2

Frequency Distributions and Graphs

STATISTICS TODAY

How Your Identity Can Be Stolen

Identity fraud is a big business today—more than 12.7 million people were victims. The total amount of the fraud in 2014 was $16 billion. The average amount of the fraud for a victim is $1260, and the average time to correct the problem is 40 hours. The ways in which a person's identity can be stolen are presented in the following table:

Government documents or benefits fraud	38.7%
Credit card fraud	17.4
Phone or utilities fraud	12.5
Bank fraud	8.2
Attempted identity theft	4.8
Employment-related fraud	4.8
Loan fraud	4.4
Other identity theft	9.2

Source: Javelin Strategy & Research; Council of Better Business Bureau, Inc.

Looking at the numbers presented in a table does not have the same impact as presenting numbers in a well-drawn chart or graph. The article did not include any graphs. This chapter will show you how to construct appropriate graphs to represent data and help you to get your point across to your audience.

See Statistics Today—Revisited at the end of the chapter for some suggestions on how to represent the data graphically.

OUTLINE

OBJECTIVES

After completing this chapter, you should be able to

1. Organize data using a frequency distribution.
2. Represent data in frequency distributions graphically, using histograms, frequency polygons, and ogives.
3. Represent data using bar graphs, Pareto charts, time series graphs, pie graphs, and dotplots.
4. Draw and interpret a stem and leaf plot.

Introduction

When conducting a statistical study, the researcher must gather data for the particular variable under study. For example, if a researcher wishes to study the number of people who were bitten by poisonous snakes in a specific geographic area over the past several years, he or she has to gather the data from various doctors, hospitals, or health departments.

To describe situations, draw conclusions, or make inferences about events, the researcher must organize the data in some meaningful way. The most convenient method of organizing data is to construct a *frequency distribution.*

After organizing the data, the researcher must present them so they can be understood by those who will benefit from reading the study. The most useful method of presenting the data is by constructing *statistical charts* and *graphs.* There are many different types of charts and graphs, and each one has a specific purpose.

This chapter explains how to organize data by constructing frequency distributions and how to present the data by constructing charts and graphs. The charts and graphs illustrated here are histograms, frequency polygons, ogives, pie graphs, Pareto charts, and time series graphs. A graph that combines the characteristics of a frequency distribution and a histogram, called a stem and leaf plot, is also explained.

2–1 Organizing Data

OBJECTIVE 1

Organize data using a frequency distribution.

Suppose a researcher wished to do a study on the ages of the 50 wealthiest people in the world. The researcher first would have to get the data on the ages of the people. In this case, these ages are listed in *Forbes Magazine*. When the data are in original form, they are called **raw data** and are listed next.

45	46	64	57	85
92	51	71	54	48
27	66	76	55	69
54	44	54	75	46
61	68	78	61	83
88	45	89	67	56
81	58	55	62	38
55	56	64	81	38
49	68	91	56	68
46	47	83	71	62

Since little information can be obtained from looking at raw data, the researcher organizes the data into what is called a *frequency distribution.*

Unusual Stats

Of Americans 50 years old and over, 23% think their greatest achievements are still ahead of them.

A **frequency distribution** is the organization of raw data in table form, using classes and frequencies.

Each raw data value is placed into a quantitative or qualitative category called a **class.** The **frequency** of a class then is the number of data values contained in a specific class. A frequency distribution is shown for the preceding data set.

Class limits	Tally	Frequency
27–35	/	1
36–44	///	3
45–53	𝍸 ////	9
54–62	𝍸 𝍸 𝍸	15
63–71	𝍸 𝍸	10
72–80	///	3
81–89	𝍸 //	7
90–98	//	2
		50

Now some general observations can be made from looking at the frequency distribution. For example, it can be stated that the majority of the wealthy people in the study are 45 years old or older.

The classes in this distribution are 27–35, 36–44, etc. These values are called *class limits.* The data values 27, 28, 29, 30, 31, 32, 33, 34, 35 can be tallied in the first class; 36, 37, 38, 39, 40, 41, 42, 43, 44 in the second class; and so on.

Two types of frequency distributions that are most often used are the *categorical frequency distribution* and the *grouped frequency distribution.* The procedures for constructing these distributions are shown now.

Categorical Frequency Distributions

The **categorical frequency distribution** is used for data that can be placed in specific categories, such as nominal- or ordinal-level data. For example, data such as political affiliation, religious affiliation, or major field of study would use categorical frequency distributions.

EXAMPLE 2–1 Distribution of Blood Types

Twenty-five army inductees were given a blood test to determine their blood type. The data set is

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Construct a frequency distribution for the data.

SOLUTION

Since the data are categorical, discrete classes can be used. There are four blood types: A, B, O, and AB. These types will be used as the classes for the distribution.

The procedure for constructing a frequency distribution for categorical data is given next.

Step 1 Make a table as shown.

A Class	B Tally	C Frequency	D Percent
A			
B			
O			
AB			

Step 2 Tally the data and place the results in column B.

Step 3 Count the tallies and place the results in column C.

Step 4 Find the percentage of values in each class by using the formula

$$\% = \frac{f}{n} \cdot 100$$

where f = frequency of the class and n = total number of values. For example, in the class of type A blood, the percentage is

$$\% = \frac{5}{25} \cdot 100 = 20\%$$

Percentages are not normally part of a frequency distribution, but they can be added since they are used in certain types of graphs such as pie graphs. Also, the decimal equivalent of a percent is called a *relative frequency.*

Step 5 Find the totals for columns C (frequency) and D (percent). The completed table is shown. It is a good idea to add the percent column to make sure it sums to 100%. This column won't always sum to 100% because of rounding.

A Class	B Tally	C Frequency	D Percent
A	𝍸	5	20
B	𝍸 //	7	28
O	𝍸 ////	9	36
AB	////	4	16
		Total 25	100%

For the sample, more people have type O blood than any other type.

Grouped Frequency Distributions

When the range of the data is large, the data must be grouped into classes that are more than one unit in width, in what is called a **grouped frequency distribution.** For example, a distribution of the blood glucose levels in milligrams per deciliter (mg/dL) for 50 randomly selected college students is shown.

Unusual Stats

Six percent of Americans say they find life dull.

Class limits	Class boundaries	Tally	Frequency
58–64	57.5–64.5	/	1
65–71	64.5–71.5	𝍸 /	6
72–78	71.5–78.5	𝍸 𝍸	10
79–85	78.5–85.5	𝍸 𝍸 ////	14
86–92	85.5–92.5	𝍸 𝍸 //	12
93–99	92.5–99.5	𝍸	5
100–106	99.5–106.5	//	2
			Total 50

The procedure for constructing the preceding frequency distribution is given in Example 2–2; however, several things should be noted. In this distribution, the values 58 and 64 of the first class are called *class limits.* The **lower class limit** is 58; it represents the smallest data value that can be included in the class. The **upper class limit** is 64; it

represents the largest data value that can be included in the class. The numbers in the second column are called **class boundaries.** These numbers are used to separate the classes so that there are no gaps in the frequency distribution. The gaps are due to the limits; for example, there is a gap between 64 and 65.

Students sometimes have difficulty finding class boundaries when given the class limits. The basic rule of thumb is that *the class limits should have the same decimal place value as the data, but the class boundaries should have one additional place value and end in a 5.* For example, if the values in the data set are whole numbers, such as 59, 68, and 82, the limits for a class might be 58–64, and the boundaries are 57.5–64.5. Find the boundaries by subtracting 0.5 from 58 (the lower class limit) and adding 0.5 to 64 (the upper class limit).

$$\text{Lower limit} - 0.5 = 58 - 0.5 = 57.5 = \text{lower boundary}$$
$$\text{Upper limit} + 0.5 = 64 + 0.5 = 64.5 = \text{upper boundary}$$

Unusual Stats

One out of every hundred people in the United States is color-blind.

If the data are in tenths, such as 6.2, 7.8, and 12.6, the limits for a class hypothetically might be 7.8–8.8, and the boundaries for that class would be 7.75–8.85. Find these values by subtracting 0.05 from 7.8 and adding 0.05 to 8.8.

Class boundaries are not always included in frequency distributions; however, they give a more formal approach to the procedure of organizing data, including the fact that sometimes the data have been rounded. You should be familiar with boundaries since you may encounter them in a statistical study.

Finally, the **class width** for a class in a frequency distribution is found by subtracting the lower (or upper) class limit of one class from the lower (or upper) class limit of the next class. For example, the class width in the preceding distribution on the distribution of blood glucose levels is 7, found from 65 – 58 = 7.

The class width can also be found by subtracting the lower boundary from the upper boundary for any given class. In this case, 64.5 – 57.5 = 7.

Note: Do not subtract the limits of a single class. It will result in an incorrect answer.

The researcher must decide how many classes to use and the width of each class. To construct a frequency distribution, follow these rules:

1. *There should be between 5 and 20 classes.* Although there is no hard-and-fast rule for the number of classes contained in a frequency distribution, it is of utmost importance to have enough classes to present a clear description of the collected data.
2. *It is preferable but not absolutely necessary that the class width be an odd number.* This ensures that the midpoint of each class has the same place value as the data. The **class midpoint** X_m is obtained by adding the lower and upper boundaries and dividing by 2, or adding the lower and upper limits and dividing by 2:

$$X_m = \frac{\text{lower boundary} + \text{upper boundary}}{2}$$

or

$$X_m = \frac{\text{lower limit} + \text{upper limit}}{2}$$

For example, the midpoint of the first class in the example with glucose levels is

$$\frac{57.5 + 64.5}{2} = 61 \quad \text{or} \quad \frac{58 + 64}{2} = 61$$

The midpoint is the numeric location of the center of the class. Midpoints are necessary for graphing (see Section 2–2). If the class width is an even number, the

midpoint is in tenths. For example, if the class width is 6 and the boundaries are 5.5 and 11.5, the midpoint is

$$\frac{5.5 + 11.5}{2} = \frac{17}{2} = 8.5$$

Rule 2 is only a suggestion, and it is not rigorously followed, especially when a computer is used to group data.

3. *The classes must be mutually exclusive.* Mutually exclusive classes have nonoverlapping class limits so that data cannot be placed into two classes. Many times, frequency distributions such as this

Age
10–20
20–30
30–40
40–50

are found in the literature or in surveys. If a person is 40 years old, into which class should she or he be placed? A better way to construct a frequency distribution is to use classes such as

Age
10–20
21–31
32–42
43–53

Recall that boundaries are mutually exclusive. For example, when a class boundary is 5.5 to 10.5, the data values that are included in that class are values from 6 to 10. A data value of 5 goes into the previous class, and a data value of 11 goes into the next-higher class.

4. *The classes must be continuous.* Even if there are no values in a class, the class must be included in the frequency distribution. There should be no gaps in a frequency distribution. The only exception occurs when the class with a zero frequency is the first or last class. A class with a zero frequency at either end can be omitted without affecting the distribution.
5. *The classes must be exhaustive.* There should be enough classes to accommodate all the data.
6. *The classes must be equal in width.* This avoids a distorted view of the data.

   One exception occurs when a distribution has a class that is **open-ended.** That is, the first class has no specific lower limit, or the last class has no specific upper limit. A frequency distribution with an open-ended class is called an **open-ended distribution.** Here are two examples of distributions with open-ended classes.

Age	Frequency
10–20	3
21–31	6
32–42	4
43–53	10
54 and above	8

Minutes	Frequency
Below 110	16
110–114	24
115–119	38
120–124	14
125–129	5

The frequency distribution for age is open-ended for the last class, which means that anybody who is 54 years or older will be tallied in the last class. The distribution for minutes is open-ended for the first class, meaning that any minute values below 110 will be tallied in that class.

The steps for constructing a grouped frequency distribution are summarized in the following Procedure Table.

Procedure Table

Constructing a Grouped Frequency Distribution

Step 1 Determine the classes.
- Find the highest and lowest values.
- Find the range.
- Select the number of classes desired.
- Find the width by dividing the range by the number of classes and rounding up.
- Select a starting point (usually the lowest value or any convenient number less than the lowest value); add the width to get the lower limits.
- Find the upper class limits.
- Find the boundaries.

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies, and find the cumulative frequencies.

Example 2–2 shows the procedure for constructing a grouped frequency distribution, i.e., when the classes contain more than one data value.

EXAMPLE 2–2 Record High Temperatures

These data represent the record high temperatures in degrees Fahrenheit (°F) for each of the 50 states. Construct a grouped frequency distribution for the data, using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Source: The World Almanac and Book of Facts.

SOLUTION

The procedure for constructing a grouped frequency distribution for numerical data follows.

Unusual Stats

America's most popular beverages are soft drinks. It is estimated that, on average, each person drinks about 52 gallons of soft drinks per year, compared to 22 gallons of beer.

Step 1 Determine the classes.

Find the highest value and lowest value: $H = 134$ and $L = 100$.

Find the range: R = highest value − lowest value $= H - L$, so

$$R = 134 - 100 = 34$$

Select the number of classes desired (usually between 5 and 20). In this case, 7 is arbitrarily chosen.

Find the class width by dividing the range by the number of classes.

$$\text{Width} = \frac{R}{\text{number of classes}} = \frac{34}{7} = 4.9$$

Round the answer up to the nearest whole number if there is a remainder: $4.9 \approx 5$. (Rounding *up* is different from rounding *off*. A number is rounded up if there is any decimal remainder when dividing. For example, $85 \div 6 = 14.167$ and is rounded up to 15. Also, $53 \div 4 = 13.25$ and is rounded up to 14. (Also, after dividing, if there is no remainder, you will need to add an extra class to accommodate all the data.)

Select a starting point for the lowest class limit. This can be the smallest data value or any convenient number less than the smallest data value. In this case, 100 is used. Add the width to the lowest score taken as the starting point to get the lower limit of the next class. Keep adding until there are 7 classes, as shown, 100, 105, 110, etc.

Subtract one unit from the lower limit of the second class to get the upper limit of the first class. Then add the width to each upper limit to get all the upper limits.

$$105 - 1 = 104$$

The first class is 100–104, the second class is 105–109, etc.

Find the class boundaries by subtracting 0.5 from each lower class limit and adding 0.5 to each upper class limit:

99.5–104.5, 104.5–109.5, etc.

