the frequency distribution.

TABLE 2.2 ■ Simple Frequency Distribution

Intervals	$f(x)$
162–179	1
144–161	0
126–143	1
108–125	3
90–107	4
72–89	3
54–71	7
36–53	8
18–35	12
0–17	11
Total	50

The number (or frequency) of values contained in each interval

The lower boundary is 0, and the upper boundary is 17 for this interval. The interval width is 18 for all intervals.

The total number of values measured in the data set. This equals the sum of the frequencies listed in the $f(x)$ column.

A simple frequency distribution for the average time in minutes that 50 healthy children spent using electronic devices per day. In this table, $f(x)$ denotes the column heading for the frequency (f) of scores (x) in each interval.

Table 2.2 shows that each lower boundary begins one degree of accuracy greater than the previous upper boundary. This means that we add one whole number because this is the degree of accuracy of the data. (If the data were to the tenths place, then we would add .1; if the data were to the hundredths place, then we would add .01; and so on.) We again make the interval width 18 for the second interval and repeat this process until all

10 intervals are constructed. In all, there are four rules for creating a simple frequency distribution:

1. Each interval is defined (it has a lower and upper boundary). Intervals such as "or more" or "less than" should not be expressed.
2. Each interval is equidistant (the interval width is the same for each interval).
3. No interval overlaps (the same score cannot occur in more than one interval).
4. All values are rounded to the same degree of accuracy measured in the original data (to the ones place for the data listed in Table 2.1).

The total counts in a frequency distribution should sum to the total number of counts made. Because we recorded data for 50 children, the frequency distribution must sum to 50; otherwise, we made a counting error. In all, the simple frequency distribution in Table 2.2 paints a picture of the data, so to speak—it conveys a more descriptive and meaningful way to look at the data.

Also, Table 2.2 shows that only two values fall at or above 126. It may be tempting here to combine the top three intervals into one **open interval**— an interval with no defined upper or lower boundary. In other words, we could list the interval as "126 and above" because only two values were counted. It is not uncommon for open intervals to be published when outliers exist in a set of data, but be aware that this is not very informative. An **outlier** is an extreme score that falls substantially above or below most other scores in a data set. In the example summarized in Table 2.2, the value of 175 minutes is an outlier because this value falls substantially above most of the other values recorded. An open interval would make it less obvious that this outlier exists because the upper boundary of the interval would not be given; instead, the upper boundary would be left open.

LEARNING CHECK 2.2

1. Which of the following is a summary display for a distribution of data organized or summarized in terms of how often (or frequently) scores occur?
   a. open interval
   b. boundaries
   c. frequency distribution
2. A researcher observes that a single parent works 42.25 hours per week. The degree of accuracy of 42.25 is
   a. the ones place
   b. the tenths place
   c. the hundredths place
3. A set of scores ranges between 4 and 23. What is the interval width for a simple frequency distribution if you choose to create 5 intervals?
   a. 23/5 = 4.6
   b. 20/5 = 4.0
   c. 19/5 = 3.8
4. An open interval can make it less obvious that an outlier exists in a data set.
   a. True
   b. False

Answers: 1. c; 2. c; 3. b; 4. a

2.3 OTHER WAYS OF SUMMARIZING GROUPED DATA IN FREQUENCY DISTRIBUTIONS

LO 2.3	Construct and explain when it is appropriate to use a cumulative frequency, relative frequency, relative percentage, cumulative relative frequency, and cumulative percentage distribution.

Although the frequency of scores in a distribution can be quite informative, we can also describe data in other ways, such as aggregated across the intervals (cumulative frequencies) or as proportions or percentages. In Example 2.2, we apply the steps for distributing the frequency of scores in a data set using a new example and also identify these other ways of summarizing data in a frequency distribution.

EXAMPLE 2.2

A topic of interest in industrial organizational psychology and fields of organizational behavior is employee safety (Cunningham & Jacobson, 2018; Michaels & Barab, 2020). To illustrate ways of summarizing data, suppose a researcher collects data on employee safety by recording the number of complaints about safety filed by employees at 45 small local businesses over the previous 3 years. The results are listed in Table 2.3. In this example, let us first construct a simple frequency distribution for these data.

TABLE 2.3 ■ The Number of Safety Complaints That Employees at 45 Local Small Businesses Filed Over the Previous 3 Years

45	98	83	50	86
66	66	88	95	73
88	55	76	115	66
92	110	79	105	101
101	85	90	92	81
55	95	91	92	
78	66	73	58	
86	92	51	63	
91	77	88	86	
94	80	102	107	

Step 1: Find the real range. The smallest value in Table 2.3 is 45, and the largest value is 115; therefore, 115 – 45 = 70. The real range is 70 + 1 = 71.

Step 2: Find the interval width. We can split the data into eight intervals (again, you choose the number of intervals). The interval width is the real range divided by the number of intervals: $\frac{71}{8} = 8.88$. The original data are listed as whole numbers, so we round up to the nearest whole number. The nearest whole number is the degree of accuracy of the data. The interval width is 9.

Step 3: Construct the frequency distribution. The frequency distribution is shown in Table 2.4. The first interval starts with the smallest value (45) and contains nine values. To construct the next interval, add one degree of accuracy, or one whole number in this example, and repeat the steps to construct the remaining intervals.

One important rule of thumb is to always summarize data in terms of how you want to describe it. For example, if you want to describe the frequency of safety complaints in discrete intervals, then a simple frequency distribution is a great way to summarize the data. But often, researchers want to describe frequencies "at or above" a certain value, or the percentage of people scoring "at least" a certain score. In these cases, it can be more effective to summarize frequency data cumulatively or as percentages. Many of these additional summaries for frequency data are described here.

TABLE 2.4 ■ Simple Frequency Distribution

Intervals	*f(x)*
108–116	2
99–107	5
90–98	11
81–89	9
72–80	7
63–71	5
54–62	3
45–53	3

A simple frequency distribution with 8 intervals and an interval width of 9. The data are the number of complaints about safety that employees of 45 local small businesses filed over the previous 3 years.

Cumulative Frequency

When researchers want to describe frequencies above or below a certain value, they often use a **cumulative frequency distribution**, which is a summary display that distributes the sum of frequencies across a series of intervals. You can add from the top or from the bottom; it really depends on how you want to discuss the data. To illustrate, we will use the data summarized in Table 2.4 for Example 2.2, which shows the frequencies of the number of complaints about safety for 45 small local businesses.

Suppose the researcher uses the following criteria to categorize these businesses as safe, at risk, or dangerous: safe = fewer than 72 complaints filed, at risk = between 72 and 89 complaints filed, and dangerous = at least 90 complaints filed. Let us explore how a cumulative frequency distribution can be a clearer way to describe these safety categories.

We can sum the frequencies beginning with the bottom frequency and adding up the table, as shown in the last column of Table 2.5. In the table, we began with the frequency in the bottom interval (3) and added the frequency above it to get 6 (3 + 3), added again to get the next frequency 11 (3 + 3 + 5), and repeated these steps until all frequencies were summed. The top frequency is equal to the total number of measures recorded (the total was 45 businesses in this example). This type of summary from the "bottom up" is most meaningful to discuss data in terms of "less than" or "at or below" a certain value or "at most." For example, "safe" businesses are those that report *fewer than* 72 complaints. In the cumulative frequency column, we find that 11 businesses are categorized as safe.

We can also sum the frequencies beginning with the top frequency and adding down the table, although this is less common. To do this, we follow the same steps shown in Table 2.5 but instead begin at the top of the table in the first column. Hence, in the table we begin with the frequency in the top interval (2) and add the frequency below it (5) to get 7 (2 + 5), add again (11) to get the next frequency 18 (2 + 5 + 11), and repeat these steps until all frequencies are summed. The bottom frequency will equal the total number of measures recorded (the total is 45 businesses in this example). This type of "top down" summary is more meaningful to discuss data in terms of "greater than" or "at or above" a certain value or "at least."

TABLE 2.5 ■ A Cumulative Frequency Distribution Table With Calculations Shown in the Center Column

Intervals	Frequency, $f(x)$	Calculation →	Cumulative Frequency (Bottom Up)
108–116	2	3 + 3 + 5 + 7 + 9 + 11 + 5 + 2	45
99–107	5	3 + 3 + 5 + 7 + 9 + 11 + 5	43
90–98	11	3 + 3 + 5 + 7 + 9 + 11	38
81–89	9	3 + 3 + 5 + 7 + 9	27
72–80	7	3 + 3 + 5 + 7	18
63–71	5	3 + 3 + 5	11
54–62	3	3 + 3	6
45–53	3	3	3
	$N = 45$		

Relative Frequency

When researchers summarize larger data sets (with thousands or even millions of counts), they often distribute the relative frequency of scores rather than counts. A **relative frequency distribution** is a summary display that distributes the proportion of scores in each interval of a frequency distribution. It is computed as the frequency in each interval divided by the total number of frequencies recorded. Note that a **proportion** is a part or portion of all measured data. The value of a proportion varies between 0 and 1.0, and the sum of all proportions for a distribution of scores is 1.0. It is often easier to list the relative frequency of scores because a list with very large frequencies in each interval can be more confusing to read. The calculation for a relative frequency is as follows:

$$\text{Relative Frequency} = \frac{\text{Observed Frequency}}{\text{Total Frequency Count}}.$$

Using the same data listed in Table 2.5, we can calculate relative frequency by dividing the frequency in each interval by the total frequency count. The relative frequency in each interval in Table 2.6 (from the top down) is

$$\frac{2}{45} = .04, \frac{5}{45} = .11, \frac{11}{45} = .24, \frac{9}{45} = .20, \frac{7}{45} = .16, \frac{5}{45} = .11, \frac{3}{45} = .07, \text{and } \frac{3}{45} = .07.$$

The sum of relative frequencies across all intervals is 1.00, or, in this example, $\frac{45}{45} = 1.00$.

Relative Percentage

A common way to summarize a relative frequency is to convert it to a relative percentage because most readers find it easier to understand percentages than decimals, perhaps because percentages are the basis for awarding grades from grade school through college.

TABLE 2.6 ■ The Relative Frequency of Scores in Each Interval (Third Column)

Intervals	Frequency, *f(x)*	Relative Frequency	Relative Percentage
108–116	2	.04	4
99–107	5	.11	11
90–98	11	.24	24
81–89	9	.20	20
72–80	7	.16	16
63–71	5	.11	11
54–62	3	.07	7
45–53	3	.07	7
	$N = 45$	Total Rel. Freq. = 1.00	Total Rel. Percentage = 100

The calculation for each relative frequency is given in the text. The relative percentage of scores in each interval is given in the last column.

A **relative percentage distribution** is a summary display that distributes the percentage of scores occurring in each interval relative to all scores distributed. To compute a relative percentage, multiply the relative frequency by 100, which moves the decimal point two places to the right:

$$\text{Relative Percentage} = \frac{\text{Observed Frequency}}{\text{Total Frequency Count}} \times 100$$

Percentages range from 0% to 100% and can never be negative. The relative percentage in each interval is given in the last column in Table 2.6. A relative percentage provides the same information as a relative frequency; it is just that many people find it easier to read percentages than decimals. The choice of which to use is up to the person compiling the data; there is no right or wrong approach because both provide the same information.