Historical Note

Florence Nightingale, a nurse in the Crimean War in 1854, used statistics to persuade government officials to improve hospital care of soldiers in order to reduce the death rate from unsanitary conditions in the military hospitals that cared for the wounded soldiers.

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies.

The completed frequency distribution is

Class limits	Class boundaries	Tally	Frequency
100–104	99.5–104.5	//	2
105–109	104.5–109.5	𝍸 ///	8
110–114	109.5–114.5	𝍸𝍸 𝍸𝍸 ///	18
115–119	114.5–119.5	𝍸𝍸 𝍸𝍸 ///	13
120–124	119.5–124.5	𝍸𝍸 //	7
125–129	124.5–129.5	/	1
130–134	129.5–134.5	/	1
			Total 50

The frequency distribution shows that the class 109.5–114.5 contains the largest number of temperatures (18) followed by the class 114.5–119.5 with 13 temperatures. Hence, most of the temperatures (31) fall between 110 and 119°F.

Sometimes it is necessary to use a *cumulative frequency distribution*. A **cumulative frequency distribution** is a distribution that shows the number of data values less than or equal to a specific value (usually an upper boundary). The values are found by adding the frequencies of the classes less than or equal to the upper class boundary of a specific class. This gives an ascending cumulative frequency. In this example, the cumulative frequency for the first class is $0 + 2 = 2$; for the second class it is $0 + 2 + 8 = 10$; for the third class it is $0 + 2 + 8 + 18 = 28$. Naturally, a shorter way to do this would be to just add the cumulative frequency of the class below to the frequency of the given class. For example, the cumulative frequency for the number of data values less than 114.5 can be

found by adding 10 + 18 = 28. The cumulative frequency distribution for the data in this example is as follows:

	Cumulative frequency
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

Cumulative frequencies are used to show how many data values are accumulated up to and including a specific class. In Example 2–2, of the total record high temperatures 28 are less than or equal to 114°F. Forty-eight of the total record high temperatures are less than or equal to 124°F.

After the raw data have been organized into a frequency distribution, it will be analyzed by looking for peaks and extreme values. The peaks show which class or classes have the most data values compared to the other classes. Extreme values, called *outliers,* show large or small data values that are relative to other data values.

When the range of the data values is relatively small, a frequency distribution can be constructed using single data values for each class. This type of distribution is called an **ungrouped frequency distribution** and is shown next.

EXAMPLE 2–3 Hours of Sleep

The data shown represent the number of hours 30 college students said they sleep per night. Construct and analyze a frequency distribution.

8	6	6	8	5	7
7	8	7	6	6	7
9	7	7	6	8	10
6	7	6	7	8	7
7	8	7	8	9	8

SOLUTION

Step 1 Determine the number of classes. Since the range is small (10 − 5 = 5), classes consisting of a single data value can be used. They are 5, 6, 7, 8, 9, and 10.

Note: If the data are continuous, class boundaries can be used. Subtract 0.5 from each class value to get the lower class boundary, and add 0.5 to each class value to get the upper class boundary.

Step 2 Tally the data.

Step 3 From the tallies, find the numerical frequencies and cumulative frequencies. The completed ungrouped frequency distribution is shown.

Class limits	Class boundaries	Tally	Frequency
5	4.5–5.5	/	1
6	5.5–6.5	𝍸 //	7
7	6.5–7.5	𝍸 𝍸 /	11
8	7.5–8.5	𝍸 ///	8
9	8.5–9.5	//	2
10	9.5–10.5	/	1

In this case, 11 students sleep 7 hours a night. Most of the students sleep between 5.5 and 8.5 hours.

The cumulative frequencies are

	Cumulative frequency
Less than 4.5	0
Less than 5.5	1
Less than 6.5	8
Less than 7.5	19
Less than 8.5	27
Less than 9.5	29
Less than 10.5	30

When you are constructing a frequency distribution, the guidelines presented in this section should be followed. However, you can construct several different but correct frequency distributions for the same data by using a different class width, a different number of classes, or a different starting point.

Interesting Fact

Male dogs bite children more often than female dogs do; however, female cats bite children more often than male cats do.

Furthermore, the method shown here for constructing a frequency distribution is not unique, and there are other ways of constructing one. Slight variations exist, especially in computer packages. But regardless of what methods are used, classes should be mutually exclusive, continuous, exhaustive, and of equal width.

In summary, the different types of frequency distributions were shown in this section. The first type, shown in Example 2–1, is used when the data are categorical (nominal), such as blood type or political affiliation. This type is called a categorical frequency distribution. The second type of distribution is used when the range is large and classes several units in width are needed. This type is called a grouped frequency distribution and is shown in Example 2–2. Another type of distribution is used for numerical data and when the range of data is small, as shown in Example 2–3. Since each class is only one unit, this distribution is called an ungrouped frequency distribution.

All the different types of distributions are used in statistics and are helpful when one is organizing and presenting data.

The reasons for constructing a frequency distribution are as follows:

1. To organize the data in a meaningful, intelligible way.
2. To enable the reader to determine the nature or shape of the distribution.
3. To facilitate computational procedures for measures of average and spread (shown in Sections 3–1 and 3–2).

4. To enable the researcher to draw charts and graphs for the presentation of data (shown in Section 2–2).
5. To enable the reader to make comparisons among different data sets.

The factors used to analyze a frequency distribution are essentially the same as those used to analyze histograms and frequency polygons, which are shown in Section 2–2.

Applying the Concepts 2–1

Ages of Presidents at Inauguration

The data represent the ages of our Presidents at the time they were first inaugurated.

57	61	57	57	58	57	61	54	68
51	49	64	50	48	65	52	56	46
54	49	51	47	55	55	54	42	51
56	55	51	54	51	60	62	43	55
56	61	52	69	64	46	54	47	

1. Were the data obtained from a population or a sample? Explain your answer.
2. What was the age of the oldest President?
3. What was the age of the youngest President?
4. Construct a frequency distribution for the data. (Use your own judgment as to the number of classes and class size.)
5. Are there any peaks in the distribution?
6. ldentify any possible outliers.
7. Write a brief summary of the nature of the data as shown in the frequency distribution.

See page 108 for the answers.

Exercises 2–1

1. List five reasons for organizing data into a frequency distribution.
2. Name the three types of frequency distributions, and explain when each should be used.
3. How many classes should frequency distributions have? Why should the class width be an odd number?
4. What are open-ended frequency distributions? Why are they necessary?

For Exercises 5–8, find the class boundaries, midpoints, and widths for each class.

5. 58–62
6. 125–131
7. 16.35–18.46
8. 16.3–18.5

For Exercises 9–12, show frequency distributions that are incorrectly constructed. State the reasons why they are wrong.

9.

Class	Frequency
10–19	1
20–29	2
30–34	0
35–45	5
46–51	8

10.

Class	Frequency
5–9	1
9–13	2
13–17	5
17–20	6
20–24	3

11.

Class	Frequency
162–164	3
165–167	7
168–170	18
174–176	0
177–179	5

12.

Class	Frequency
9–13	1
14–19	6
20–25	2
26–28	5
29–32	9

13. Favorite Coffee Flavor A survey was taken asking the favorite flavor of a coffee drink a person prefers. The responses were V = Vanilla, C = Caramel, M = Mocha, H = Hazelnut, and P = Plain. Construct a categorical frequency distribution for the data. Which class has the most data values and which class has the fewest data values?

V	C	P	P	M	M	P	P	M	C
M	M	V	M	M	M	V	M	M	M
P	V	C	M	V	M	C	P	M	P
M	M	M	P	M	M	C	V	M	C
C	P	M	P	M	H	H	P	H	P

14. Trust in Internet Information A survey was taken on how much trust people place in the information they read on the Internet. Construct a categorical frequency distribution for the data. A = trust in all that they read, M = trust in most of what they read, H = trust in about one-half of what they read, S = trust in a small portion of what they read. (Based on information from the *UCLA Internet Report.*)

M	M	M	A	H	M	S	M	H	M
S	M	M	M	M	A	M	M	A	M
M	M	H	M	M	M	H	M	H	M
A	M	M	M	H	M	M	M	M	M

15. Eating at Fast Food Restaurants A survey was taken of 50 individuals. They were asked how many days per week they ate at a fast-food restaurant. Construct a frequency distribution using 8 classes (0–7). Based on the distribution, how often did most people eat at a fast-food restaurant?

1	3	4	0	4
5	2	2	3	1
2	2	2	2	2
2	2	2	2	3
2	2	5	2	4
2	4	5	2	1
4	1	3	2	2
2	0	7	2	3
2	2	2	5	2
3	3	4	1	3

16. Ages of Dogs The ages of 20 dogs in a pet shelter are shown. Construct a frequency distribution using 7 classes.

5	8	7	6	3
9	4	4	5	8
7	4	7	5	7
3	5	8	4	9

17. Maximum Wind Speeds The data show the maximum wind speeds in miles per hour recorded for 40 states. Construct a frequency distribution using 7 classes.

59	78	62	72	67
76	92	77	64	83
64	70	67	75	75
78	75	71	72	93
68	69	76	72	85
64	70	77	74	72
53	67	48	76	59
87	53	77	70	63

Source: NOAA

18. Stories in the World's Tallest Buildings The number of stories in each of a sample of the world's 30 tallest buildings follows. Construct a grouped frequency distribution and a cumulative frequency distribution with 7 classes.

88	88	110	88	80	69	102	78	70	55
79	85	80	100	60	90	77	55	75	55
54	60	75	64	105	56	71	70	65	72

Source: New York Times Almanac.

19. Ages of Declaration of Independence Signers The ages of the signers of the Declaration of Independence are shown. (Age is approximate since only the birth year appeared in the source, and one has been omitted since his birth year is unknown.) Construct a grouped frequency distribution and a cumulative frequency distribution for the data, using 7 classes.

41	54	47	40	39	35	50	37	49	42	70	32
44	52	39	50	40	30	34	69	39	45	33	42
44	63	60	27	42	34	50	42	52	38	36	45
35	43	48	46	31	27	55	63	46	33	60	62
35	46	45	34	53	50	50					

Source: The Universal Almanac.

20. Salaries of Governors Here are the salaries (in dollars) of the governors of 25 randomly selected states. Construct a grouped frequency distribution with 6 classes.

112,895	117,312	140,533	110,000	115,331
95,000	177,500	120,303	139,590	150,000
173,987	130,000	133,821	144,269	142,542
150,000	145,885	105,000	93,600	166,891
130,273	70,000	113,834	117,817	137,092

Source: World Almanac.

21. Charity Donations A random sample of 30 large companies in the United States shows the amount,

in millions of dollars, that each company donated to charity for a specific year. Construct a frequency distribution for the data, using 9 classes.

26	25	19	31	14
48	35	43	25	46
17	21	57	58	34
41	12	27	15	53
16	63	82	23	52
56	75	19	26	88

22. Unclaimed Expired Prizes The number of unclaimed expired prizes (in millions of dollars) for lottery tickets bought in a sample of states is shown. Construct a frequency distribution for the data, using 5 classes.

28.5	51.7	19	5
2	1.2	14	14.6
0.8	11.6	3.5	30.1
1.7	1.3	13	14

23. Scores in the Rose Bowl The data show the scores of the winning teams in the Rose Bowl. Construct a frequency distribution for the data using a class width of 7.

24	20	45	21	26	38	49	32	41	38
28	34	37	34	17	38	21	20	41	38
21	38	34	46	17	22	20	22	45	20
45	24	28	23	17	17	27	14	23	18

Source: The World Almanac.

24. Consumption of Natural Gas Construct a frequency distribution for the energy consumption of natural gas (in billions of Btu) by the 50 states and the District of Columbia. Use 9 classes.

474	475	205	639	197	344	3	409	247	66
377	87	747	1166	223	248	958	406	251	3462
2391	514	371	58	224	530	317	267	769	9
188	289	76	678	331	52	214	165	255	319
34	1300	284	834	114	1082	73	62	95	393
146									

Source: Time Almanac.

25. Average Wind Speeds A sample of 40 large cities was selected, and the average of the wind speeds was computed for each city over one year. Construct a frequency distribution, using 7 classes.

12.2	9.1	11.2	9.0
10.5	8.2	8.9	12.2
9.5	10.2	7.1	11.0
6.2	7.9	8.7	8.4
8.9	8.8	7.1	10.1
8.7	10.5	10.2	10.7
7.9	8.3	8.7	8.7
10.4	7.7	12.3	10.7
7.7	7.8	11.8	10.5
9.6	9.6	8.6	10.3

Source: World Almanac and Book of Facts.

26. Percentage of People Who Completed 4 or More Years of College Listed by state are the percentages of the population who have completed 4 or more years of a college education. Construct a frequency distribution with 7 classes.

21.4	26.0	25.3	19.3	29.5	35.0	34.7	26.1	25.8	23.4
27.1	29.2	24.5	29.5	22.1	24.3	28.8	20.0	20.4	26.7
35.2	37.9	24.7	31.0	18.9	24.5	27.0	27.5	21.8	32.5
33.9	24.8	31.7	25.6	25.7	24.1	22.8	28.3	25.8	29.8
23.5	25.0	21.8	25.2	28.7	33.6	33.6	30.3	17.3	25.4

Source: New York Times Almanac.

Extending the Concepts

27. JFK Assassination A researcher conducted a survey asking people if they believed more than one person was involved in the assassination of John F. Kennedy. The results were as follows: 73% said yes, 19% said no, and 9% had no opinion. Is there anything suspicious about the results?

28. The Value of Pi The ratio of the circumference of a circle to its diameter is known as π (pi). The value of π is an irrational number, which means that the decimal part goes on forever and there is no fixed sequence of numbers that repeats. People have found the decimal part of π to over a million places. We can statistically study the number. Shown here is the value of π to 40 decimal places. Construct an ungrouped frequency distribution for the digits. Based on the distribution, do you think each digit appears equally in the number?