Cumulative Relative Frequency and Cumulative Percentage

It is also useful to summarize relative frequencies and percentages cumulatively, for the same reasons described for cumulative frequencies. A **cumulative relative frequency distribution** is a summary display that distributes the sum of relative frequencies across a series of intervals. To distribute the cumulative relative frequency, add each relative frequency beginning at the top or bottom of the table. Table 2.7 lists the bottom-up cumulative relative frequency distribution for the business safety data, with calculations given in the table. To add from the bottom up in the table, we summed the relative frequency in each interval as we moved up the table. For the top-down summary, we would follow these same steps, except we would begin at the top of the table and sum down. The total cumulative relative frequency is equal to 1.00 (give or take rounding errors).

We can also summarize data as percentages. A **cumulative percentage distribution** is a summary display that distributes the sum of relative percentages across a series of intervals.

TABLE 2.7 ■ A Cumulative Relative Frequency Distribution Table With Calculations Shown in the Fourth Column

Intervals	Frequency, *f(x)*	Relative Frequency	Calculation →	Cumulative Relative Frequency (Bottom Up)
108–116	2	.04	.07 + .07 + .11 + .16 + .20 + .24 + .11 + .04	1.00
99–107	5	.11	.07 + .07 + .11 + .16 + .20 + .24 + .11	.96
90–98	11	.24	.07 + .07 + .11 + .16 + .20 + .24	.85
81–89	9	.20	.07 + .07 + .11 + .16 + .20	.61
72–80	7	.16	.07 + .07 + .11 + .16	.41
63–71	5	.11	.07 + .07 + .11	.25
54–62	3	.07	.07 + .07	.14
45–53	3	.07	.07	.07
	N = 45			

To distribute cumulative percentages, we sum the relative percentage in each interval, following the same procedures for adding as we did for the cumulative relative frequencies. In a cumulative percentage distribution, shown in the last column of Table 2.8, intervals are typically summed from the smallest to the largest score in a distribution (bottom up). The total cumulative percentage is equal to 100% (give or take rounding errors). In the next section we will discuss the usefulness of a cumulative percentage distribution for identifying percentile points and ranks.

TABLE 2.8 ■ A Cumulative Percentage of Scores Adding From the Bottom Up

Intervals	Frequency, *f(x)*	Relative Percentage	Cumulative Percentage (Bottom Up)
108–116	2	4	100
99–107	5	11	96
90–98	11	24	85
81–89	9	20	61
72–80	7	16	41
63–71	5	11	25
54–62	3	7	14
45–53	3	7	7
	N = 45	Total Rel. Percentage = 100	

LEARNING CHECK 2.3

1. In which direction do we summarize data in a cumulative frequency distribution when we want to describe data as "at or below" a given score?
   a. Top down
   b. Bottom up
2. In which direction do we summarize data in a cumulative frequency distribution when we want to describe data as "at or above" a given score?
   a. Top down
   b. Bottom up
3. Why do researchers often choose to construct a relative frequency distribution to summarize data?
   a. To summarize the total frequencies in a distribution.
   b. To summarize relatively small data sets.
   c. To summarize large data sets.
4. The sum of relative frequencies across all intervals is equal to
   a. the total frequency count
   b. 100%
   c. 1.00
5. A study evaluating consumer behavior classifies products being studied as either tangible goods or services. Out of 60 total products, 15 were classified as tangible goods. The relative frequency for services is equal to
   a. .15
   b. .25
   c. .60
   d. .75

Answers: 1. b; 2. a; 3. c; 4. c; 5. d.

2.4 IDENTIFYING PERCENTILE POINTS AND PERCENTILE RANKS

LO 2.4	Identify percentile points and percentile ranks in a cumulative percentage distribution.

In some cases, it may be useful to identify the position or rank of an individual within a frequency distribution. You find cases like this in standardized testing, for example, in which you are given scores that indicate your rank as a percentage relative to others who took the same exam. Class standing is also based upon ranks, with students in the "top percentage of the class" being among the best or highest-performing students. To identify the position or rank of an individual, we convert a frequency distribution to a cumulative percentage distribution, then apply the steps identified in this section.

A cumulative percentage distribution is used to identify percentiles, which are measures of the relative position of individuals or scores within a larger distribution. A percentile, called a **percentile point**, is the value of a score on a measurement scale below which a specified percentage of scores in a distribution fall. The corresponding percentile of a percentile point is the **percentile rank** of that score—it is the percentage of scores with values that fall below a specified score, or percentile point, in a distribution. Thus, the 75th percentile point, for example, is the value (the percentile point) below which 75% of scores in a distribution fall (the percentile rank). To find the percentile point in a cumulative percentage distribution, we follow four basic steps. In this section, we will apply these steps to identify the percentile point in a frequency distribution at the 75th percentile for the business safety data, which are reproduced in Table 2.9.

TABLE 2.9 ■ The Cumulative Percentage of Scores for the Business Safety Data

Intervals	Frequency, *f(x)*	Cumulative Percentage (Bottom Up)
108–116	2	100
99–107	5	96
90–98	**11**	**85**
81–89	9	61
72–80	7	41
63–71	5	25
54–62	3	14
45–53	3	7
	$N = 45$	

← The 75th percentile falls at this interval

Step 1: Identify the interval within which a specified percentile point falls. In our example, we want to identify the 75th percentile point, which falls in the interval of 90–98. Note that each percentile given in Table 2.9 is the top percentage in each interval.

Step 2: Identify the real range for the interval identified. In our example, the interval that contains the 75th percentile point is the interval of 90–98 (this is the observed range). The real limits for this interval are 0.5 less than the lower limit and 0.5 greater than the upper limit. Hence, the real limits are 89.5 to 98.5. The width of the real range is therefore 9 points, or 1 point greater than the observed range. For the percentages, the range width is 24 percentage points (from 61% to 85%).

Step 3: Find the position of the percentile point within the interval. To identify the position of the percentile point, first find the distance of the 75th percentile from the top of the interval. The 75th percentile is 10 points from the top of the interval. Next, divide 10 by the total range width of the percentages. Hence, 75% is 10 out of 24, or $\frac{10}{24}$ of the total interval.

Interval	Percentages
98.5	85
?	**75**
89.5	61

← We are looking for the percentile point at the 75th percentile.

As a final part for this step, multiply the fraction by the width of the real range, which is 9 points:

$$\frac{10}{24} \times 9 = 3.75 \text{ points.}$$

Hence, the position of the percentile point is 3.75 points from the top of the interval.

Step 4: Identify the percentile point. In this example, the top of the interval is 98.5. We subtract 3.75 from that value to identify the percentile point at the 75th percentile: 98.5 – 3.75 = 94.75. Thus, the percentile point at the 75th percentile is 94.75.

In all, you can follow these basic steps to find the percentile point at any percentile rank in a frequency distribution.

LEARNING CHECK 2.4

1. What type of distribution is used to identify percentiles, which are measures of the relative position of individuals or scores within a larger distribution?
   a. simple frequency distribution
   b. cumulative frequency distribution
   c. cumulative percentage distribution
2. A ______ is the value of a score on a measurement scale below which a specified percentage of scores in a distribution fall. The corresponding percentile of a percentile point is the ______ of that score. [Fill in the blanks]
   a. percentile point; percentile rank
   b. percentile rank; percentile point
3. A student, Jasmine, scores in the 80th percentile on an exam. Which of the following correctly explains what this means?
   a. 80% of students scored higher on the exam than Jasmine.
   b. Jasmine scored higher than 80% of all others who took the same exam.

---

answers: 1. c; 2. a; 3. b.

2.5 SPSS IN FOCUS: FREQUENCY DISTRIBUTIONS FOR QUANTITATIVE DATA

SPSS LO 2.5 Construct a frequency distribution for quantitative data using SPSS.

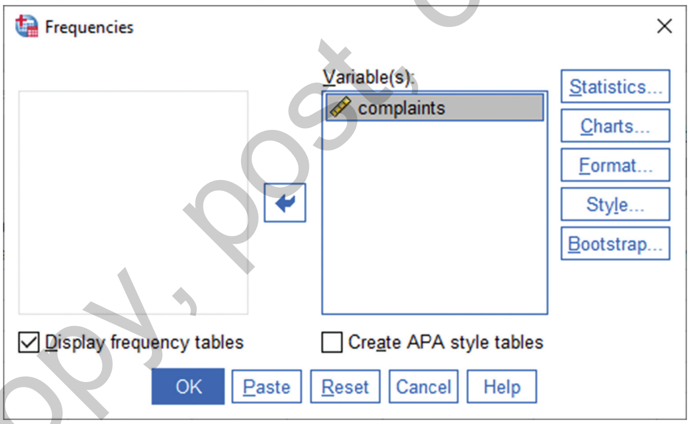
SPSS can be used to construct frequency distributions. Here we will start with data entry (Steps 1 and 2), then construct a frequency distribution for quantitative data (Steps 3–5). In this section, we will construct a frequency distribution table for the business safety data first listed in Table 2.3, which are reproduced here for reference.

1. Click on the Variable View tab and enter *complaints* in the Name column. We will enter whole numbers, so go to the Decimals column and reduce the value to 0. In the Measure column label the variable as Scale.
2. Click on the Data View tab and enter the 45 values from Table 2.3 in the column labeled *complaints*. You can enter the data in any order you wish, but make sure all the data are entered correctly.
3. Go to the menu bar and click Analyze, then Descriptive Statistics and Frequencies, to display the dialog box shown in Figure 2.2.

Data Reproduced From Table 2.3

45	98	83	50	86
66	66	88	95	73
88	55	76	115	66
92	110	79	105	101
101	85	90	92	81
55	95	91	92	
78	66	73	58	
86	92	51	63	
91	77	88	86	
94	80	102	107	

FIGURE 2.2 ■ SPSS Dialog Box



4. In the dialog box, select the *complaints* variable. When you click the arrow in the center, it will move *complaints* into the Variable(s) box to the right. Make sure the option to display frequency tables is selected.

5. Select OK, or select Paste and click the Run command.

Table 2.10 shows the SPSS output table display. SPSS did not distribute these data into intervals as we did. Instead, every value in the original data set is listed (in numerical order from least to most) with frequencies, relative percentages (middle two columns), and cumulative percentages given for each value. Note that SPSS automatically groups data into intervals only with very large data sets.

TABLE 2.10 ■ SPSS Output Display

complaints

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	45	1	2.2	2.2	2.2
	50	1	2.2	2.2	4.4
	51	1	2.2	2.2	6.7
	55	2	4.4	4.4	11.1
	58	1	2.2	2.2	13.3
	63	1	2.2	2.2	15.6
	66	4	8.9	8.9	24.4
	73	2	4.4	4.4	28.9
	76	1	2.2	2.2	31.1
	77	1	2.2	2.2	33.3
	78	1	2.2	2.2	35.6
	79	1	2.2	2.2	37.8
	80	1	2.2	2.2	40.0
	81	1	2.2	2.2	42.2
	83	1	2.2	2.2	44.4
	85	1	2.2	2.2	46.7
	86	3	6.7	6.7	53.3
	88	3	6.7	6.7	60.0
	90	1	2.2	2.2	62.2
	91	2	4.4	4.4	66.7
	92	4	8.9	8.9	75.6
	94	1	2.2	2.2	77.8
	95	2	4.4	4.4	82.2
	98	1	2.2	2.2	84.4
	101	2	4.4	4.4	88.9
	102	1	2.2	2.2	91.1
	105	1	2.2	2.2	93.3
	107	1	2.2	2.2	95.6
	110	1	2.2	2.2	97.8
	115	1	2.2	2.2	100.0
	Total	45	100.0	100.0	

LEARNING CHECK 2.5

1. Using SPSS, which options in the menu bar do you select to construct a frequency distribution for quantitative data?
   a. Analyze, then Descriptive Statistics and Explore
   b. Analyze, then Descriptive Statistics and Descriptives
   c. Analyze, then Descriptive Statistics and Frequencies

2. For a given variable, all data for that variable are entered into a single column in the Data View tab.
- **a.** True, all data for that variable are entered into a single column in the Data View tab.
- **b.** False, data should always be entered in the Variable View tab.

A portion of an SPSS output display is given in the image below for data observing the duration of time (in seconds) that participants attend to a task. [The display gives all the information needed to enter these data yourself. Please feel free to enter these data in SPSS and follow the directions given in Section 2.5 to reproduce the SPSS output shown.] Based on the data summarized in the table, answer Questions 3 and 4.

duration

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.88	1	10.0	10.0	10.0
	1.05	1	10.0	10.0	20.0
	1.32	1	10.0	10.0	30.0
	2.03	1	10.0	10.0	40.0
	2.20	1	10.0	10.0	50.0
	2.56	1	10.0	10.0	60.0
	3.00	1	10.0	10.0	70.0
	3.87	1	10.0	10.0	80.0
	5.03	1	10.0	10.0	90.0
	5.32	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

3. What is the cumulative frequency for a duration at or above 3.00?
- **a.** 4
- **b.** 7
- **c.** 10

4. What is the cumulative frequency for a duration at or below 3.00?
- **a.** 4
- **b.** 7
- **c.** 10

Answers: 1. c; 2. a; 3. a; 4. b.

2.6 FREQUENCY DISTRIBUTIONS FOR UNGROUPED DATA

LO 2.6 Construct a frequency distribution for ungrouped or categorical data.

Although grouping data into intervals is a practical way to summarize large, quantitative data sets, it is not always necessary to group data. In other words, it can be appropriate to leave the data ungrouped. **Ungrouped data** are a set of scores or categories distributed individually, in which the frequency for each individual score or category is counted. For example, it is appropriate to leave data ungrouped when the dependent variable is qualitative or categorical, or when the number of different scores is small (i.e., the scores cannot easily fit into at least 5 intervals). For cases in which data are ungrouped, each measured score or category is listed in a frequency table; intervals are not constructed. In Example 2.3, we look at a new data to construct a frequency distribution for ungrouped data.

EXAMPLE 2.3

To construct a frequency distribution for ungrouped data, let us look at research on the benefits of napping among children, which is a topic of interest in the behavioral sciences (Horváth & Plunkett, 2018; Staton et al., 2020). To explore how many naps children actually take, you ask at random a sample of 40 primary caregivers of a child younger than age 3 how many naps their child takes per day, on average. Table 2.11 lists the hypothetical results.

TABLE 2.11 ■ A List of the Number of Naps That Children (Younger Than 3 Years) Take Each Day

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

For the napping data, the caregivers gave one of five responses: 0, 1, 2, 3, or 4 naps per day. Grouping data with only five different responses makes little sense, especially because the recommended minimum number of intervals is five. Instead, the data should remain ungrouped. The frequency of ungrouped data is simply listed in a frequency table. We skip all the steps for creating the intervals and go straight to counting the frequency of each value. Each score represents its own count in a frequency distribution table, as shown in Table 2.12. Notice the important distinction: Grouped data have intervals, and ungrouped data do not.

Ungrouped data can be summarized using relative frequency, cumulative relative frequency, relative percentage, and cumulative relative percentage, just as grouped data can. It again depends on how you want to describe the data. Summarizing ungrouped data is especially practical for data sets with only a few different scores and for qualitative or categorical variables. Data obtained for opinion and marketing polls, health categories (lean, healthy, overweight, and obese), or college year (freshman, sophomore, junior, and senior) are often summarized as ungrouped data.

TABLE 2.12 ■ A Simple Frequency Distribution for Ungrouped Data

Number of Naps	Frequency
0	8
1	8
2	15
3	8
4	1
	$N = 40$

The data are the number of naps that children, younger than age 3, take per day.

DATA IN RESEARCH

SUMMARIZING DEMOGRAPHIC INFORMATION

© iStock.com/gorodenkoff

Lau (2017) conducted a study to examine how social media use and social media multitasking impact academic performance among an undergraduate college sample of 342 students. To describe general characteristics of the students they sampled, they included data, shown in Table 2.13, summarizing the subjects being studied and class year for students (ungrouped, categorical data). The table includes the frequency and relative percentage for each categorical variable.

Most academic journals require that researchers report relevant demographic information for human participants. In this study, the researcher reported demographic data that were relevant to the study. This study showed that using social media for academic purposes did not predict academic performance (i.e., cumulative grade point average), whereas using social media for nonacademic purposes (e.g., video gaming) and multitasking between many types of social media predicted worse academic performance.

TABLE 2.13 ■ Demographic Information of Subject Studied and Class Year of Undergraduate Student Participants

Variable		Frequency	Relative Percent
Subject	Arts	51	14.9%
	Business	73	21.3%
	Education	19	5.6%
	Engineering	29	8.5%
	Law	8	2.3%
	Medicine	53	15.5%
	Science	47	13.7%
	Social Science	62	18.1%
Class Year	First year	123	36.0%
	Second year	85	24.9%
	Third year	75	21.9%
	Fourth year	57	16.7%
	Fifth year	2	0.6%

Source: Adapted from Lau, W. W. F. (2017). Effects of social media usage and social media multitasking on the academic performance of university students. *Computers in Human Behavior, 68*, 286–291.

LEARNING CHECK 2.6

1. When will a researcher summarize ungrouped data in a frequency distribution?
   - **a.** When the dependent variable is qualitative or categorical.
   - **b.** When the number of different scores is small (i.e., the scores cannot be fit easily into at least 5 intervals).
   - **c.** Both a and b.
2. Ungrouped data can be summarized using what type of frequency distribution?
   - **a.** Relative frequency
   - **b.** Cumulative relative frequency
   - **c.** Relative percentage
   - **d.** Cumulative relative percentage
   - **e.** All options above.
3. A professor gives a quiz that is out of 5 possible points (1 point for each correct answer with no partial credit possible). What type of frequency distribution is most appropriate?
   - **a.** A frequency distribution for grouped data.
   - **b.** A frequency distribution for ungrouped data.

Answers: 1. c; 2. e; 3. b.

2.7 SPSS IN FOCUS: FREQUENCY DISTRIBUTIONS FOR CATEGORICAL DATA

SPSS LO 2.7 Construct a frequency distribution for categorical data using SPSS.

SPSS can be used to summarize categorical data that are ungrouped. Here we will start with data entry (Steps 1–4), then construct a frequency distribution for categorical data (Steps 5–7). As an example for completing these steps, let us create a frequency distribution for the following: A group of health practitioners identify children in a local public school system as being lean, healthy, overweight, or obese—this is a common classification (Centers for Disease Control and Prevention, 2019). In their assessment, they classified 15 children as lean, 30 as healthy, 35 as overweight, and 20 as obese.

1. Click on the Variable View tab and enter *categories* in the Name column. In the second row, enter *frequencies* in the Name column. We will enter whole numbers, so go to the Decimals column and reduce the value to 0 for both rows.
2. We must code the data for the categorical variable. In the first row, click on the Values cell and click on the small gray box with three dots. In the dialog box, enter *1* in the value cell and *lean* in the label cell, and then click Add. Repeat these steps by entering *2* for *healthy, 3* for *overweight*, and *4* for *obese*, and then select OK. Now each level for the categorical variable is coded.
3. Click on the Data View tab and enter *1, 2, 3*, and *4* in the *categories* column. In the *frequencies* column, enter *15, 30, 35*, and *20* next to the corresponding numeric code.
4. Go to Data then Weight cases . . . to open up a dialog box. Select Weight cases by and move *frequencies* into the Frequency Variable box. Now each frequency is matched to each level of the variable. Select OK.
5. Go to the menu bar and click Analyze, then Descriptive Statistics and Frequencies, to display a dialog box.
6. In the dialog box, select the *categories* variable and click the arrow in the center to move *categories* into the box on the right labeled Variable(s). Make sure the option to display frequency tables is selected.
7. Select OK, or select Paste and click the Run command.

Notice that SPSS does not list the values as 1, 2, 3, and 4 in the output table shown in Table 2.14, although you entered these values in the *categories* column in Data View. Instead, SPSS lists the data as you labeled them in Step 2 in Variable View. This format makes it much easier to read the output file. Also, every category in the original data set is listed with frequencies, relative percentages, and cumulative percentages given.

TABLE 2.14 ■ SPSS Output Display

Statistics

categories

N	Valid	100
	Missing	0

categories

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	lean	15	15.0	15.0	15.0
	healthy	30	30.0	30.0	45.0
	overweight	35	35.0	35.0	80.0
	obese	20	20.0	20.0	100.0
	Total	100	100.0	100.0	

LEARNING CHECK 2.7

1. Using SPSS, which options in the menu bar do you select to construct a frequency distribution for categorical data?
   a. Analyze, then Descriptive Statistics and Explore
   b. Analyze, then Descriptive Statistics and Descriptives
   c. Analyze, then Descriptive Statistics and Frequencies
2. It is not necessary to code categorical variables in SPSS.
   a. True, coding is not appropriate in SPSS.
   b. False, categorical variables are coded in SPSS.

A portion of an SPSS output display is given in the image below for data observing the number of times a penalty kick in soccer was directed to the middle, left, and right side of the goal. [The display gives all the information needed to enter these data yourself. Please feel free to enter these data in SPSS and follow the directions given in Section 2.7 to reproduce the SPSS output shown.] Based on the data summarized in the table, answer Questions 3 and 4.

Direction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Left	101	50.5	50.5	50.5
	Middle	12	6.0	6.0	56.5
	Right	87	43.5	43.5	100.0
	Total	200	100.0	100.0	

3. In which direction was a soccer ball kicked most often?
   a. left
   b. middle
   c. right
4. How many times was a soccer ball kicked to the right side of the goal?
   a. 12
   b. 87
   c. 101

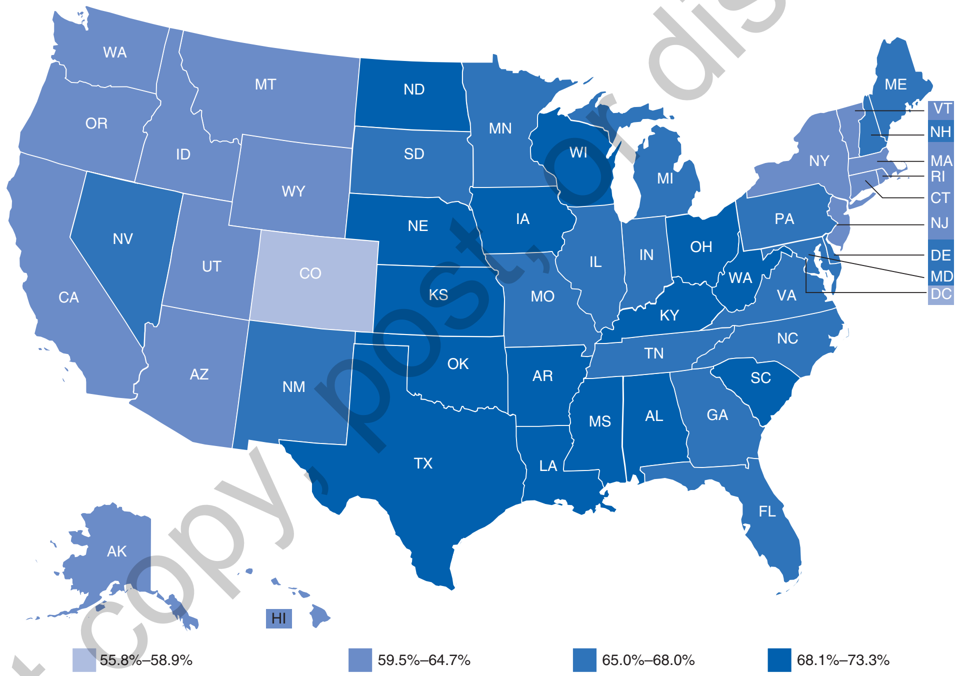
Answers: 1. c; 2. b; 3. a; 4. b.