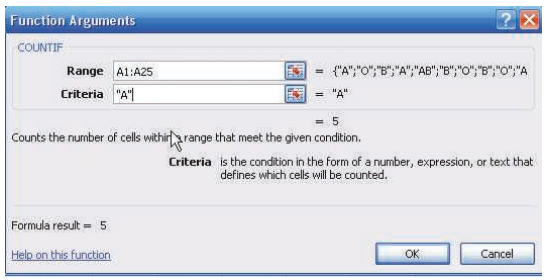
3.1415926535897932384626433832795028841971

Technology Step by Step

EXCEL Step by Step

Categorical Frequency Table (Qualitative or Discrete Data)

1. In an open workbook, select cell A1 and type in all the blood types from Example 2–1 down column A.
2. Type in the variable name **Blood Type** in cell B1.
3. Select cell B2 and type in the four different blood types down the column.
4. Type in the name **Count** in cell C1.
5. Select cell C2. From the toolbar, select the Formulas tab on the toolbar.
6. Select the Insert Function icon , then select the Statistical category in the Insert Function dialog box.
7. Select the Countif function from the function name list.
8. In the dialog box, type **A1:A25** in the **Range** box. Type in the blood type "A" in quotes in the **Criteria** box. The count or frequency of the number of data corresponding to the blood type should appear below the input. Repeat for the remaining blood types.
9. After all the data have been counted, select cell C6 in the worksheet.
10. From the toolbar select Formulas, then AutoSum and type in C2:C5 to insert the total frequency into cell C6.



After entering data or a heading into a worksheet, you can change the width of a column to fit the input. To automatically change the width of a column to fit the data:

1. Select the column or columns that you want to change.
2. On the Home tab, in the Cells group, select Format.
3. Under Cell Size, click Autofit Column Width.

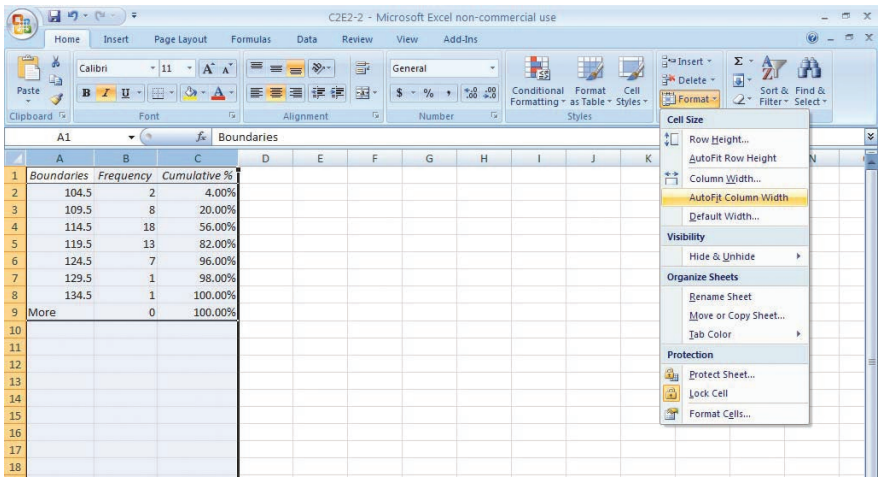
Making a Grouped Frequency Distribution (Quantitative Data)

1. Press [**Ctrl**]-**N** for a new workbook.
2. Enter the raw data from Example 2–2 in column A, one number per cell.
3. Enter the upper class boundaries in column B.
4. From the toolbar select the Data tab, then click Data Analysis.
5. In the Analysis Tools, select Histogram and click [OK].
6. In the Histogram dialog box, type **A1:A50** in the Input Range box and type **B1:B7** in the Bin Range box.
7. Select New Worksheet Ply, and check the Cumulative Percentage option. Click [OK].
8. You can change the label for the column containing the upper class boundaries and expand the width of the columns automatically after relabeling:

   Select the Home tab from the toolbar.

Highlight the columns that you want to change.

Select Format, then AutoFit Column Width.



Note: By leaving the Chart Output unchecked, a new worksheet will display the table only.

MINITAB

Step by Step

Make a Categorical Frequency Table (Qualitative or Discrete Data)

1. Type in all the blood types from Example 2–1 down C1 of the worksheet.
   A B B AB O O O B AB B B B O A O A O O O AB AB A O B A
2. Click above row 1 and name the column **BloodType.**
3. Select **Stat>Tables>Tally Individual Values.**
   The cursor should be blinking in the Variables dialog box. If not, click inside the dialog box.
4. Double-click C1 in the Variables list.
5. Check the boxes for the statistics: Counts, Percents, and Cumulative percents.
6. Click [OK]. The results will be displayed in the Session Window as shown.

Tally for Discrete Variables: BloodType

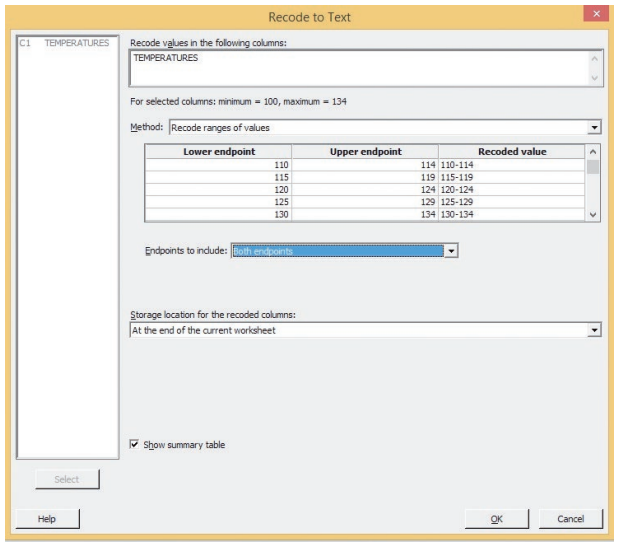
BloodType	Count	Percent	CumPct
A	5	20.00	20.00
AB	4	16.00	36.00
B	7	28.00	64.00
O	9	36.00	100.00
N=	25		

Make a Grouped Frequency Distribution (Quantitative Variable)

1. Select **File>New>Minitab Worksheet**. A new worksheet will be added to the project.
2. Type the data used in Example 2–2 into C1. Name the column **TEMPERATURES.**
3. Use the instructions in the textbook to determine the class limits of 100 to 134 in increments of 5.
   In the next step you will create a new column of data, converting the numeric variable to text categories that can be tallied.

4. Select **Data>Recode>to Text**.
   a) The cursor should be blinking in Recode values in the following columns. If not, click inside the box, then double-click C1 Temperatures in the list. Only quantitative variables will be shown in this list.
   b) Click inside the Method: box and select **Recode ranges of values.**
   c) Press [Tab] to move to the table.
   d) Type 100 in the Lower endpoint column, press [Tab], type 104 in the Upper endpoint column.
   e) Press [Tab] to move to the Recoded value column, and type the text category **100–104.**
   f) Continue to tab to each dialog box, typing the lower endpoint and upper endpoint and then the category until the last category has been entered.
   g) Click inside the Endpoints to include: box and select **Both endpoints.**

   The dialog box should look like the one shown.



5. Click [OK]. In the worksheet, a new column of data will be created in the first empty column, C2. This new variable will contain the category for each value in C1. The column C2-T contains alphanumeric data.
6. Click Stat>Tables>Tally Individual Values, then double-click Recoded TEMPERATURES in the Variables list.
   a) Check the boxes for the desired statistics, such as Counts, Percents, and Cumulative percents.
   b) Click [OK].

   The table will be displayed in the Session Window. Eighteen states have high temperatures between 110 and 114°F. Eighty-two percent of the states have record high temperatures less than or equal to 119°F.

Tally for Discrete Variables: Recoded TEMPERATURES

Recoded TEMPERATURES	Count	Percent	CumPct
100–104	2	4.00	4.00
105–109	8	16.00	20.00
110–114	18	36.00	56.00
115–119	13	26.00	82.00
120–124	7	14.00	96.00
125–129	1	2.00	98.00
130–134	1	2.00	100.00
N=	50		

7. Click File>Save Project As . . . , and type the name of the project file, **Ch2-1.** This will save the two worksheets and the Session Window.

2–2 Histograms, Frequency Polygons, and Ogives

OBJECTIVE 2

Represent data in frequency distributions graphically, using histograms, frequency polygons, and ogives.

After you have organized the data into a frequency distribution, you can present them in graphical form. The purpose of graphs in statistics is to convey the data to the viewers in pictorial form. It is easier for most people to comprehend the meaning of data presented graphically than data presented numerically in tables or frequency distributions. This is especially true if the users have little or no statistical knowledge.

Statistical graphs can be used to describe the data set or to analyze it. Graphs are also useful in getting the audience's attention in a publication or a speaking presentation. They can be used to discuss an issue, reinforce a critical point, or summarize a data set. They can also be used to discover a trend or pattern in a situation over a period of time.

The three most commonly used graphs in research are

1. The histogram.
2. The frequency polygon.
3. The cumulative frequency graph, or ogive (pronounced o-jive).

The steps for constructing the histogram, frequency polygon, and the ogive are summarized in the procedure table.

Procedure Table

Constructing a Histogram, Frequency Polygon, and Ogive

Step 1 Draw and label the x and y axes.

Step 2 On the x axis, label the class boundaries of the frequency distribution for the histogram and ogive. Label the midpoints for the frequency polygon.

Step 3 Plot the frequencies for each class, and draw the vertical bars for the histogram and the lines for the frequency polygon and ogive.

(*Note:* Remember that the lines for the frequency polygon begin and end on the x axis while the lines for the ogive begin on the x axis.)

Historical Note

Karl Pearson introduced the histogram in 1891. He used it to show time concepts of various reigns of Prime Ministers.

The Histogram

The **histogram** is a graph that displays the data by using contiguous vertical bars (unless the frequency of a class is 0) of various heights to represent the frequencies of the classes.

EXAMPLE 2–4 Record High Temperatures

Construct a histogram to represent the data shown for the record high temperatures for each of the 50 states (see Example 2–2).

Class boundaries	Frequency
99.5–104.5	2
104.5–109.5	8
109.5–114.5	18
114.5–119.5	13
119.5–124.5	7
124.5–129.5	1
129.5–134.5	1

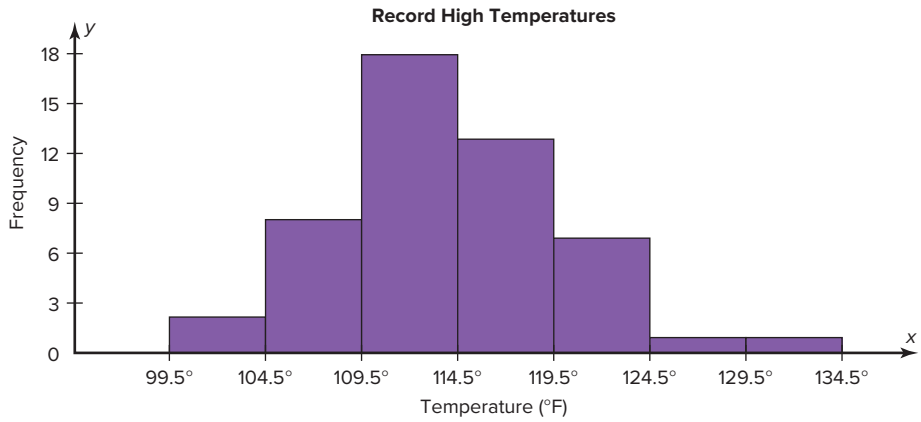
SOLUTION

Step 1 Draw and label the x and y axes. The x axis is always the horizontal axis, and the y axis is always the vertical axis.

Step 2 Represent the frequency on the y axis and the class boundaries on the x axis.

Step 3 Using the frequencies as the heights, draw vertical bars for each class. See Figure 2–1.

FIGURE 2–1 Histogram for Example 2–4



As the histogram shows, the class with the greatest number of data values (18) is 109.5–114.5, followed by 13 for 114.5–119.5. The graph also has one peak with the data clustering around it.

Historical Note

Graphs originated when ancient astronomers drew the position of the stars in the heavens. Roman surveyors also used coordinates to locate landmarks on their maps.

The development of statistical graphs can be traced to William Playfair (1759–1823), an engineer and drafter who used graphs to present economic data pictorially.

The Frequency Polygon

Another way to represent the same data set is by using a frequency polygon.

The **frequency polygon** is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points.

Example 2–5 shows the procedure for constructing a frequency polygon. Be sure to begin and end on the x axis.

EXAMPLE 2–5 Record High Temperatures

Using the frequency distribution given in Example 2–4, construct a frequency polygon.

SOLUTION

Step 1 Find the midpoints of each class. Recall that midpoints are found by adding the upper and lower boundaries and dividing by 2:

$$\frac{99.5 + 104.5}{2} = 102 \qquad \frac{104.5 + 109.5}{2} = 107$$

and so on. The midpoints are

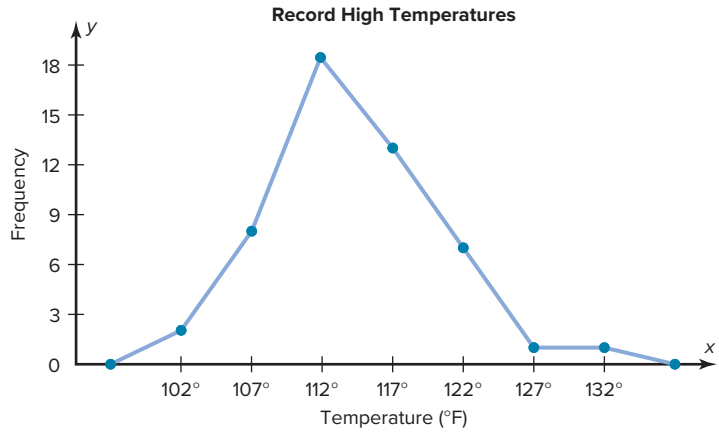
Class boundaries	Midpoints	Frequency
99.5–104.5	102	2
104.5–109.5	107	8
109.5–114.5	112	18
114.5–119.5	117	13
119.5–124.5	122	7
124.5–129.5	127	1
129.5–134.5	132	1

Step 2 Draw the x and y axes. Label the x axis with the midpoint of each class, and then use a suitable scale on the y axis for the frequencies.

Step 3 Using the midpoints for the x values and the frequencies as the y values, plot the points.

Step 4 Connect adjacent points with line segments. Draw a line back to the x axis at the beginning and end of the graph, at the same distance that the previous and next midpoints would be located, as shown in Figure 2–2.

FIGURE 2–2
Frequency Polygon for Example 2–5



The frequency polygon and the histogram are two different ways to represent the same data set. The choice of which one to use is left to the discretion of the researcher.

The Ogive

The third type of graph that can be used represents the cumulative frequencies for the classes. This type of graph is called the *cumulative frequency graph,* or *ogive.* The **cumulative frequency** is the sum of the frequencies accumulated up to the upper boundary of a class in the distribution.

The **ogive** is a graph that represents the cumulative frequencies for the classes in a frequency distribution.

Example 2–6 shows the procedure for constructing an ogive. Be sure to start on the x axis.

EXAMPLE 2–6 Record High Temperatures

Construct an ogive for the frequency distribution described in Example 2–4.

SOLUTION

Step 1 Find the cumulative frequency for each class.

	Cumulative frequency
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

Step 2 Draw the *x* and *y* axes. Label the *x* axis with the class boundaries. Use an appropriate scale for the *y* axis to represent the cumulative frequencies. (Depending on the numbers in the cumulative frequency columns, scales such as 0, 1, 2, 3, . . . , or 5, 10, 15, 20, . . . , or 1000, 2000, 3000, . . . can be used. Do *not* label the *y* axis with the numbers in the cumulative frequency column.) In this example, a scale of 0, 5, 10, 15, . . . will be used.

Step 3 Plot the cumulative frequency at each upper class boundary, as shown in Figure 2–3. Upper boundaries are used since the cumulative frequencies represent the number of data values accumulated up to the upper boundary of each class.

Step 4 Starting with the first upper class boundary, 104.5, connect adjacent points with line segments, as shown in Figure 2–4. Then extend the graph to the first lower class boundary, 99.5, on the *x* axis.

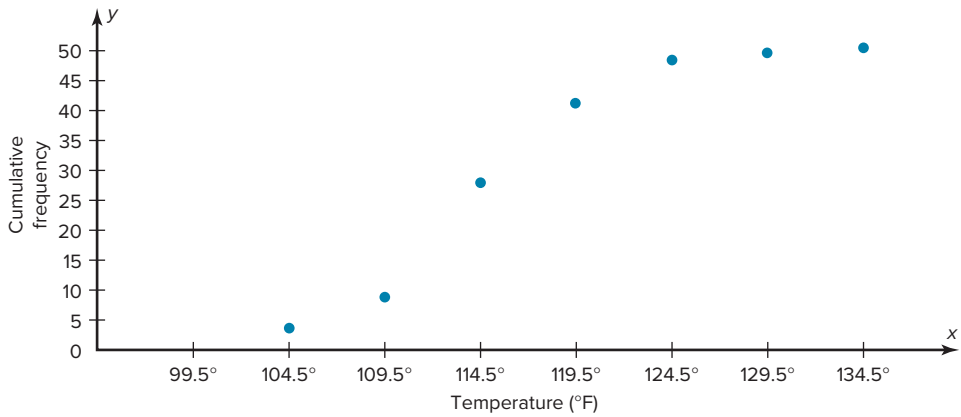


FIGURE 2–3
Plotting the Cumulative Frequency for Example 2–6

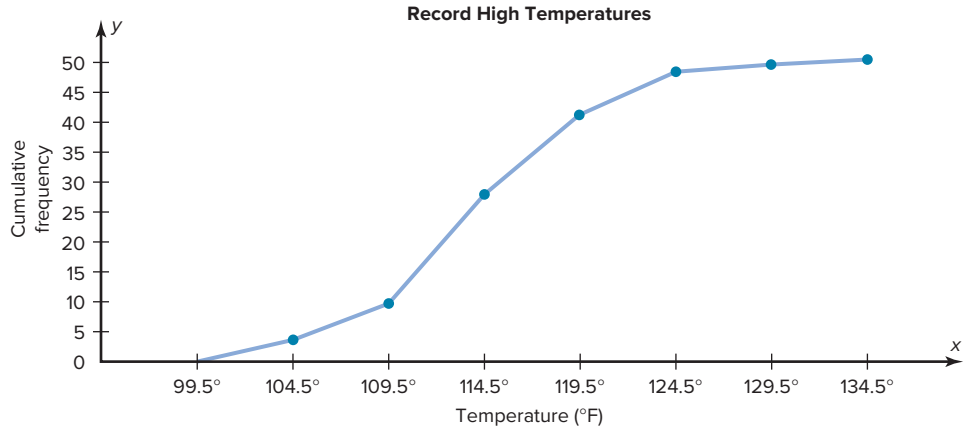
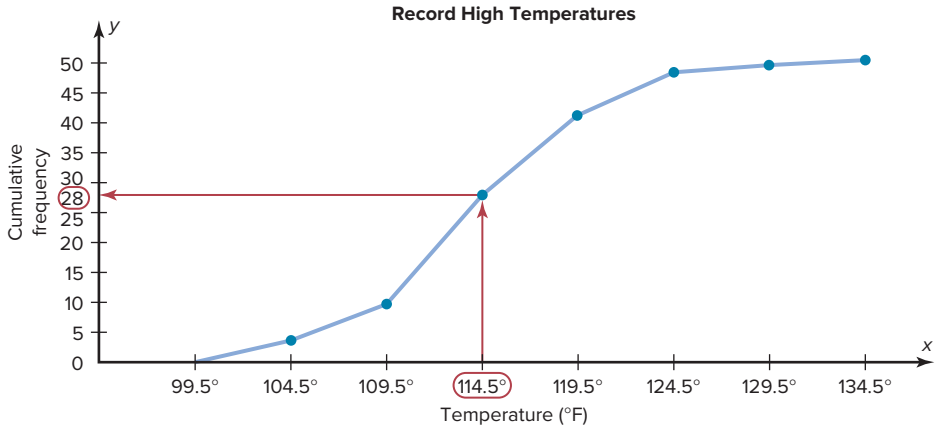


FIGURE 2–4
Ogive for Example 2–6

FIGURE 2–5
Finding a Specific Cumulative Frequency



Unusual Stats
Twenty-two percent of Americans sleep 6 hours a day or less.

Cumulative frequency graphs are used to visually represent how many values are below a certain upper class boundary. For example, to find out how many record high temperatures are less than 114.5°F, locate 114.5°F on the *x* axis, draw a vertical line up until it intersects the graph, and then draw a horizontal line at that point to the *y* axis. The *y* axis value is 28, as shown in Figure 2–5.

Relative Frequency Graphs

The histogram, the frequency polygon, and the ogive shown previously were constructed by using frequencies in terms of the raw data. These distributions can be converted to distributions using *proportions* instead of raw data as frequencies. These types of graphs are called **relative frequency graphs.**

Graphs of relative frequencies instead of frequencies are used when the proportion of data values that fall into a given class is more important than the actual number of data values that fall into that class. For example, if you wanted to compare the age distribution of adults in Philadelphia, Pennsylvania, with the age distribution of adults of Erie, Pennsylvania, you would use relative frequency distributions. The reason is that since the population of Philadelphia is 1,526,006 and the population of Erie is 101,786, the bars using the actual data values for Philadelphia would be much taller than those for the same classes for Erie.

To convert a frequency into a proportion or relative frequency, divide the frequency for each class by the total of the frequencies. The sum of the relative frequencies will always be 1. These graphs are similar to the ones that use raw data as frequencies, but the values on the *y* axis are in terms of proportions. Example 2–7 shows the three types of relative frequency graphs.

EXAMPLE 2–7 Ages of State Governors

Construct at histogram, frequency polygon, and ogive using relative frequencies for the distribution shown. This is a grouped frequency distribution using the ages (at the time of this writing) of the governors of the 50 states of the United States.

Class boundaries	Frequency
42.5–47.5	4
47.5–52.5	4
52.5–57.5	11
57.5–62.5	14
62.5–67.5	9
67.5–72.5	5
72.5–77.5	3
	Total 50

SOLUTION

Step 1 Convert each frequency to a proportion or relative frequency by dividing the frequency for each class by the total number of observations.

For the class 42.5–47.5 the relative frequency $= \frac{4}{50} = 0.08$; for the class 47.5–52.5, the relative frequency is $\frac{4}{50} = 0.08$; for the class 52.5–57.5, the relative frequency is $\frac{11}{50} = 0.22$, and so on.

Place these values in the column labeled Relative Frequency. Also, find the midpoints, as shown in Example 2-5, for each class and place them in the midpoint column

Class boundaries	Midpoints	Relative frequency
42.5–47.5	45	0.08
47.5–52.5	50	0.08
52.5–57.5	55	0.22
57.5–62.5	60	0.28
62.5–67.5	65	0.18
67.5–72.5	70	0.10
72.5–77.5	75	0.06

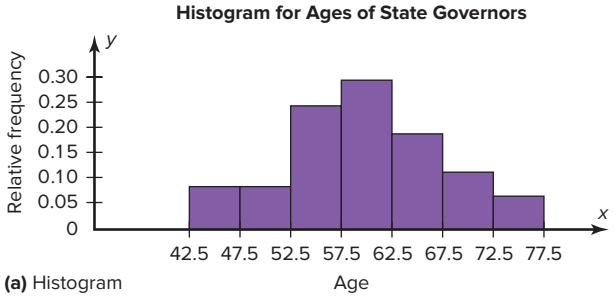
Step 2 Find the cumulative relative frequencies. To do this, add the frequency in each class to the total frequency of the preceding class. In this case, 0.00 + 0.08 = 0.08, 0.08 + 0.08 = 0.16, 0.16 + 0.22 = 0.38, 0.28 + 0.38 = 0.66, etc. Place these values in a column labeled Cumulative relative frequency.

An alternative method would be to change the cumulative frequencies for the classes to relative frequencies. (Divide each by the total).

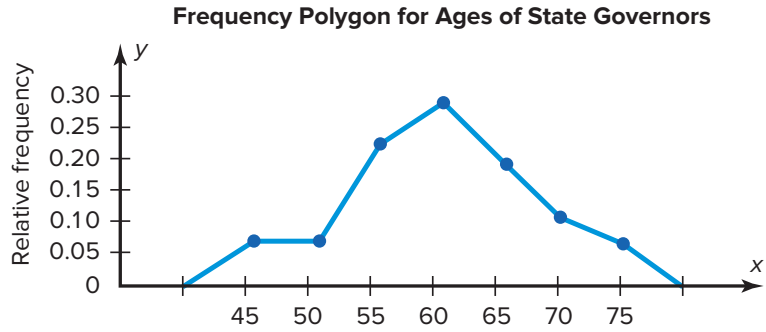
	Cumulative frequency	Cumulative relative frequency
Less than 42.5	0	0.00
Less than 47.5	4	0.08
Less than 52.5	8	0.16
Less than 57.5	19	0.38
Less than 62.5	33	0.66
Less than 67.5	42	0.84
Less than 72.5	47	0.94
Less than 77.5	50	1.00

Step 3 Draw each graph as shown in Figure 2-6. For the histogram and ogive, use the class boundaries along the *x* axis. For the frequency, use the midpoints on the *x* axis. For the scale on the *y* axis, use proportions.

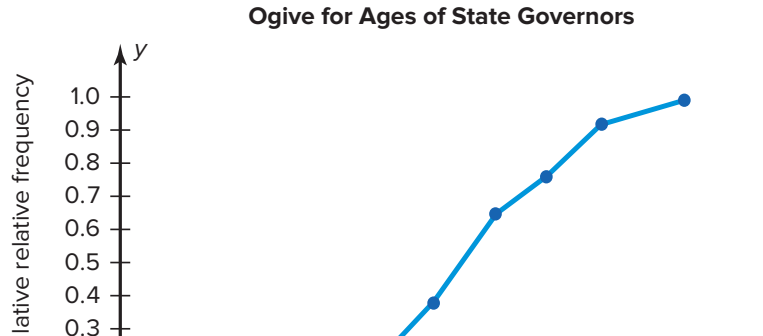
FIGURE 2–6
Graphs for Example 2–7



(a) Histogram



(b) Frequency polygon



(c) Ogive

Distribution Shapes

When one is describing data, it is important to be able to recognize the shapes of the distribution values. In later chapters, you will see that the shape of a distribution also determines the appropriate statistical methods used to analyze the data.

A distribution can have many shapes, and one method of analyzing a distribution is to draw a histogram or frequency polygon for the distribution. Several of the most common shapes are shown in Figure 2–7: *the bell-shaped or mound-shaped, the uniform-shaped, the J-shaped, the reverse J-shaped, the positively or right-skewed shape, the negatively or left-skewed shape, the bimodal-shaped, and the U-shaped.*

Distributions are most often not perfectly shaped, so it is not necessary to have an exact shape but rather to identify an overall pattern.