2.8 PICTORIAL FREQUENCY DISTRIBUTIONS

LO 2.8 Identify how presenting images can meaningfully illustrate data.

Pictures often help people make sense of frequency data, inasmuch as the reader can relate to the images presented. This type of pictorial presentation is called a **pictogram**, also called a **pictograph**, which is a summary display that uses symbols or illustrations to represent a concept, object, place, or event that typically corresponds to or reflects the data being reported. By relating frequency data to images that correspond to or reflect the data being reported, pictograms can help "bring to life" or illustrate the data in a clearer way, in the same way data presented on a dashboard in a car can help to make sense of information quickly, such as your speed, mileage, and more. In research, economic or health-related data given by state, for example, are often presented as pictures, such as that shown in Figure 2.3. The figure shows a map of the United States with the distribution of Americans who are overweight and obese given as a percentage for each state. This map groups the relative percentage for each state into one of four groups: Each interval is coded by color at the bottom. Although it is not recommended, the intervals are not equal in this example. The reason is probably that the original percentages were published with this distribution, and readers could view the actual relative percentage for each state in the list in addition to viewing the map. To find where your state stands on this or other health-related topics, visit http://kff.org/statedata/ and choose a category or a location. If you visit that website, pay attention to the graphs and tables to explore how they measure up to the standards described in this chapter.

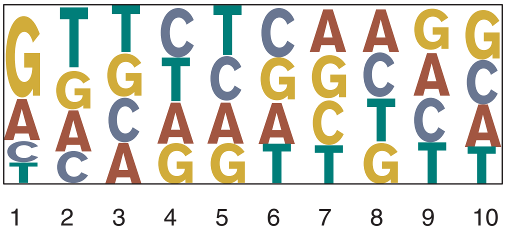
FIGURE 2.3 ■ A Map Showing the Percentage of U.S. Adults Who Are Overweight and Obese, 2018



Source: "Percentage of Adults Who Are Overweight or Obese, 2018," The Henry J. Kaiser Family Foundation, 2018. This information was reprinted with permission from the Henry J. Kaiser Family Foundation. The Kaiser Family Foundation is a nonprivate private operating foundation, based in Menlo Park, California, dedicated to producing and communicating the best possible analysis and information on health issues.

A pictogram is an effective way to summarize frequency data. A pictogram uses symbols or illustrations to represent a concept, object, place, or event. Pictograms are often seen in magazines, newspapers, government documents or reports, and even research journal articles. For instance, Figure 2.4, which was published in the academic journal *BMC Genomics*, illustrates the relative percentage frequency distribution for each nucleotide (adenine [A], thymine [T], cytosine [C], and guanine [G]) in a section of human DNA. In this study, researchers used images of the nucleotide letters instead of tables to distribute their frequency data. In this way, pictograms are effective at summarizing everything from health facts by state to nucleotides in a human DNA sequence.

FIGURE 2.4 ■ A Pictogram of a Specific Section of Human DNA



The size of a letter on the pictogram is proportional to the frequency of the relevant nucleotide. The relative percentile of the nucleotide with the largest frequency is given below each nucleotide position.

Source: Adapted from Dong, H., Zhang, L., Zheng, K. et al. (2012). A Gaijin-like miniature inverted repeat transposable element is mobilized in rice during cell differentiation. *BMC Genomics 13,* 135.

LEARNING CHECK 2.8

1. A ______ is a summary display that uses symbols or illustrations to represent a concept, object, place, or event. [Fill in the blank]
   - **a.** pictogram
   - **b.** symbolism
   - **c.** cumulative distribution
2. A pictogram is an effective way to summarize frequency data.
   - **a.** True, a pictogram is an effective way to summarize frequency data.
   - **b.** False, pictogram is an ineffective way to summarize frequency data.
3. What types of publications are pictograms often reported in?
   - **a.** magazines
   - **b.** newspapers
   - **c.** research journals
   - **d.** all of the above

Answers: 1. a; 2. a; 3. d.

2.9 GRAPHING DISTRIBUTIONS: CONTINUOUS DATA

LO 2.9 Construct and interpret graphs for distributions of continuous data.

Researchers can also display frequency data graphically instead of in a table. Although using a table or graph to summarize frequency data is equally effective for the most part, graphs have the advantage of being more visual and less intimidating than tables in many cases. In this section, we look at several ways to graph distributions of continuous data: histograms, frequency polygons, and ogives.

Histograms

Grouped data are often summarized graphically using a **histogram**, which is a graphical display used to summarize the frequency of continuous data that are distributed in numeric intervals. Histograms are graphs that distribute the intervals along the horizontal scale (x-axis) and list the frequency of scores in each interval on the vertical scale (y-axis). To illustrate, we can construct a histogram for the data given in Figure 2.5, which shows the frequency table and the respective histogram for the time (in months) it took a sample of 200 college graduates to find employment. To construct a histogram, we follow three rules:

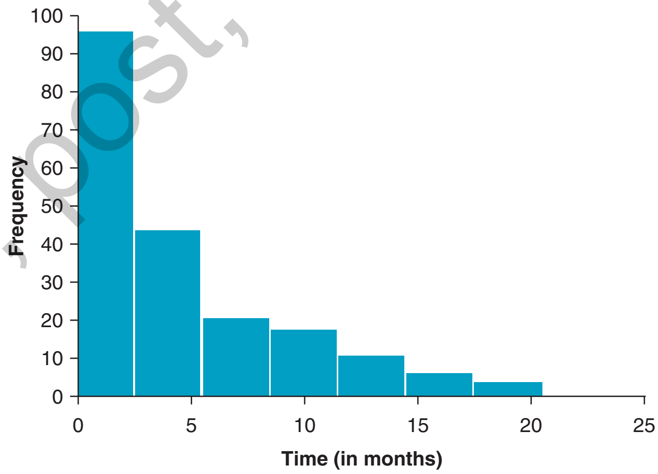
Rule 1: A vertical rectangle represents each interval, and the height of the rectangle equals the frequency recorded for each interval. This rule implies that the y-axis should be labeled as a number or count. The y-axis reflects the frequency of scores for each interval.

Rule 2: The base of each rectangle begins and ends at the upper and lower boundaries of each interval. This rule means that histograms cannot be constructed for open intervals because open intervals do not have an upper or a lower boundary. Also, each rectangle should have the same interval width.

Rule 3: Each rectangle touches adjacent rectangles at the boundaries of each interval. Histograms are used to summarize continuous data, such as the time (in months) it takes to find employment. The adjacent rectangles touch because the data are continuous. In other words, the data were measured along a continuum.

FIGURE 2.5 ■ A Histogram

Intervals (in months)	$f(x)$
18–20	4
15–17	6
12–14	11
9–11	18
6–8	21
3–5	44
0–2	96

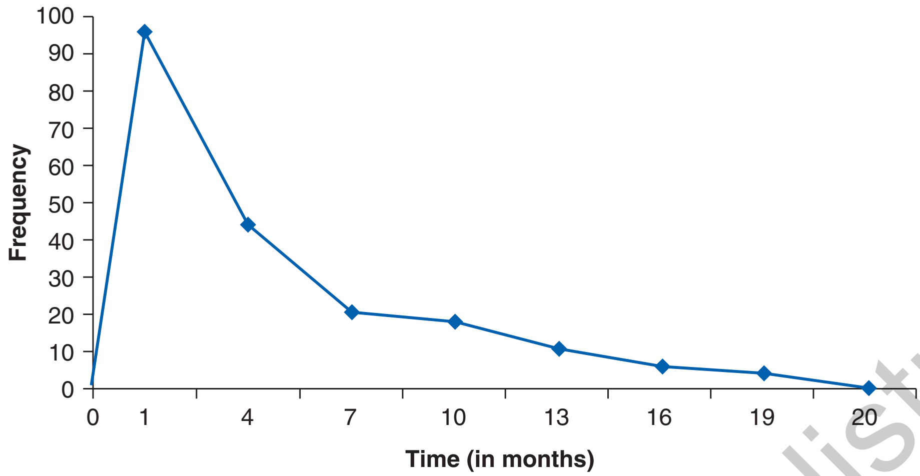


A frequency table (left) and histogram (right) summarizing the frequency distribution for the time (in months) that it took a sample of 200 college graduates to find employment.

Frequency Polygons

Another graph that can be used to summarize grouped data is the **frequency polygon**. A frequency polygon is a dot-and-line graph used to summarize the frequency of continuous data at the midpoint of each interval. In a frequency polygon, a dot is plotted at the midpoint of each interval and a line connects each dot. The midpoint of an interval is distributed along the x-axis and is calculated by adding the upper and lower boundary of an interval and then dividing by 2. Figure 2.6 illustrates a frequency polygon for the same data used to construct the histogram in Figure 2.5. A histogram and a frequency polygon are equally effective at summarizing these data—the choice between the two depends on how you prefer to summarize the data.

FIGURE 2.6 ■ A Frequency Polygon



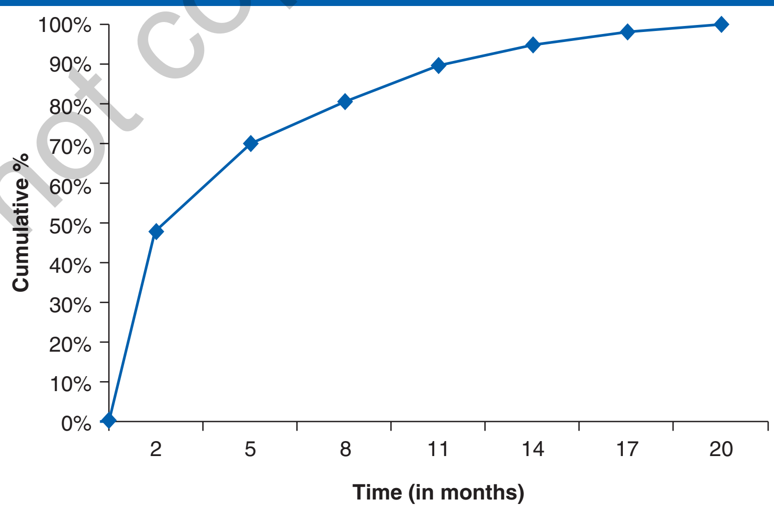
A frequency polygon summarizing the frequency distribution for the time (in months) that it took a sample of 200 college graduates to find employment.

Ogives

A dot-and-line graph can also be useful to summarize cumulative data. One type of graph, called an **ogive** (pronounced *oh-jive*), is a dot-and-line graph used to summarize the cumulative percentages or cumulative frequencies of data at the upper boundary of each interval. In an ogive, a dot is thus plotted at the upper boundary of each interval, and a line connects each dot. As an example, Figure 2.7 shows an ogive for the cumulative percentage distribution, from the bottom up, for the same data used to construct the histogram in Figure 2.5. The *y*-axis of an ogive always ranges from 0% to 100% of the data.

Notice that each dot in an ogive is plotted at the upper boundary of each interval. Each dot represents the cumulative percentage of scores at each interval. Plotting at the upper boundary of each interval is necessary because this point represents or contains all the scores in that interval.

FIGURE 2.7 ■ An Ogive



An ogive summarizing the cumulative percentage distribution for the time (in months) that it took a sample of 200 college graduates to find employment.

LEARNING CHECK 2.9

1. Histograms are used to summarize discrete data, which is why each vertical rectangle touches the other.
   a. True, histograms are used to summarize discrete data.
   b. False, histograms are used to summarize continuous data.
2. A(n) _________ is a dot-and-line graph plotted at the midpoint of each interval, whereas a(n) _________ is a dot-and-line graph plotted at the upper boundary of each interval. [Fill in the blanks]
   a. frequency polygon; ogive
   b. ogive; frequency polygon
3. An ogive graphically summarizes what type of frequency distribution?
   a. A cumulative percentage distribution.
   b. A relative percentage distribution.
   c. Both a and b.

Answers: 1. b; 2. a; 3. c.

2.10 STEM-AND-LEAF DISPLAYS

LO 2.10 Summarize discrete or continuous data in a stem-and-leaf display.

A **stem-and-leaf display**, also called a **stem-and-leaf plot**, is a graphic method of displaying data that can be used for discrete or continuous data, in which each individual score from an original data set is listed. This type of display can be particularly useful with smaller data sets because each individual score is listed in the display. Using a stem-and-leaf display, the data are organized such that the common digits shared by all scores are listed to the left (in the *stem*), with the remaining digits for each score listed to the right (in the *leaf*).

To illustrate this type of display, consider the data listed in Table 2.15 for the number of times per day that 20 patients with obsessive-compulsive disorder washed their hands.

TABLE 2.15 ■ A List of the Number of Times per Day That 20 Patients With Obsessive-Compulsive Disorder Washed Their Hands

12	14	10	47
33	23	16	52
24	32	26	44
42	46	29	19
11	50	30	15

To construct the stem-and-leaf display for these data, we must first recognize the beginning digits of these numbers. For example, 12 begins with the number 1, and 50 begins with a 5, so let us reorganize the data so that every number with the same first digit is placed in the same row, as shown in Table 2.16a for this discrete data set. Here it is apparent that seven numbers begin with the digit 1 (10, 11, 12, 14, 15, 16, and 19), four begin with 2 (23, 24, 26, and 29), three begin with 3 (30, 32, and 33), four begin with 4 (42, 44, 46, and 47), and two begin with 5 (50 and 52). To simplify this further, we can place the first digit in its own column (separated by a vertical line) so that it is never repeated in a single row, as shown in Table 2.16b. This arrangement of data is called a stem-and-leaf display.

TABLE 2.16 ■ Stem-and-Leaf Displays

(a)						
10	11	12	14	15	16	19
23	24	26	29			
30	32	33				
42	44	46	47			
50	52					
(b)						
1	0	1	2	4	5	6
2	3	4	6	9		
3	0	2	3			
4	2	4	6	7		
5	0	2				
(c)						
1.2	0	3	4	8		
2.2	2	5				
(d)						
1	.20	.23	.24	.28		
2	.22	.25				

A stem-and-leaf-display in which every number with the same first digit is placed in the same row (a) and the first digit in each row is placed in its own column (b). Using new data, (c) displays two digits in the stem, and (d) displays two digits in the leaf.

In a stem-and-leaf display, numbers to the right of the vertical line are the **leaf**, and numbers to the left of the vertical line are the **stem**. The leaf lists the last digit or digits for each number in each row; the stem lists the first digit or digits for each number in each row. The names *stem* and *leaf* derive from their appearance, which in many cases looks like half a leaf, with a stem to the left. There are many ways to display data in a stem-and-leaf display. For example, stems can be more than one digit, as shown in Table 2.16c for continuous data, which displays 1.20, 1.23, 1.24, 1.28, 2.22, and 2.25 with two-digit stems. Also, the leaves can have more than one digit, as shown in Table 2.16d, which displays the same numbers, but this time with leaves that have two digits.

In all, a stem-and-leaf display is similar to a histogram in that both display the shape of a distribution for continuous and discrete data. Unlike histograms, though, the stem-and-leaf display lists the actual scores in the distribution, instead of frequencies, allowing a person to read the raw data immediately from the display and usually in ascending order. In this way, the stem-and-leaf display retains the value of each data point. The only information we lose is the order in which the data were originally obtained. This makes stem-and-leaf displays a very useful tool for displaying data.

LEARNING CHECK 2.10

1. In a stem-and-leaf display, the ____ lists the first digit or digits for each number in each row to the left of the vertical line, and the ____ lists the last digit or digits for each number in each row to the right of the vertical line. [Fill in the blanks]
   a. leaf; stem
   b. stem; leaf
2. What is a key distinction between a stem-and-leaf display and a histogram?
   a. A stem-and-leaf display lists actual scores in a distribution; a histogram lists frequencies.
   b. A histogram lists actual scores in a distribution; a stem-and-leaf display lists frequencies.
   c. There is no distinction; a stem-and-leaf display and a histogram summarize data in the exact same way.
3. What are the original scores listed in the following stem-and-leaf display.

1	3	3	7
5	4	5	

   a. 13, 17, 54, and 55.
   b. 13, 13, 17, 54, and 45.
   c. 13, 13, 17, 54, and 55.

Answers: 1. b; 2.a; 3. c.

2.11 GRAPHING DISTRIBUTIONS: DISCRETE AND CATEGORICAL DATA

LO 2.11 Construct and interpret graphs for distributions of discrete or categorical data.

Researchers often measure discrete variables—variables measured in whole units or categories. For example, the number of traffic accidents and the number of college graduates are discrete variables because these variables are measured by counting one traffic accident or college graduate at a time. Researchers also measure categorical variables, which vary by class. Examples include race, gender, and marital status. Discrete and categorical data are graphed differently than continuous data because the data are measured in whole units or classes and not along a continuum. Two types of graphs for discrete and categorical data described here are bar charts and pie charts.

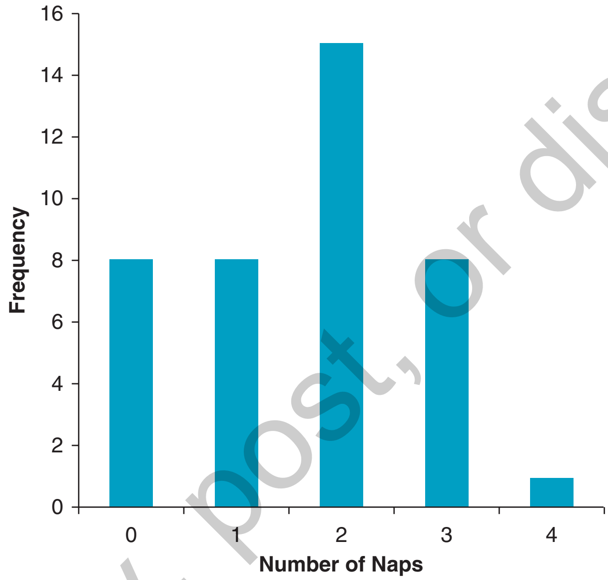
Bar Charts

A **bar chart**, or **bar graph**, is a graphical display used to summarize the frequency of discrete and categorical data that are distributed in whole units or classes, such that (1) a vertical rectangle represents each whole unit or class, (2) the height of the rectangle equals the frequency recorded for each whole unit or class, (3) each rectangle does not touch adjacent rectangles at the boundaries of each interval. Bar charts are much like histograms, except that the bars are separated from one another, whereas the vertical rectangles or bars on histograms touch each other. The separation between bars reflects the separation or "break" between the whole numbers or categories being summarized. For this reason, bar charts are appropriate for summarizing distributions of discrete and categorical data.

To construct a bar chart, list the whole units or categories along the *x*-axis and distribute the frequencies along the *y*-axis. To illustrate, Figure 2.8 gives the frequency distribution and the respective bar chart for the number of naps that mothers give their children daily. (The original data for Figure 2.8 are given in Table 2.11.) The bar chart has two characteristics: (1) Each class or category is represented by a rectangle, and (2) each rectangle is separated along the *x*-axis. Again, the bar chart is nothing more than a histogram with the bars separated, which makes it more appropriate for summarizing discrete and categorical data.

FIGURE 2.8 ■ A Bar Chart

Number of Naps	$f(x)$
0	8
1	8
2	15
3	8
4	1



A frequency table (left) and a bar chart (right) summarizing the average number of naps per day that mothers give their children, who are younger than age 3.

Pie Charts

A **pie chart** is a graphical display in the shape of a circle that depicts the relative percentage of discrete and categorical data as segments of the circle. It is used almost exclusively with discrete and categorical data. Educators often teach children subtraction and other mathematical operations by "slicing up pieces of pie." Similarly, you can think of pie charts as slices or pieces of data. To construct a pie chart, we typically distribute data as relative percentages. Consider Table 2.17, which displays the educational attainment in the United States of a sample of Americans in 2018.

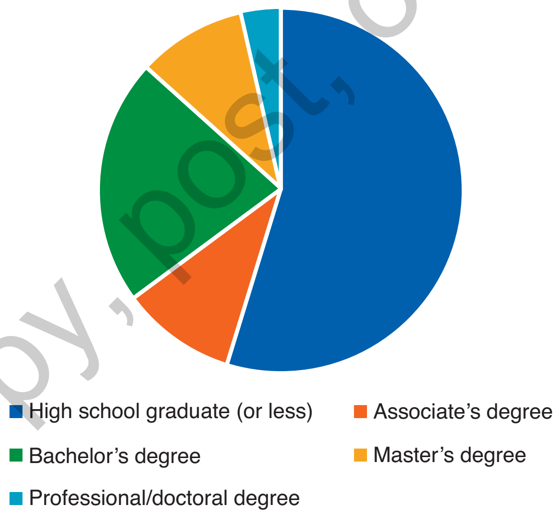
Converting this distribution to a pie chart is simply a matter of finding the correct angles for each slice of pie. There are 360 degrees in a complete circle; therefore, we multiply each percentage by 3.6 (because 100% × 3.6 = 360°) to find the central angles of each **sector**—the portion of the pie chart that represents the relative percentage of a particular class or category. The central angles for the data in Table 2.17 (from the top down and rounded to the nearest tenths place) are 54.8 × 3.6 = 197.28 (rounded up to 197.3), 10.2 × 3.6 = 36.7, 21.9 × 3.6 = 78.9, 9.6 × 3.6 = 34.6, and 3.5 × 3.6 = 12.6. The total of all central angles will equal 360 degrees. Now dust off your protractor or use a computer program (such as Excel or SPSS) to construct the pie chart by slicing the pie into each angle you just calculated. The result is shown in Figure 2.9.

TABLE 2.17 ■ The Frequency and Relative Percentage of Educational Attainment of the Population 25 Years and Older in the United States, 2018

Level of Education	Frequency	Relative Percent
High school graduate (or less)	120,538	54.8%
Associate's degree	22,369	10.2%
Bachelor's degree	48,235	21.9%
Master's degree	21,048	9.6%
Professional/doctoral degree	7,640	3.5%
Total	219,830	100.0%

Source: Table created based on data from the U.S. Census Bureau, Current Population Survey, 2018 Annual Social and Economic Supplement.