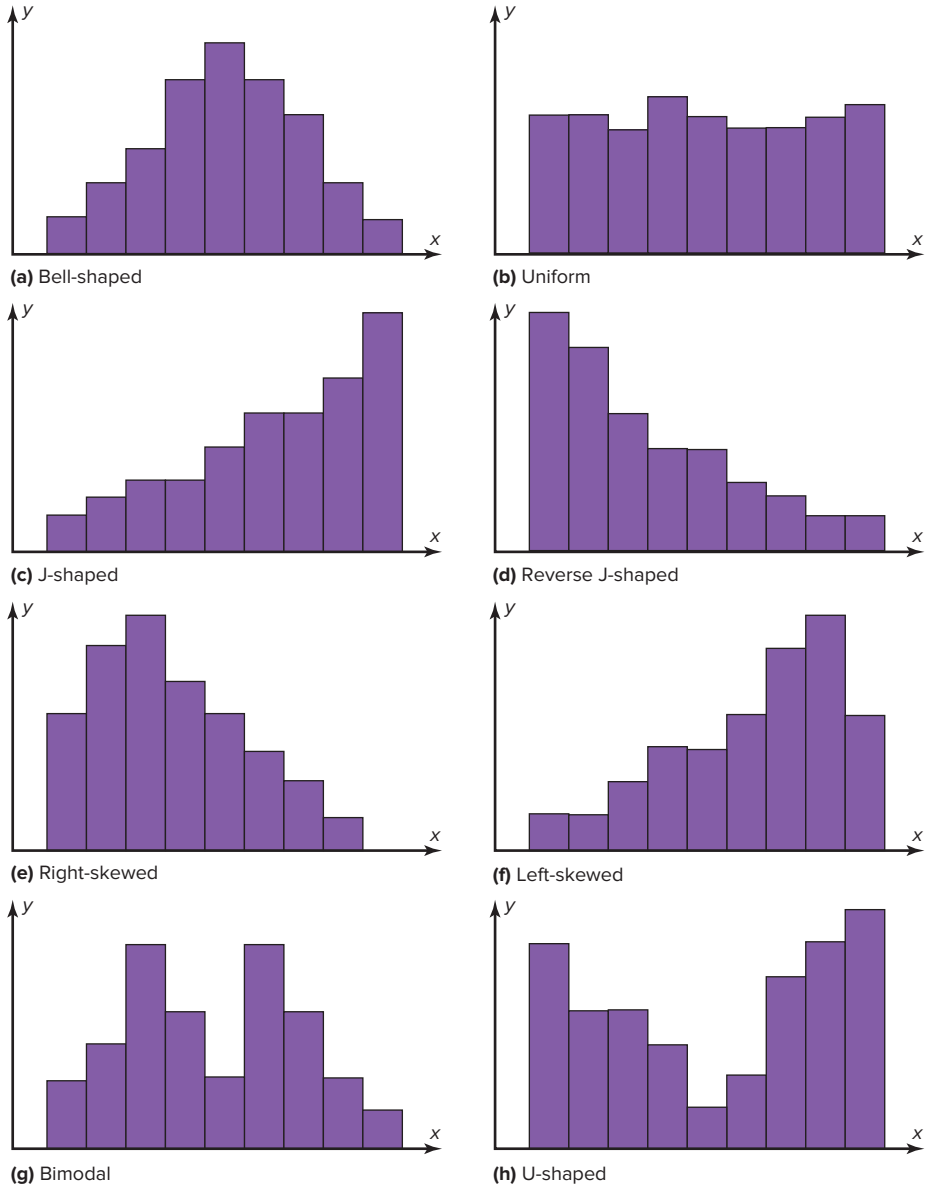
A *bell-shaped distribution* shown in Figure 2–7(a) has a single peak and tapers off at either end. It is approximately symmetric; i.e., it is roughly the same on both sides of a line running through the center.

A *uniform distribution* is basically flat or rectangular. See Figure 2–7(b).

A *J-shaped distribution* is shown in Figure 2–7(c), and it has a few data values on the left side and increases as one moves to the right. A *reverse J-shaped distribution* is the opposite of the J-shaped distribution. See Figure 2–7(d).

When the peak of a distribution is to the left and the data values taper off to the right, a distribution is said to be *positively or right-skewed.* See Figure 2–7(e). When

FIGURE 2–7
Distribution Shapes



the data values are clustered to the right and taper off to the left, a distribution is said to be *negatively or left-skewed.* See Figure 2–7(f). Skewness will be explained in detail in Chapter 3. Distributions with one peak, such as those shown in Figure 2–7(a), (e), and (f), are said to be *unimodal.* (The highest peak of a distribution indicates where the mode of the data values is. The mode is the data value that occurs more often than any other data value. Modes are explained in Chapter 3.) When a distribution has two peaks of the same height, it is said to be *bimodal.* See Figure 2–7(g). Finally, the graph shown in Figure 2–7(h) is a *U-shaped* distribution.

Distributions can have other shapes in addition to the ones shown here; however, these are some of the more common ones that you will encounter in analyzing data.

When you are analyzing histograms and frequency polygons, look at the shape of the curve. For example, does it have one peak or two peaks? Is it relatively flat, or is it U-shaped? Are the data values spread out on the graph, or are they clustered around the center? Are there data values in the extreme ends? These may be *outliers.* (See Section 3–3 for an explanation of outliers.) Are there any gaps in the histogram, or does

the frequency polygon touch the *x* axis somewhere other than at the ends? Finally, are the data clustered at one end or the other, indicating a *skewed distribution*?

For example, the histogram for the record high temperatures in Figure 2–1 shows a single peaked distribution, with the class 109.5–114.5 containing the largest number of temperatures. The distribution has no gaps, and there are fewer temperatures in the highest class than in the lowest class.

Applying the Concepts 2–2

Selling Real Estate

Assume you are a realtor in Bradenton, Florida. You have recently obtained a listing of the selling prices of the homes that have sold in that area in the last 6 months. You wish to organize those data so you will be able to provide potential buyers with useful information. Use the following data to create a histogram, frequency polygon, and cumulative frequency polygon.

142,000	127,000	99,600	162,000	89,000	93,000	99,500
73,800	135,000	119,500	67,900	156,300	104,500	108,650
123,000	91,000	205,000	110,000	156,300	104,000	133,900
179,000	112,000	147,000	321,550	87,900	88,400	180,000
159,400	205,300	144,400	163,000	96,000	81,000	131,000
114,000	119,600	93,000	123,000	187,000	96,000	80,000
231,000	189,500	177,600	83,400	77,000	132,300	166,000

1. What questions could be answered more easily by looking at the histogram rather than the listing of home prices?
2. What different questions could be answered more easily by looking at the frequency polygon rather than the listing of home prices?
3. What different questions could be answered more easily by looking at the cumulative frequency polygon rather than the listing of home prices?
4. Are there any extremely large or extremely small data values compared to the other data values?
5. Which graph displays these extremes the best?
6. Is the distribution skewed?

See page 108 for the answers.

Exercises 2–2

1. Do Students Need Summer Development? For 108 randomly selected college applicants, the following frequency distribution for entrance exam scores was obtained. Construct a histogram, frequency polygon, and ogive for the data. (The data for this exercise will be used for Exercise 13 in this section.)

Class limits	Frequency
90–98	6
99–107	22
108–116	43
117–125	28
126–134	9
	Total 108

Applicants who score above 107 need not enroll in a summer developmental program. In this group, how many students do not have to enroll in the developmental program?

2. Bear Kills The number of bears killed in 2014 for 56 counties in Pennsylvania is shown in the frequency distribution. Construct a histogram, frequency polygon, and ogive for the data. Comment on the skewness of the distribution. How many counties had 75 or fewer bears killed? (The data for this exercise will be used for Exercise 14 of this section.)

Class limits	Frequency
1–25	16
26–50	14
51–75	9
76–100	8
101–125	5
126–150	0
151–175	1
176–200	1
201–225	0
226–250	0
251–275	2
	Total 56

Source: Pennsylvania State Game Commission.

3. Pupils Per Teacher The average number of pupils per teacher in each state is shown. Construct a grouped frequency distribution with 6 classes. Draw a histogram, frequency polygon, and ogive. Analyze the distribution.

16	16	15	12	14
13	16	14	15	14
18	18	18	12	15
15	16	16	15	15
25	19	15	12	22
18	14	13	17	9
13	14	13	16	12
14	16	10	22	20
12	14	18	15	14
16	12	12	13	15

Source: U.S. Department of Education.

4. Number of College Faculty The number of faculty listed for a sample of private colleges that offer only bachelor's degrees is listed below. Use these data to construct a frequency distribution with 7 classes, a histogram, a frequency polygon, and an ogive. Discuss the shape of this distribution. What proportion of schools have 180 or more faculty?

165	221	218	206	138	135	224	204
70	210	207	154	155	82	120	116
176	162	225	214	93	389	77	135
221	161	128	310				

Source: World Almanac and Book of Facts.

5. Railroad Crossing Accidents The data show the number of railroad crossing accidents for the 50 states of the United States for a specific year. Construct a histogram, frequency polygon, and ogive for the data. Comment on the skewness of the distribution. (The data in this exercise will be used for Exercise 15 in this section.)

Class limits	Frequency
1–43	24
44–86	17
87–129	3
130–172	4
173–215	1
216–258	0
259–301	0
302–344	1
	Total 50

Source: Federal Railroad Administration.

6. NFL Salaries The salaries (in millions of dollars) for 31 NFL teams for a specific season are given in this frequency distribution.

Construct a histogram, a frequency polygon, and an ogive for the data; and comment on the shape of the distribution. (The data for this exercise will be used for Exercise 16 of this section.)

Class limits	Frequency
39.9–42.8	2
42.9–45.8	2
45.9–48.8	5
48.9–51.8	5
51.9–54.8	12
54.9–57.8	5
	Total 31

Source: NFL.com

7. Suspension Bridges Spans The following frequency distribution shows the length (in feet) of the main spans of the longest suspension bridges in the United States. Construct a histogram, frequency polygon, and ogive for the distribution. Describe the shape of the distribution.

Class limits	Frequency
1260–1734	12
1735–2209	6
2210–2684	3
2685–3159	1
3160–3634	1
3635–4109	1
4110–4584	2

Source: U.S. Department of Transportation.

8. Costs of Utilities The frequency distribution represents the cost (in cents) for the utilities of states that supply much of their own power. Construct a histogram, frequency polygon, and ogive for the data. Is the distribution skewed?

Class limits	Frequency
6–8	12
9–11	16
12–14	3
15–17	1
18–20	0
21–23	0
24–26	1
	Total 33

9. Air Pollution One of the air pollutants that is measured in selected cities is sulfur dioxide. This pollutant occurs when fossil fuels are burned. This pollutant is measured in micrograms per cubic meter ($\mu g/m^3$). The results obtained from a sample of 24 cities are shown in the frequency distributions. One sample was taken recently, and the other sample of the same cities was taken

5 years ago. Construct a histogram and compare the two distributions.

Class limits	Frequency (now)	Frequency (5 years ago)
10–14	6	5
15–19	4	4
20–24	3	2
25–29	2	3
30–34	5	6
35–39	1	2
40–44	2	1
45–49	1	1
	Total 24	Total 24

10. Making the Grade The frequency distributions shown indicate the percentages of public school students in fourth-grade reading and mathematics who performed at or above the required proficiency levels for the 50 states in the United States. Draw histograms for each, and decide if there is any difference in the performance of the students in the subjects.

Class	Reading frequency	Math frequency
17.5–22.5	7	5
22.5–27.5	6	9
27.5–32.5	14	11
32.5–37.5	19	16
37.5–42.5	3	8
42.5–47.5	1	1
	Total 50	Total 50

Source: National Center for Educational Statistics.

11. Blood Glucose Levels The frequency distribution shows the blood glucose levels (in milligrams per deciliter) for 50 patients at a medical facility. Construct a histogram, frequency polygon, and ogive for the data. Comment on the shape of the distribution. What range of glucose levels did most patients fall into?

Class limits	Frequency
60–64	2
65–69	1
70–74	5
75–79	12
80–84	18
85–89	6
90–94	5
95–99	1
	Total 50

12. Waiting Times The frequency distribution shows the waiting times (in minutes) for 50 patients at a walk-in medical facility. Construct a histogram, frequency polygon, and ogive for the data. Is the distribution skewed? How many patients waited longer than 30 minutes?

Class limits	Frequency
11–15	7
16–20	9
21–25	15
26–30	9
31–35	5
36–40	3
41–45	2
	Total 50

13. Construct a histogram, frequency polygon, and ogive, using relative frequencies for the data in Exercise 1 of this section.

14. Construct a histogram, frequency polygon, and ogive, using relative frequencies for the data in Exercise 2 of this section.

15. Construct a histogram, frequency polygon, and ogive, using relative frequencies for the data in Exercise 5 of this section.

16. Construct a histogram, frequency polygon, and ogive, using relative frequencies for the data in Exercise 6 of this section.

17. Home Runs The data show the most number of home runs hit by a batter in the American League over the last 30 seasons. Construct a frequency distribution using 5 classes. Draw a histogram, a frequency polygon, and an ogive for the date, using relative frequencies. Describe the shape of the histogram.

40	43	40
53	47	46
44	57	43
43	52	44
54	47	51
39	48	36
37	56	42
54	56	49
54	52	40
48	50	40

Source: World Almanac and Book of Facts.

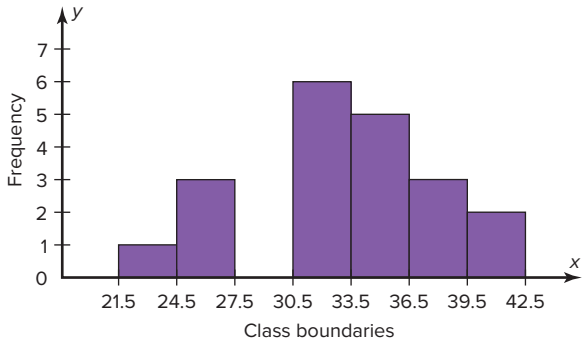
18. Protein Grams in Fast Food The amount of protein (in grams) for a variety of fast-food sandwiches is reported here. Construct a frequency distribution, using 6 classes. Draw a histogram, a frequency polygon, and an ogive for the data, using relative frequencies. Describe the shape of the histogram.

23	30	20	27	44	26	35	20	29	29
25	15	18	27	19	22	12	26	34	15
27	35	26	43	35	14	24	12	23	31
40	35	38	57	22	42	24	21	27	33

Source: The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.

Extending the Concepts

19. Using the histogram shown here, do the following.



a. Construct a frequency distribution; include class limits, class frequencies, midpoints, and cumulative frequencies.

b. Construct a frequency polygon.

c. Construct an ogive.

20. Using the results from Exercise 19, answer these questions.

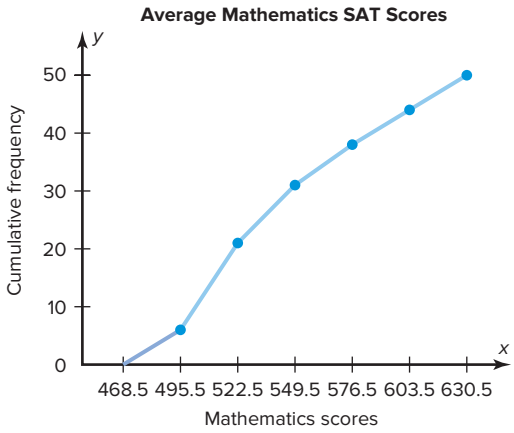
a. How many values are in the class 27.5–30.5?

b. How many values fall between 24.5 and 36.5?

c. How many values are below 33.5?

d. How many values are above 30.5?