FIGURE 2.9 ■ A Pie Chart for the Distribution of Educational Attainment of the Population 25 Years and Older in the United States, 2018



Source: Table created based on data from the U.S. Census Bureau, Current Population Survey, 2018 Annual Social and Economic Supplement.

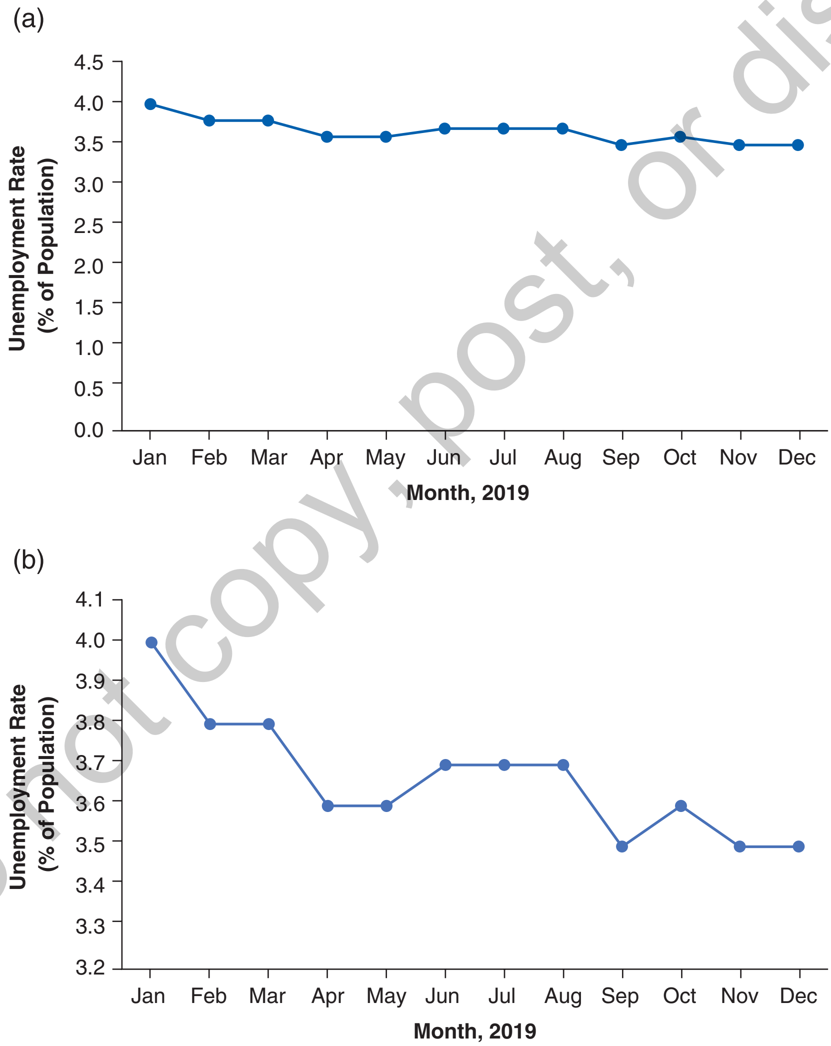
MAKING SENSE

DECEPTION DUE TO THE DISTORTION OF DATA

It was Mark Twain who once said *there are lies, damned lies, and statistics.* His statement identified that statistics can be deceiving—and so can interpreting them. Descriptive statistics are used to inform us. Therefore, being able to identify statistics and correctly interpret what they mean is an important part of the research process. Presenting data can be an ethical concern when the data are distorted in any way, whether by accident or intentionally. The distortion of data can occur for data presented graphically or as summary statistics. Here, we will describe how the presentation of data can be distorted.

When a graph is distorted, it can deceive the reader into thinking differences exist when, in truth, differences are negligible (Frankfort-Nachmias & Leon-Guerrero, 2006; Privitera, 2019). Three common distortions to look for in graphs are (1) displays with an unlabeled axis, (2) displays with one axis altered in relation to the other axis, and (3) displays in which the vertical axis (y-axis) does not begin with 0. As an example of how a graphical display can be distorted, Figure 2.10 displays a frequency polygon for U.S. unemployment rates in 2019. Figure 2.10a displays the data correctly with the y-axis starting at 0%; Figure 2.10b displays the same data with the y-axis distorted and beginning at 3.2%. When the graph is distorted in this way, it can make the slope of the line appear steeper, as if unemployment rates are substantially declining, although it is clear from Figure 2.10a that this is not the case—although unemployment rates in the U.S. did decline slightly in 2019. To avoid misleading or deceiving readers, pay attention to how data are displayed in graphs to make sure that the data are accurately and appropriately presented.

FIGURE 2.10 ■ Two Graphical Displays for the Same Data



(a) This graph is a correct display and (b) this is a display that is distorted because the y-axis does not begin at 0%. Data are of actual unemployment rates in the United States in 2019.

Source: U.S. Bureau of Labor Statistics, 2020.

Distortion can also occur when presenting summary statistics. Two common distortions to look for with summary statistics are when data are omitted or differences are described in a way that gives the impression of larger differences than really are meaningful in the data. It can sometimes be difficult to determine whether data are misleading or have been omitted, although some data can naturally be reported together—such as reporting the sample size with percentile distributions. For example, if we report that 75% of those surveyed preferred Product A to Product B, you may be inclined to conclude that Product A is a better product. However, if you were also informed that only four people were sampled, then 75% may not seem as convincing. Anytime you read a claim about results in a study, it is important to refer to the data to confirm the extent to which the data support the claim being made by the author or authors of the research study.

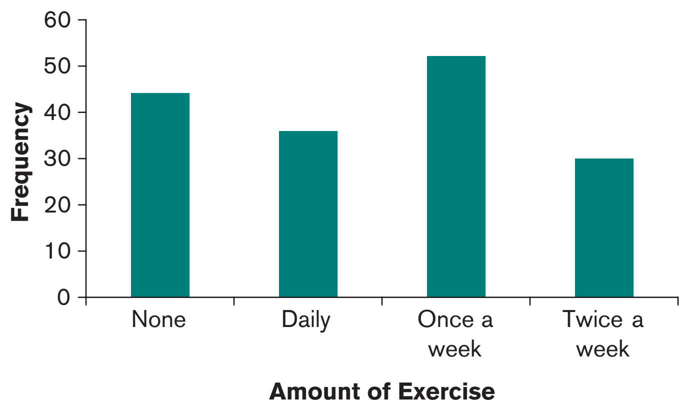
DATA IN RESEARCH

FREQUENCIES AND PERCENTAGES

Although graphs are often used to help the reader understand frequency data, bar charts and histograms are not always equally effective at summarizing percentage data. For example, Hollands and Spence (1992, 1998) asked adult participants to identify relative percentages displayed in bar charts and pie charts (similar to those presented in this chapter). Their studies showed that participants required more time and made larger errors looking at bar charts than when they looked at pie charts. They went on to show that participants also required more time as the number of bars in the graph increased, whereas increasing the number of slices in a pie chart did not have this effect. They explained that most bar graphs, especially for frequency data, are not distributed in percentage units; hence, the reader cannot clearly estimate a proportion by simply viewing the scale. This research suggests that when you want to convey data as percentages, pie charts (and even ogives) would be a better choice for displaying the data.

LEARNING CHECK 2.11

1. Which of the following best describes histograms and bar charts?
   - **a.** They are similar except that the vertical bars do not touch in a bar chart.
   - **b.** They are similar except that the vertical bars do not touch in a histogram.
   - **c.** Both are used to summarize qualitative and discrete data.
2. In the following bar chart summarizing the frequency of exercise among a sample of college students, which category of exercise has the largest frequency?



a. None.
b. Daily.
c. Once a week.
d. Twice a week.

3. The 360 degrees in a pie chart correspond to 100% of a data distribution.
   a. True.
   b. False.

Answers: 1. a; 2. c; 3. a.

2.12 SPSS IN FOCUS: HISTOGRAMS, BAR CHARTS, PIE CHARTS, AND STEM-AND-LEAF DISPLAYS

SPSS LO 2.12 Construct a histogram, bar chart, pie chart, and stem-and-leaf display using SPSS.

We can use SPSS to construct a histogram, bar chart, pie chart, and stem-and-leaf display. Here we will start with data entry (Steps 1 and 2), then construct each graph (Steps 3–6). To review, histograms are used for continuous or quantitative data; bar charts and pie charts are used for discrete, categorical, or qualitative data; and stem-and-leaf displays can be used to summarize data that are discrete or continuous. As an exercise to construct these graphs in SPSS, let us treat the data as a simple set of general values. Suppose we measure the data shown in Table 2.18.

TABLE 2.18 ■ A Sample of 20 Values

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

Because we are not defining these values, we can just call the variable "numbers." Here are the steps:

1. Click on the Variable View tab and enter *numbers* in the Name column. We will enter whole numbers, so go to the Decimals column and reduce the value to 0.
2. Click on the Data View tab and enter the 20 values in the column you labeled *numbers*. You can enter the data in any order you wish, but make sure all the data are entered correctly.
3. To construct a histogram, bar chart, and pie chart, go to the menu bar and click Analyze, then Descriptive Statistics and Frequencies, to display a dialog box.
4. In the dialog box, select the *numbers* variable and click the arrow in the center to move *numbers* into the box on the right labeled Variable(s). Because we only want the graphs and charts in this example, make sure the option to display frequency tables is not selected.

5. Click the Charts option to display a new dialog box. In the dialog box, you have the option to select bar charts, pie charts, or histograms. Select each option to explore how each is displayed; however, you can only select one option at a time. After you make your selection, click Continue.
6. Select OK, or select Paste and click the Run command to construct each graph.

In this example, you can also display the frequency table by keeping the option to display frequency tables selected. Also, for many graphical options, such as histograms, there are other commands you can choose as well. For example, selecting Analyze, then Descriptive Statistics and Explore will give you the option for not only a histogram in the Plots option, but also gives you the option to create a stem-and-leaf display for your data. In this way, SPSS gives you many options for summarizing data using tables and graphs.

LEARNING CHECK 2.12

1. Using SPSS, which options in the menu bar do you select to construct a histogram, bar chart, and pie chart?
   a. Analyze, then Descriptive Statistics and Explore
   b. Analyze, then Descriptive Statistics and Descriptives
   c. Analyze, then Descriptive Statistics and Frequencies
2. Using SPSS, which options in the menu bar can you select to also have the option to create a stem-and-leaf display?
   a. Analyze, then Descriptive Statistics and Explore
   b. Analyze, then Descriptive Statistics and Descriptives
   c. Analyze, then Descriptive Statistics and Frequencies
3. True or false: SPSS gives you many options for summarizing data using tables and graphs.
   a. True.
   b. False.

Answers: 1. c; 2. a; 3. a.

CHAPTER SUMMARY

LO 2.1 Define frequency and explain why it is valuable to summarize data.

- A frequency is the number of times or how often a category, score, or range of scores occurs.
- Summarizing data is valuable because it can make the presentation and interpretation of a distribution of data clearer. For large data sets, summarizing data can be even more valuable to clarify patterns or outcomes in the data that may otherwise go unnoticed if only the raw data set were presented.

LO 2.2 Construct a simple frequency distribution for grouped data.