21. Math SAT Scores Shown is an ogive depicting the cumulative frequency of the average mathematics SAT scores by state. Use it to construct a histogram and a frequency polygon.



Technology

TI-84 Plus

Step by Step

Step by Step

Constructing a Histogram

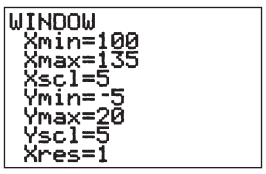
To display the graphs on the screen, enter the appropriate values in the calculator, using the WINDOW menu. The default values are $X_{min} = -10$, $X_{max} = 10$, $Y_{min} = -10$, and $Y_{max} = 10$.

The X_{scl} changes the distance between the tick marks on the x axis and can be used to change the class width for the histogram.

To change the values in the WINDOW:

1. Press **WINDOW.**
2. Move the cursor to the value that needs to be changed. Then type in the desired value and press **ENTER.**
3. Continue until all values are appropriate.
4. Press **[2nd] [QUIT]** to leave the WINDOW menu.

Input

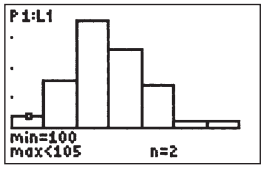


Input

To plot the histogram from raw data:

1. Enter the data in L_1.
2. Make sure WINDOW values are appropriate for the histogram.
3. Press **[2nd] [STAT PLOT] ENTER.**
4. Press **ENTER** to turn the plot 1 on, if necessary.
5. Move cursor to the Histogram symbol and press **ENTER,** if necessary. The histogram is the third option.
6. Make sure Xlist is L_1.
7. Make sure Freq is 1.
8. Press **GRAPH** to display the histogram.
9. To obtain the frequency (number of data values in each class), press the **TRACE** key, followed by ◄ or ► keys.

Output



Example TI2–1

Plot a histogram for the following data from Example 2–2.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Press **TRACE** and use the arrow keys to determine the number of values in each group.

To graph a histogram from grouped data:

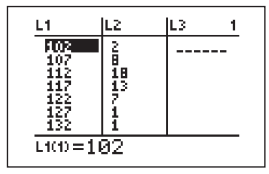
1. Enter the midpoints into L_1.
2. Enter the frequencies into L_2.
3. Make sure WINDOW values are appropriate for the histogram.
4. Press **[2nd] [STAT PLOT] ENTER.**
5. Press **ENTER** to turn the plot on, if necessary.
6. Move cursor to the histogram symbol, and press **ENTER,** if necessary.
7. Make sure Xlist is L_1.
8. Make sure Freq is L_2.
9. Press **GRAPH** to display the histogram.

Example TI2–2

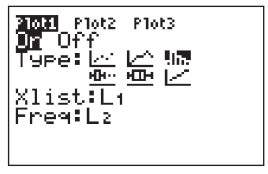
Plot a histogram for the data from Examples 2–4 and 2–5.

Class boundaries	Midpoints	Frequency
99.5–104.5	102	2
104.5–109.5	107	8
109.5–114.5	112	18
114.5–119.5	117	13
119.5–124.5	122	7
124.5–129.5	127	1
129.5–134.5	132	1

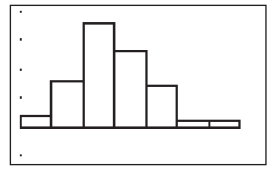
Input



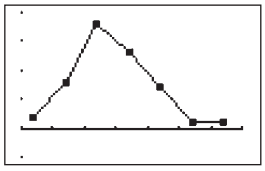
Input



Output

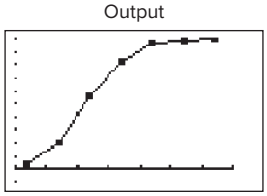


Output



To graph a frequency polygon from grouped data, follow the same steps as for the histogram except change the graph type from histogram (third graph) to a line graph (second graph).

Output



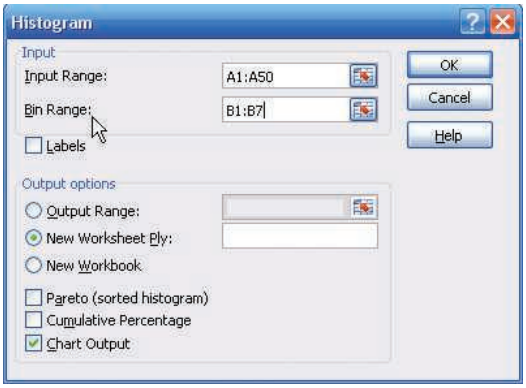
To graph an ogive from grouped data, modify the procedure for the histogram as follows:

1. Enter the upper class boundaries into L_1.
2. Enter the cumulative frequencies into L_2.
3. Change the graph type from histogram (third graph) to line (second graph).
4. Change the Y_{max} from the WINDOW menu to the sample size.

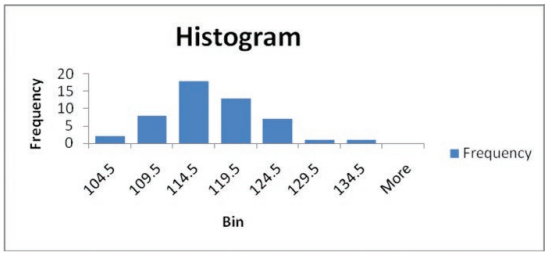
EXCEL
Step by Step

Constructing a Histogram

1. Press **[Ctrl]-N** for a new workbook.
2. Enter the data from Example 2–2 in column A, one number per cell.
3. Enter the upper boundaries into column B.
4. From the toolbar, select the Data tab, then select Data Analysis.
5. In Data Analysis, select Histogram and click [OK].
6. In the Histogram dialog box, type **A1:A50** in the Input Range box and type **B1:B7** in the Bin Range box.



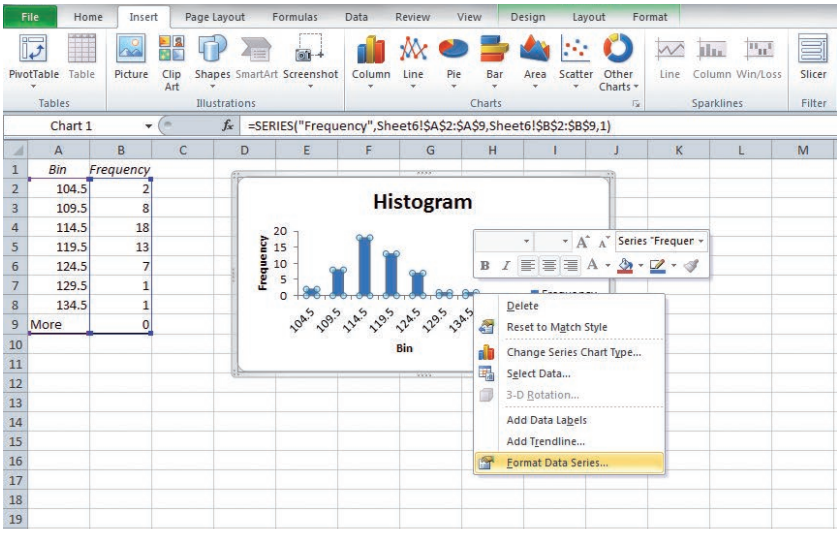
7. Select New Worksheet Ply and Chart Output. Click [OK].



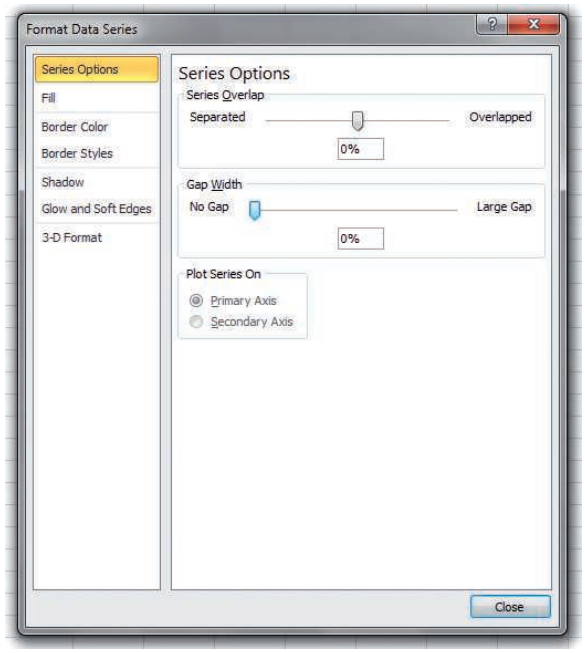
Editing the Histogram

To move the vertical bars of the histogram closer together:

1. Right-click one of the bars of the histogram, and select Format Data Series.

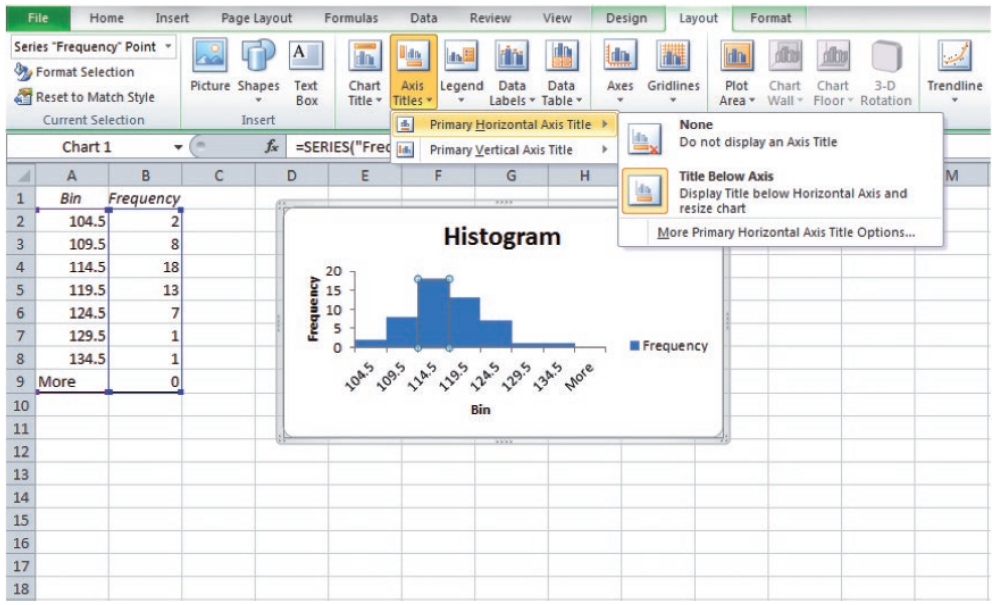


2. Move the Gap Width slider all the way to the left to change the gap width of the bars in the histogram to 0.



To change the label for the horizontal axis:

1. Left-click the mouse over any part of the histogram.
2. Select the Chart Tools tab from the toolbar.
3. Select the Layout tab, Axis Titles and Primary Horizontal Axis Title.



Once the Axis Titles text box is selected, you can type in the name of the variable represented on the horizontal axis.

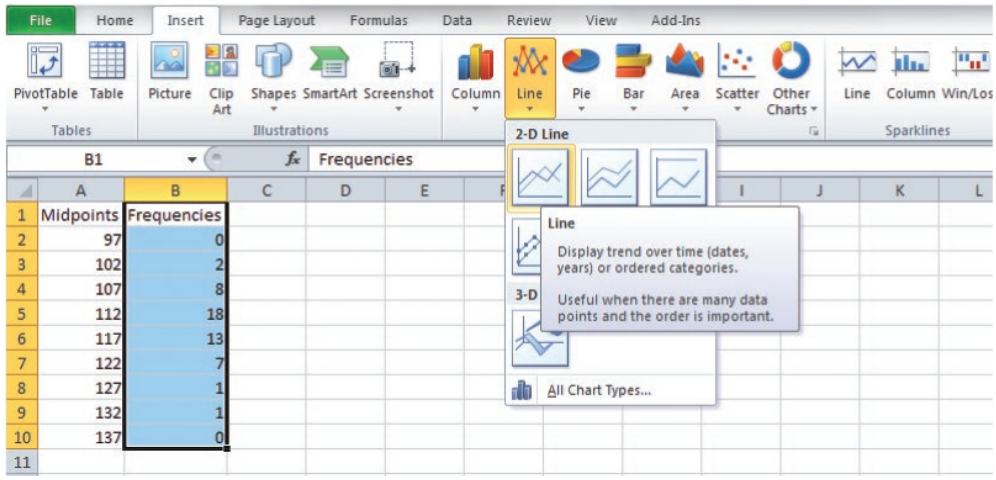
Constructing a Frequency Polygon

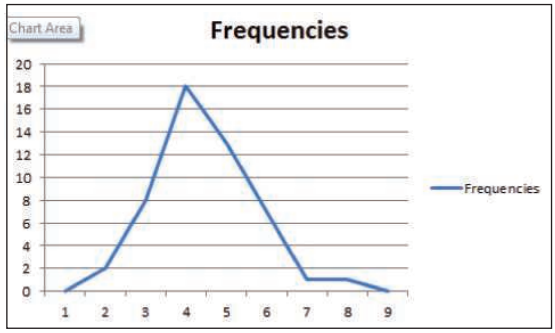
1. Press **[CTRL]-N** for a new notebook.
2. Enter the midpoints of the data from Example 2–2 into column A and the frequencies into column B, including labels.

Note: Classes with frequency 0 have been added at the beginning and the end to "anchor" the frequency polygon to the horizontal axis.

	A	B
1	Midpoints	Frequencies
2	97	0
3	102	2
4	107	8
5	112	18
6	117	13
7	122	7
8	127	1
9	132	1
10	137	0
11		

3. Press and hold the left mouse button, and drag over the Frequencies (including the label) from column B.
4. Select the Insert tab from the toolbar and the Line Chart option.
5. Select the 2-D line chart type.





We will need to edit the graph so that the midpoints are on the horizontal axis.

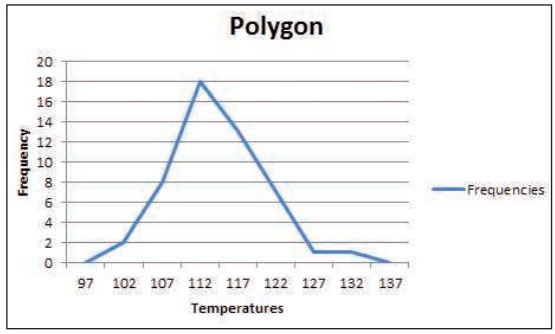
1. Right click the mouse on any region of the chart.
2. Choose Select Data.
3. Select Edit below the Horizontal (Category) Axis Labels panel on the right.
4. Press and hold the left mouse button, and drag over the midpoints (not including the label) for the Axis label range, then click [OK].
5. Click [OK] on the Select Data Source box.

Inserting Labels on the Axes

1. Click the mouse on any region of the graph.
2. Select Chart Tools and then Layout on the toolbar.
3. Select Axis Titles to open the horizontal and vertical axis text boxes. Then manually type in labels for the axes.

Changing the Title

1. Select Chart Tools, Layout from the toolbar.
2. Select Chart Title.
3. Choose one of the options from the Chart Title menu and edit.

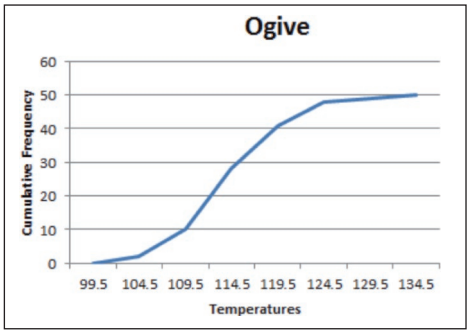


Constructing an Ogive

To create an ogive, use the upper class boundaries (horizontal axis) and cumulative frequencies (vertical axis) from the frequency distribution.

1. Type the upper class boundaries (including a class with frequency 0 before the lowest class to anchor the graph to the horizontal axis) and corresponding cumulative frequencies into adjacent columns of an Excel worksheet.
2. Press and hold the left mouse button, and drag over the Cumulative Frequencies from column B.
3. Select Line Chart, then the 2-D Line option.

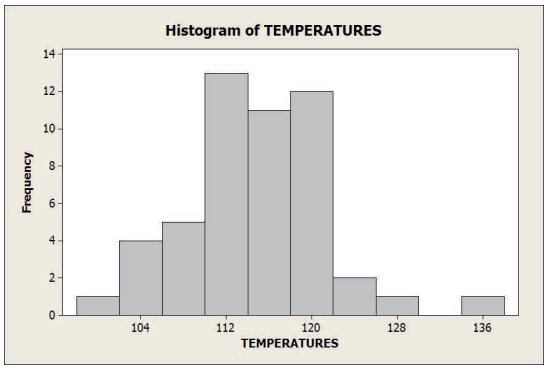
As with the frequency polygon, you can insert labels on the axes and a chart title for the ogive.



MINITAB Step by Step

Construct a Histogram

1. Enter the data from Example 2–2, the high temperatures for the 50 states, into C1.
2. Select **Graph>Histogram**.
3. Select [Simple], then click [OK].
4. Click C1 TEMPERATURES in the Graph variables dialog box.
5. Click [OK]. A new graph window containing the histogram will open.
6. Click the **File** menu to print or save the graph.

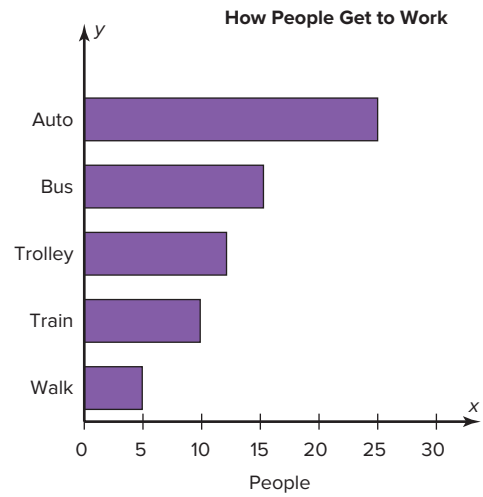


7. Click **File>Exit**.
8. Save the project as **Ch2-3.mpj.**

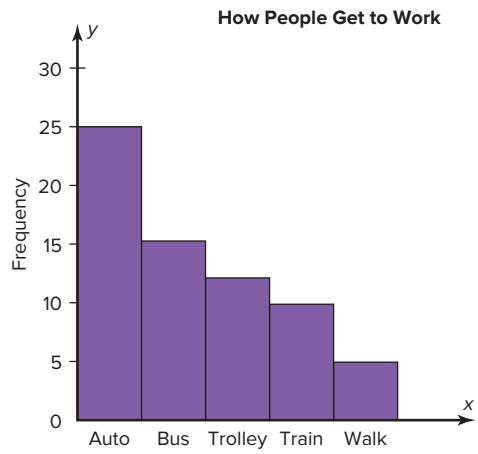
2–3 Other Types of Graphs

In addition to the histogram, the frequency polygon, and the ogive, several other types of graphs are often used in statistics. They are the bar graph, Pareto chart, time series graph, pie graph, and the dotplot. Figure 2–8 shows an example of each type of graph.

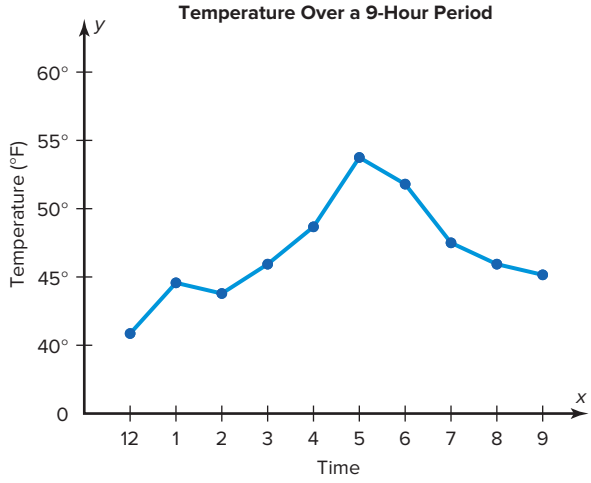
FIGURE 2–8 Other Types of Graphs Used in Statistics



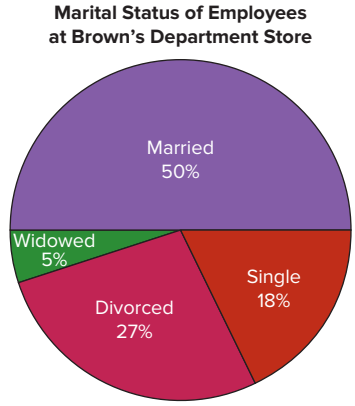
(a) Bar graph



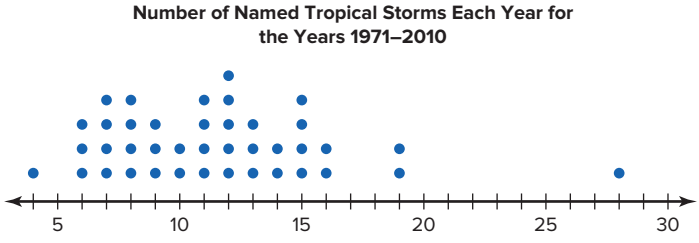
(b) Pareto chart



(c) Time series graph



(d) Pie graph



(e) Dotplot

OBJECTIVE 3

Represent data using bar graphs, Pareto charts, time series graphs, pie graphs, and dotplots.

Bar Graphs

When the data are qualitative or categorical, bar graphs can be used to represent the data. A bar graph can be drawn using either horizontal or vertical bars.

A **bar graph** represents the data by using vertical or horizontal bars whose heights or lengths represent the frequencies of the data.

EXAMPLE 2–8 College Spending for First-Year Students

The table shows the average money spent by first-year college students. Draw a horizontal and vertical bar graph for the data.

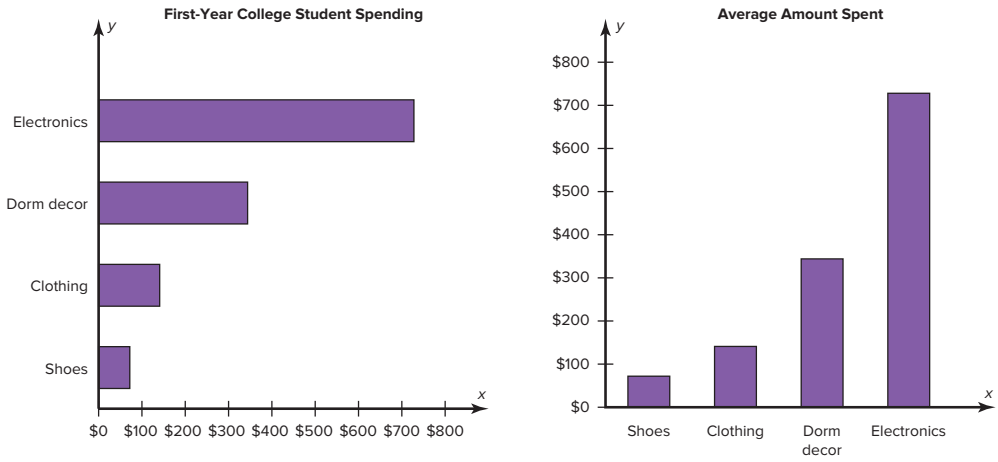
Electronics	$728
Dorm decor	344
Clothing	141
Shoes	72

Source: The National Retail Federation.

SOLUTION

1. Draw and label the x and y axes. For the horizontal bar graph place the frequency scale on the x axis, and for the vertical bar graph place the frequency scale on the y axis.
2. Draw the bars corresponding to the frequencies. See Figure 2–9.

FIGURE 2–9 Bar Graphs for Example 2–8



The graphs show that first-year college students spend the most on electronic equipment.

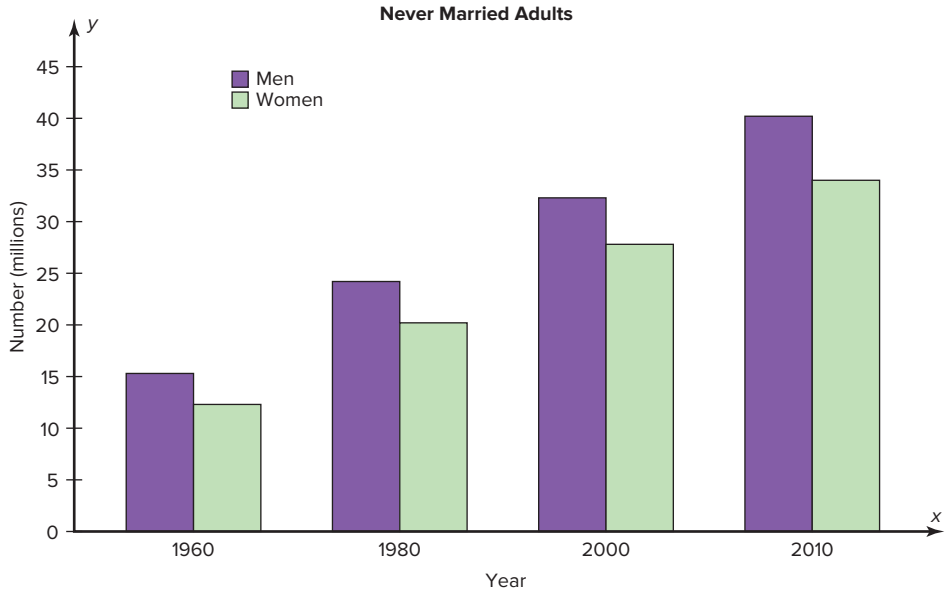
Bar graphs can also be used to compare data for two or more groups. These types of bar graphs are called *compound bar graphs.* Consider the following data for the number (in millions) of never married adults in the United States.

Year	Males	Females
1960	15.3	12.3
1980	24.2	20.2
2000	32.3	27.8
2010	40.2	34.0

Source: U.S. Census Bureau.

Figure 2–10 shows a bar graph that compares the number of never married males with the number of never married females for the years shown. The comparison is made by placing the bars next to each other for the specific years. The heights of the bars can be compared. This graph shows that there have consistently been more never married

FIGURE 2–10
Example of a Compound Bar Graph



males than never married females and that the difference in the two groups has increased slightly over the last 50 years.

Pareto Charts

When the variable displayed on the horizontal axis is qualitative or categorical, a *Pareto chart* can also be used to represent the data.

A **Pareto chart** is used to represent a frequency distribution for a categorical variable, and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

EXAMPLE 2–9 Traffic Congestion

The data shown consist of the average number of hours that a commuter spends in traffic congestion per year in each city. Draw and analyze a Pareto chart for the data.

City	Hours
Atlanta	52
Boston	64
Chicago	61
New York	74
Washington, D. C.	82

Source: 2015 Urban Mobility Scorecard

SOLUTION

Step 1 Arrange the data from the largest to the smallest according to the number of hours.

City	Hours
Washington, D.C.	82
New York	74
Boston	64
Chicago	61
Atlanta	52

Historical Note

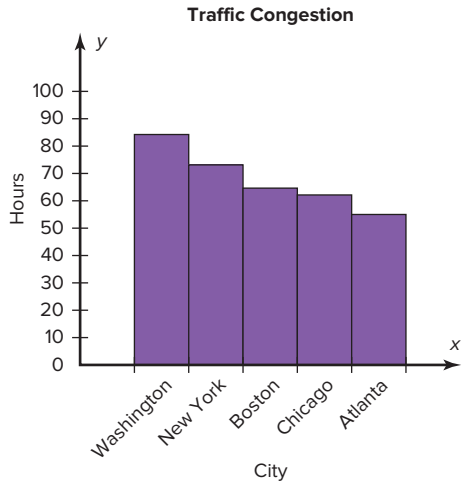
Vilfredo Pareto (1848–1923) was an Italian scholar who developed theories in economics, statistics, and the social sciences. His contributions to statistics include the development of a mathematical function used in economics. This function has many statistical applications and is called the Pareto distribution. In addition, he researched income distribution, and his findings became known as Pareto's law.

Step 2 Draw and label the *x* and *y* axes.