- A frequency distribution is a summary display for a distribution of data organized or summarized in terms of how often or frequently scores occur.
- A simple frequency distribution for grouped data displays the frequency of data in intervals. Each interval is equidistant, no interval overlaps, and the degree of accuracy for each interval is the same as in the original data. This distribution can be constructed using three steps:
  Step 1: Find the real range.
  Step 2: Find the interval width.
  Step 3: Construct the frequency distribution.

LO 2.3 Construct and explain when it is appropriate to use a cumulative frequency, relative frequency, relative percentage, cumulative relative frequency, and cumulative percentage distribution.

- A cumulative frequency is a summary display that distributes the sum of frequencies across a series of intervals. You can sum from the top or the bottom depending on how you want to discuss the data. You add from the bottom up when discussing the data in terms of "less than" or "at or below" a certain value or "at most." You add from the top down when discussing the data in terms of "greater than" or "at or above" a certain value or "at least."
- A relative frequency is a summary display that distributes the proportion of scores in each interval. To compute a relative frequency, divide the frequency in each interval by the total number of scores counted. The relative frequency is reported when summarizing large data sets. To convert relative frequencies to relative percentages, multiply each relative frequency by 100. Both summary displays convey the same information.
- Cumulative relative frequencies are summary displays for the sum of relative frequencies from the top down or the bottom up. These can be converted to cumulative relative percentages by multiplying the cumulative relative frequency in each interval by 100. Cumulative relative percentages can be distributed as percentile ranks, which indicate the percentage of scores at or below a given score.

LO 2.4 Identify percentile points and percentile ranks in a cumulative percentage distribution.

- A cumulative percentage distribution identifies percentiles, which are measures of the relative position of individuals or scores within a larger distribution. A percentile, specifically a percentile point, is the value of an individual or score within a larger distribution. The corresponding percentile of a percentile point is the percentile rank of that score.
- To find the percentile point in a cumulative percentage distribution, follow four basic steps:

  Step 1: Identify the interval within which a specified percentile point falls.

  Step 2: Identify the real range for the interval identified.

  Step 3: Find the position of the percentile point within the interval.

  Step 4: Identify the percentile point.

SPSS LO 2.5 Construct a frequency distribution for quantitative data using SPSS.

- SPSS can be used to create frequency distributions for quantitative data. Quantitative data are typically entered by column. Frequency distributions for quantitative data are created using the Analyze, Descriptive Statistics, and Frequencies options in the menu bar (for more details, refer to Section 2.5).

LO 2.6 Construct a frequency distribution for ungrouped or categorical data.

- A simple frequency distribution for ungrouped data displays the frequency of categories or whole units when the number of different values collected is small. Because constructing intervals is not necessary for ungrouped data, skip straight to Step 3 for constructing a simple frequency distribution to construct this type of frequency distribution.

SPSS LO 2.7 **Construct a frequency distribution for categorical data using SPSS.**

- SPSS can be used to create frequency distributions for categorical data. Categorical data (which typically require coding) are entered by row. Frequency distributions for categorical data are created using the Analyze, Descriptive Statistics, and Frequencies options in the menu bar. When the levels of a variable are coded, a Weight cases option must also be selected from the menu bar (for more details, refer to Section 2.7).

LO 2.8 **Identify how presenting images can meaningfully illustrate data.**

- A pictogram is a summary display that uses symbols or illustrations to represent a concept, object, place, or event. Pictures often help people make sense of frequency data, inasmuch as the reader can relate to the images presented. Relating frequency data to images that correspond to or reflect the data being reported can help "bring to life" or illustrate the data in a clearer way.

LO 2.9 **Construct and interpret graphs for distributions of continuous data.**

- A histogram is a graphical display used to summarize the frequency of continuous data distributed in numeric intervals (grouped). Histograms are constructed by distributing the intervals along the x-axis and listing the frequencies of scores on the y-axis, with each interval connected by vertical bars or rectangles. The height of each rectangle reflects the frequency of scores in a given interval. Three rules for constructing histograms are as follows:

  Rule 1: A vertical rectangle represents each interval, and the height of the rectangle equals the frequency recorded for each interval.

  Rule 2: The base of each rectangle begins and ends at the upper and lower boundaries of each interval.

  Rule 3: Each rectangle touches adjacent rectangles at the boundaries of each interval.
- A frequency polygon is a dot-and-line graph in which the dot is the midpoint of each interval and the line connects each dot. The midpoint of an interval is distributed along the x-axis and is calculated by summing the upper and lower boundary of an interval and then dividing by 2.
- An ogive is a dot-and-line graph used to summarize the cumulative percentage of continuous data at the upper boundary of each interval.

LO 2.10 **Summarize discrete or continuous data in a stem-and-leaf display.**

- A stem-and-leaf display is a graphical display in which each individual score from an original set of data is listed. The data are organized such that the common digits shared by all scores are listed to the left (in the stem), with the remaining digits for each score listed to the right (in the leaf).

LO 2.11 **Construct and interpret graphs for distributions of discrete or categorical data.**

- Bar charts are used to summarize discrete and categorical data. Bar charts are similar to histograms, except that the bars or rectangles are separated to indicate discrete units or classes. To construct a bar chart, list the whole units or categories along the x-axis, and distribute the frequencies along the y-axis.
- A pie chart is a graphical display in the shape of a circle that is used to summarize the relative percentage of discrete and categorical data in segments. Converting proportions to a pie chart requires finding the correct angles for each slice of the pie. To find the central angles of each sector (or category), multiply each relative percentage by 3.6 ($100\% \times 3.6 = 360°$).

SPSS LO 2.12 Construct a histogram, bar chart, pie chart, and stem-and-leaf display using SPSS.

- SPSS can be used to create histograms, bar charts, and pie charts. Each graph is created using the Analyze, Descriptive Statistics, and Frequencies options in the menu bar. This option will display a dialog box that will allow you to identify your variable and select the Charts option, which gives you the option to select bar charts, pie charts, or histograms. Select each option to explore how each summary is displayed. Note that selecting Analyze, then Descriptive Statistics and Explore also gives you the option to create a stem-and-leaf display in the Plots option (for more details, refer to Section 2.12).

KEY TERMS

bar chart
bar graph
cumulative frequency distribution
cumulative percentage distribution
cumulative relative frequency distribution
frequency
frequency distribution
frequency polygon
grouped data
histogram
interval
interval boundaries
interval width
leaf
lower boundary
ogive
open interval
outlier
percentile point
percentile rank
pictogram
pictograph
pie chart
proportion
real range
relative frequency distribution
relative percentage distribution
sector
simple frequency distribution
stem
stem-and-leaf display
stem-and-leaf plot
ungrouped data
upper boundary

END-OF-CHAPTER PROBLEMS

1. A professor has 100 students in their class, and 40 students scored between 80 and 89 for a B on the first exam. Which value(s) represents the *frequency* in this example?
   a. 40
   b. 100
   c. 80–89
2. Which of the following explains why it is valuable to summarize data?
   a. It can make the presentation of a distribution of data clearer.
   b. It can make the interpretation of a distribution of data clearer.
   c. Both a and b are correct.
3. Below is the number of times a commercial was shown displaying high fat, high sugar foods during children's programming over a one-month period. For these data, construct a simple frequency distribution for grouped data with four intervals. Which interval has the largest frequency?
   21, 8, 11, 9, 12, 10, 10, 5, 9, 18, 17, 3, 6, 14, 18, 16, 19, 3, 22, 7
   a. 18–22
   b. 13–17
   c. 8–12
   d. 3–7

4. The following table shows a frequency distribution for grouped data. Notice that the frequency of scores, *f*(*x*), does not add up to 15. If a total of 15 scores were actually counted, which of the following gives an explanation for why the frequencies do not add up to 15?

Intervals	Frequency
0–5	4
5–10	6
10–15	3
15–20	5

a. There are too many intervals for the data counted, which might lead to a larger total than the number of scores counted.
b. There are too few intervals for the data counted, which might lead to a smaller total than the number of scores counted.
c. The intervals overlap at the upper and lower boundaries, which might lead to some scores being counted in more than one interval.

5. The lower boundaries for the number of missing assignments among students in a large class are 1, 4, 7, 10, 13, and 16. Which of the following lists the value for each upper boundary in this distribution?
a. 1, 4, 7, 10, 13, and 16
b. 3, 6, 9, 12, 15, and 18
c. 2, 5, 8, 11, 14, and 17

6. The upper boundaries for a distribution of waiting times (in seconds) in a grocery store aisle are 45, 56, 67, and 78. Which of the following lists the value for each lower boundary in this distribution?
a. 35, 46, 57, and 68
b. 45, 56, 67, and 78
c. 44, 55, 66, and 77

A researcher reports the following frequency distribution for the time (in minutes) that 70 college students spent on social networking websites during class time. For Questions 7 and 8, enter the values for A and B to complete the frequency distribution.

Time	Frequency
0–9	14
A–19	18
20–29	**B**
30–39	12

7. A = ___ [enter whole number]

8. B = ___ [enter whole number]

9. When cumulating frequencies from the bottom up, the data are discussed in terms of
a. at most
b. at least
c. at or above

10. When cumulating frequencies from the top down, the data are discussed in terms of
   a. at most
   b. at least
   c. at or below

Use the simple frequency distribution table below to answer Questions 11 and 12.

Intervals	Frequency
44–46	6
41–43	4
38–40	5
35–37	10
32–34	5

11. If we convert this frequency distribution to relative percentages, which of the following gives the corresponding relative percentages in each interval?
   a. 20%, 33%, 50%, 67%, and 100%
   b. 20%, 13%, 17%, 33%, and 17%
   c. 20, .33, .50, .67, and 1.00
   d. 20, .13, .17, .33, and .17

12. What is the cumulative percentage of scores at or below 37 in this distribution?
   Answer: ___% [enter whole number]

13. What is the percentile point at the 90th percentile in the following distribution?

Intervals	Percentile Rank
9–11	100%
6–8	80%
3–5	40%
0–2	20%

   Answer: ___ [enter whole number]

14. A researcher records the number of dreams that 50 college freshman students recalled during the night prior to a final exam. For these data, what is the number of dreams at the 60th percentile? [Hint: First convert this table to a percentile rank distribution.]

Number of Dreams	Cumulative Frequency
4	50
3	44
2	30
1	12
0	5

   Answer: ___ dreams [enter whole number]

Use the following example to answer Questions 15 and 16: A researcher records the duration of time (in seconds) that it took 14 participants to complete an assessment. The SPSS output is given below for the 14 times recorded. [After you answer Question 15, please feel free to enter these data in SPSS and follow the directions given in Section 2.5 to reproduce the SPSS output table shown.]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	60	1	7.1	7.1	7.1
	65	2	14.3	14.3	21.4
	76	3	21.4	21.4	42.9
	80	1	7.1	7.1	50.0
	82	2	14.3	14.3	64.3
	88	1	7.1	7.1	71.4
	92	1	7.1	7.1	78.6
	94	2	14.3	14.3	92.9
	98	1	7.1	7.1	100.0
	Total	14	100.0	100.0	

15. Which of the following lists the original 14 times recorded?
- **a.** 1, 2, 3, 1, 2, 1, 1, 2, and 1.
- **b.** 60, 65, 76, 80, 82, 88, 92, 94, and 98.
- **c.** 60, 65, 65, 76, 76, 76, 80, 82, 88, 88, 92, 94, 94, and 98.
- **d.** 60, 65, 65, 76, 76, 76, 80, 82, 82, 88, 92, 94, 94, and 98.

16. Half of the times recorded fall at or below and half fall above what value in this distribution? Answer: ___ seconds [enter whole number]

17. A researcher observed a rat respond for a food reward by pressing one of three levers in a cage. Pressing the lever to the right (R) produced no food reward, pressing the lever to the left (L) produced a single food pellet, and pressing the lever at the center (C) produced two food pellets. Because the center lever produced the largest reward, the researcher hypothesized that the rat would press this lever most often. Each trial ended when the rat pressed a lever. Below is the number of lever presses recorded for 30 trials. Do these data support the hypothesis?
L, L, R, L, R, C, R, L, C, L, L, C, C, C, R,
C, R, C, L, C, C, L, C, C, C, L, C, C, C, C
- **a.** No, the rat pressed the left lever most often.
- **b.** No, the rat pressed the right lever most often.
- **c.** Yes, the rat pressed the center lever most often.

18. The following is an incomplete frequency distribution table for the number of mistakes made during a series of military combat readiness training exercises. If 0, 1, 2, 3, 4, or 5 mistakes were counted, then which of the following correctly identifies the missing value for A and for B?

Number of Mistakes	Frequency
5	1
4	3
A	2
2	2
1	6
0	8
	N = **B**

a. A = 3; B = 22
b. A = 3; B = 21
c. A = 2; B = 22
d. A = 2; B = 21

19. In a study on romantic relationships, 240 romantically involved participants were asked to choose their preference for an ideal night out with their partner. The frequency of participants choosing (1) dinner and a movie, (2) a sporting event, (3) gambling/gaming, or (4) going out for drinks was recorded. What type of frequency distribution is most appropriate to summarize these data?
a. A simple frequency distribution for grouped data.
b. A simple frequency distribution for ungrouped data.
c. A simple frequency distribution displayed in intervals.

Answer Questions 20 and 21 for the following example: A pollster surveys participants and records the type of dance each participant stated they would be most interested in learning. The SPSS output is given below for the 90 responses recorded. [The table below gives all information needed; please feel free to enter these data in SPSS and follow the directions given in Section 2.7 to reproduce the SPSS output table shown.]

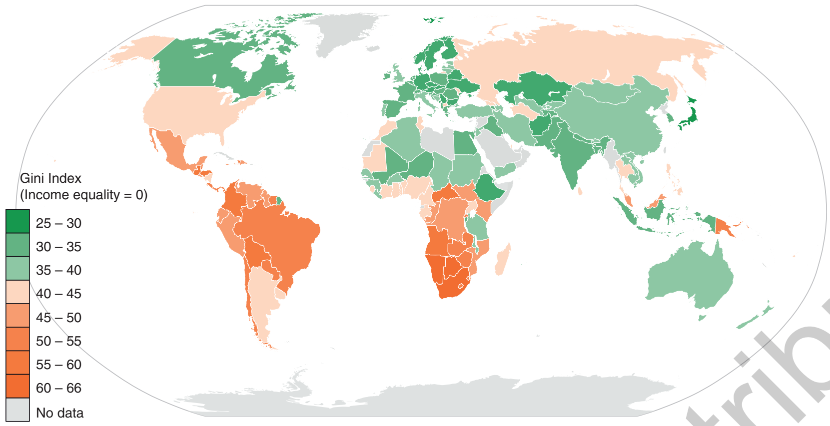
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Salsa	17	18.9	18.9	18.9
	Tango	15	16.7	16.7	35.6
	Cha-Cha	10	11.1	11.1	46.7
	Swing	18	20.0	20.0	66.7
	Hip-Hop	30	33.3	33.3	100.0
	Total	90	100.0	100.0	

20. Which type of dance were participants in this survey most interested in learning?
a. Salsa
b. Tango
c. Cha-Cha
d. Swing
e. Hip-Hop

21. Which type of dance were participants in this survey least interested in learning?
a. Salsa
b. Tango
c. Cha-Cha
d. Swing
e. Hip-Hop

22. What type of summary display uses symbols or illustrations to represent a concept, object, place, or event?
a. pictogram
b. bar chart
c. histogram

Answer Questions 23 and 24 for the following example: The Gini Index is a measure of the extent to which income or wealth is equally distributed in a given population. The lower the Gini Index score, the more equally distributed wealth is in the population. The following figure represents the distribution of wealth based on this measure (reported at http://en.wikipedia.org/wiki/Gini_coefficient).



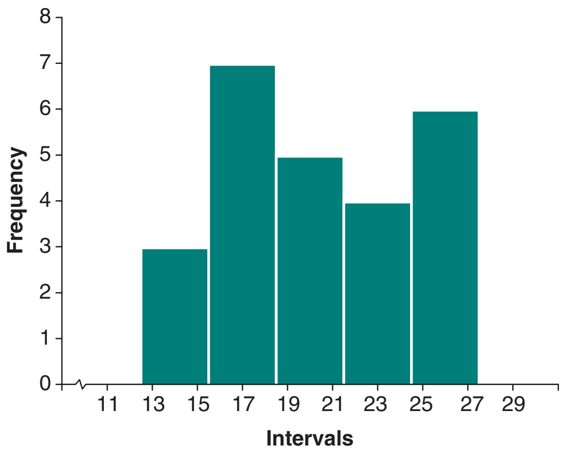
23. In which range does the United States of America fall in terms of income equality using the Gini Index?
  a. 35–40
  b. 40–45
  c. 45–50

24. Which of the following countries has the greatest income equality based on the Gini Index score?
  a. United States of America
  b. South America
  c. Canada

25. A histogram has the following intervals: 2-4, 5-7, 8-10, and 11-13. If the histogram is converted to a frequency polygon, at what value will the interval for 8–10 be plotted? Answer: ___ [enter whole number]

26. Frequency polygons are plotted at the ________ of each interval, whereas ogives are plotted at the ________ of each interval. [Fill in the blanks]
  a. midpoint; upper boundary
  b. upper boundary; midpoint
  c. right; left
  d. left; right

27. The following histogram lists scores contained within intervals from 12.5 to 27.5. For these data, how many scores fall at or below 21.5?



Answer: ___ [enter whole number]

For Questions 28–30, identify the type of graphical display that is most appropriate to summarize the data listed. Answer H for histogram and B for bar chart.

28. Consumer spending (in dollars) during the holidays.

29. The grade point average (GPA) of students graduating from college.

30. The direction of a person's eye movement (right, left, or center).

31. The following set of scores was summarized in a stem-and-leaf display. Which of the following correctly lists the original set of scores in this data set?

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

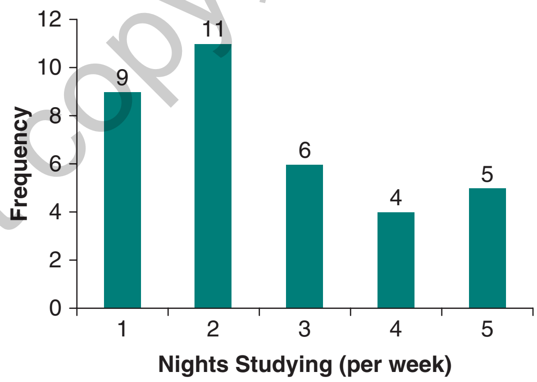
- **a.** 13, 17, 32, 40, 41, 49, 66, and 68.
- **b.** 13, 17, 32, 32, 32, 41, 49, 66, and 68.
- **c.** 13, 17, 32, 32, 32, 40, 41, 49, 66, and 68.

32. A researcher records the following data: 225, 114, 153, 117, 223, 152, 159, 227, 110, 119, 155, 159, 226, 153, 223, 114, 158, 221, 115, 220. These data are summarized in the following stem-and-leaf display. Identify the missing value for A.

1	10	14	14	15	17	19	52	53	**A**	55	58	59	59
2	20	21	23	23	25	26	27						

Answer: ___ [enter whole number]

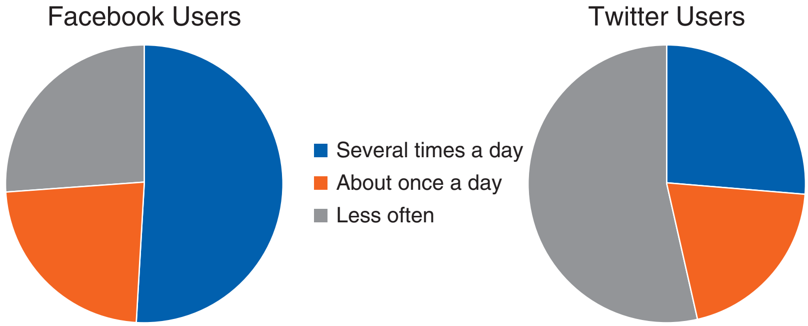
33. The following bar graph summarizes the number of nights per week a sample of college students spent studying. Based on the data shown, how many students studied at least 4 nights per week?



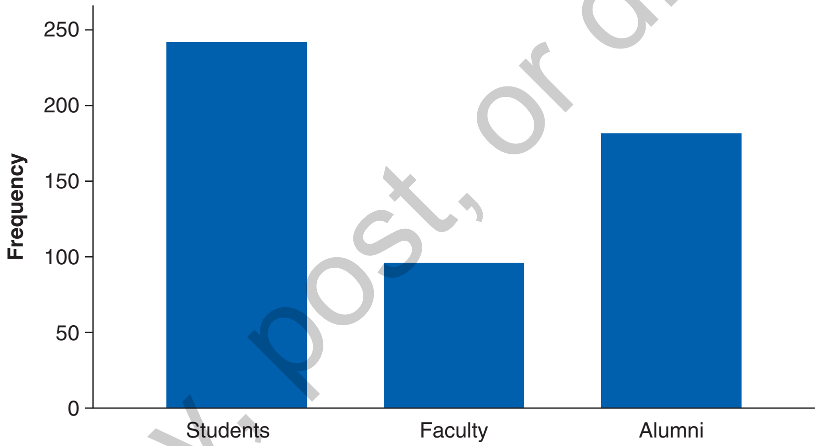
Answer: ___ students [enter whole number]

34. Given below are two pie charts for how often users of Facebook (left) and Twitter (right) visit these platforms, among U.S. adults. Using the pie charts, identify which of the following statements are true for the data summarized (reported by the Pew Research Center at https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/).

- **a.** A majority of Twitter users visit Twitter more often than Facebook users visit Facebook.
- **b.** A majority of Facebook users visit Facebook more often than Twitter users visit Twitter.
- **c.** The number of times per day that U.S. adults use Facebook and Twitter platforms are about the same.



Answer Questions 35 and 36 for the following example: An SPSS output is given below for a graph summarizing the number tweets posted by students, faculty, and school alumni immediately following a big sports victory for their school. The number of tweets sent were 240, 95, and 180 tweets, respectively. [Please feel free to enter these data in SPSS and follow the directions given in Section 2.12 to reproduce the SPSS output shown.]



35. What type of graphical display is shown in the SPSS output image?
- **a.** bar chart
- **b.** histogram
- **c.** pie chart

36. Who sent the greatest number of tweets immediately following the big sports victory?
- **a.** students
- **b.** faculty
- **c.** alumni

Answers for even numbers are in Appendix E.

- The following end-of-chapter problems have full data provided and are designated as problems to test the SPSS learning objectives: 15-16, 20-21, and 35-36.

Visit **edge.sagepub.com/priviterastats4e** to access resources including datasets and SPSS screencasts.