Step 3 Draw the vertical bars according to the number of hours (large to small).

The graph shown in Figure 2–11 shows that Washington, D.C. has the longest congestion times and Atlanta has the shortest times for the selection of cities.

FIGURE 2–11 Pareto Chart for Example 2–9



Suggestions for Drawing Pareto Charts

1. Make the bars the same width.
2. Arrange the data from largest to smallest according to frequency.
3. Make the units that are used for the frequency equal in size.

When you analyze a Pareto chart, make comparisons by looking at the heights of the bars.

The Time Series Graph

When data are collected over a period of time, they can be represented by a time series graph.

A **time series graph** represents data that occur over a specific period of time.

Example 2–10 shows the procedure for constructing a time series graph.

EXAMPLE 2–10 Price of an Advertisement for the Academy Awards Show

The data show the average cost (in millions of dollars) of a 30-second television ad on the Academy Awards show. Draw and analyze a time series graph for the data.

Year	2010	2011	2012	2013	2014	2015
Cost	1.40	1.55	1.61	1.65	1.78	1.90

Source: Kantar Media, *USA TODAY* RESEARCH

Historical Note

Time series graphs are over 1000 years old. The first ones were used to chart the movements of the planets and the sun.

SOLUTION

Step 1 Draw and label the x and y axes.

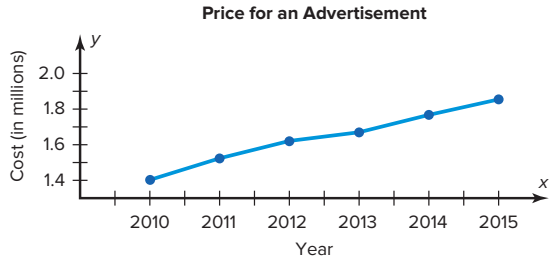
Step 2 Label the x axis for years and label the y axis for cost.

Step 3 Plot each point for the values shown in the table.

Step 4 Draw line segments connecting adjacent points. Do not try to fit a smooth curve through the data points. See Figure 2–12.

The data show that there has been an increase every year. The largest increase (shown by the steepest line segment) occurred for the year 2011 compared to 2010. The increases for the years 2011, 2012, and 2013 were relatively small compared to the increases from 2010 to 2014 and 2014 to 2015.

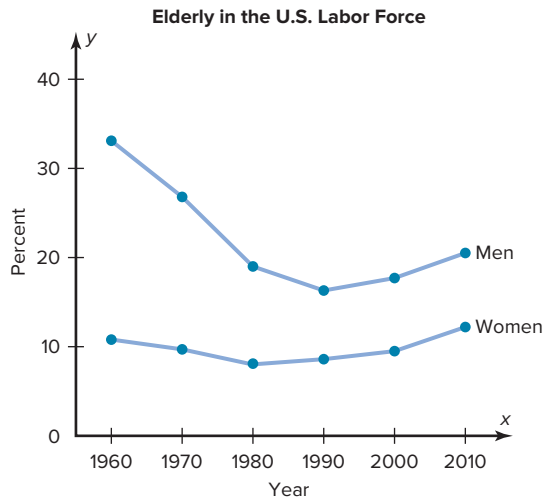
FIGURE 2–12 Figure for Example 2–10



When you analyze a time series graph, look for a trend or pattern that occurs over the time period. For example, is the line ascending (indicating an increase over time) or descending (indicating a decrease over time)? Another thing to look for is the slope, or steepness, of the line. A line that is steep over a specific time period indicates a rapid increase or decrease over that period.

Two or more data sets can be compared on the same graph called a *compound time series graph* if two or more lines are used, as shown in Figure 2–13. This graph shows

FIGURE 2–13
Two Time Series Graphs for Comparison



Source: Bureau of Census, U.S. Department of Commerce.

the percentage of elderly males and females in the U.S. labor force from 1960 to 2010. It shows that the percentage of elderly men decreased significantly from 1960 to 1990 and then increased slightly after that. For the elderly females, the percentage decreased slightly from 1960 to 1980 and then increased from 1980 to 2010.

The Pie Graph

Pie graphs are used extensively in statistics. The purpose of the pie graph is to show the relationship of the parts to the whole by visually comparing the sizes of the sections. Percentages or proportions can be used. The variable is nominal or categorical.

A **pie graph** is a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution.

Example 2–11 shows the procedure for constructing a pie graph.

EXAMPLE 2–11 Super Bowl Snack Foods

This frequency distribution shows the number of pounds of each snack food eaten during the Super Bowl. Construct a pie graph for the data.

Snack	Pounds (frequency)
Potato chips	11.2 million
Tortilla chips	8.2 million
Pretzels	4.3 million
Popcorn	3.8 million
Snack nuts	2.5 million
	Total $n = 30.0$ million

Source: USA TODAY Weekend.

SOLUTION

Step 1 Since there are 360° in a circle, the frequency for each class must be converted to a proportional part of the circle. This conversion is done by using the formula

$$\text{Degrees} = \frac{f}{n} \cdot 360°$$

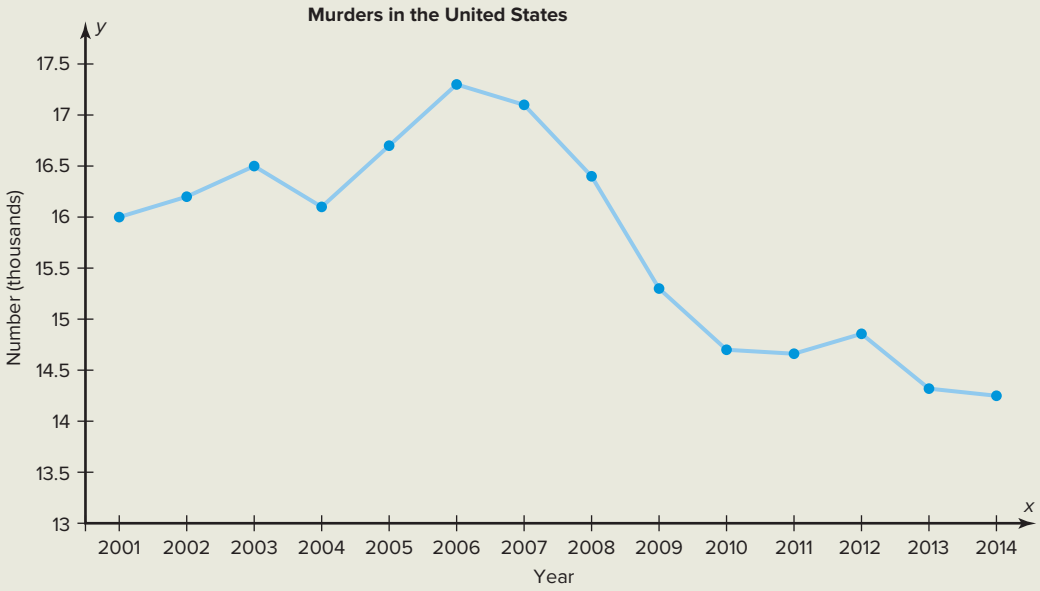
where f = frequency for each class and n = sum of the frequencies. Hence, the following conversions are obtained. The degrees should sum to 360°.[1]

Potato chips	$\frac{11.2}{30} \cdot 360° = 134°$
Tortilla chips	$\frac{8.2}{30} \cdot 360° = 98°$
Pretzels	$\frac{4.3}{30} \cdot 360° = 52°$
Popcorn	$\frac{3.8}{30} \cdot 360° = 46°$
Snack nuts	$\frac{2.5}{30} \cdot 360° = 30°$
Total	360°

[1]*Note:* The degrees column does not always sum to 360° due to rounding.

SPEAKING OF STATISTICS Murders in the United States

The graph shows the number of murders (in thousands) that have occurred in the United States since 2001. Based on the graph, do you think the number of murders is increasing, decreasing, or remaining the same?



Source: Crime in the United States 2015, FBI, Department of Justice.

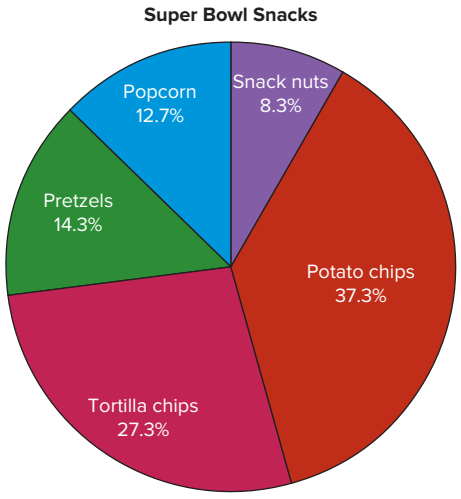
Step 2 Each frequency must also be converted to a percentage. Recall from Example 2–1 that this conversion is done by using the formula

$$\% = \frac{f}{n} \cdot 100$$

Hence, the following percentages are obtained. The percentages should sum to 100%.[2]

Potato chips	$\frac{11.2}{30} \cdot 100 = 37.3\%$	
Tortilla chips	$\frac{8.2}{30} \cdot 100 = 27.3\%$	
Pretzels	$\frac{4.3}{30} \cdot 100 = 14.3\%$	
Popcorn	$\frac{3.8}{30} \cdot 100 = 12.7\%$	
Snack nuts	$\frac{2.5}{30} \cdot 100 = 8.3\%$	
Total		99.9%

FIGURE 2–14 Pie Graph for Example 2–11



Step 3 Next, using a protractor and a compass, draw the graph, using the appropriate degree measures found in Step 1, and label each section with the name and percentages, as shown in Figure 2–14.

[2]*Note:* The percent column does not always sum to 100% due to rounding.

EXAMPLE 2–12 Police Calls

Construct and analyze a pie graph for the calls received each shift by a local municipality for a recent year. (Data obtained by author.)

Shift	Frequency
1. Day	2594
2. Evening	2800
3. Night	2436
	7830

SOLUTION

Step 1 Find the number of degrees for each shift, using the formula:

$$\text{Degrees} = \frac{f}{n} \cdot 360°$$

For each shift, the following results are obtained:

Day: $\frac{2594}{7830} \cdot 360° = 119°$

Evening: $\frac{2800}{7830} \cdot 360° = 129°$

Night: $\frac{2436}{7830} \cdot 360° = 112°$

Step 2 Find the percentages:

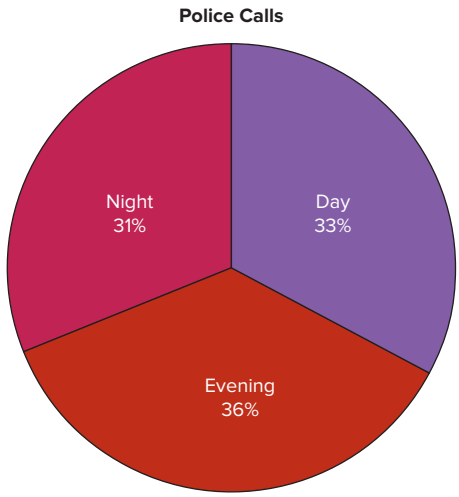
Day: $\frac{2594}{7830} \cdot 100 = 33\%$

Evening: $\frac{2800}{7830} \cdot 100 = 36\%$

Night: $\frac{2436}{7830} \cdot 100 = 31\%$

Step 3 Using a protractor, graph each section and write its name and corresponding percentage as shown in Figure 2–15.

FIGURE 2–15
Figure for Example 2–12



To analyze the nature of the data shown in the pie graph, look at the size of the sections in the pie graph. For example, are any sections relatively large compared to the rest? Figure 2–15 shows that the number of calls for the three shifts are about equal, although slightly more calls were received on the evening shift.

Note: Computer programs can construct pie graphs easily, so the mathematics shown here would only be used if those programs were not available.

Dotplots

A dotplot uses points or dots to represent the data values. If the data values occur more than once, the corresponding points are plotted above one another.

A **dotplot** is a statistical graph in which each data value is plotted as a point (dot) above the horizontal axis.

Dotplots are used to show how the data values are distributed and to see if there are any extremely high or low data values.

EXAMPLE 2–13 Named Storms

The data show the number of named storms each year for the last 40 years. Construct and analyze a dotplot for the data.

19	15	14	7	6	11	11
9	16	8	8	11	9	8
16	12	13	14	13	12	7
15	15	19	11	4	6	13
10	15	7	12	6	10	
28	12	8	7	12	9	

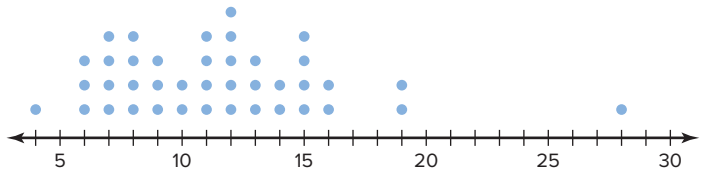
Source: NOAA.

Step 1 Find the lowest and highest data values, and decide what scale to use on the horizontal axis. The lowest data value is 4 and the highest data value is 28, so a scale from 4 to 28 is needed.

Step 2 Draw a horizontal line, and draw the scale on the line.

Step 3 Plot each data value above the line. If the value occurs more than once, plot the other point above the first point. See Figure 2–16.

FIGURE 2–16 Figure for Example 2–13



The graph shows that the majority of the named storms occur with frequency between 6 and 16 per year. There are only 3 years when there were 19 or more named storms per year.

Stem and Leaf Plots

The stem and leaf plot is a method of organizing data and is a combination of sorting and graphing. It has the advantage over a grouped frequency distribution of retaining the actual data while showing them in graphical form.

OBJECTIVE 4

Draw and interpret a stem and leaf plot.

A **stem and leaf plot** is a data plot that uses part of the data value as the stem and part of the data value as the leaf to form groups or classes.

For example, a data value of 34 would have 3 as the stem and 4 as the leaf. A data value of 356 would have 35 as the stem and 6 as the leaf.

Example 2–14 shows the procedure for constructing a stem and leaf plot.

EXAMPLE 2–14 Out Patient Cardiograms

At an outpatient testing center, the number of cardiograms performed each day for 20 days is shown. Construct a stem and leaf plot for the data.

25	31	20	32	13
14	43	02	57	23
36	32	33	32	44
32	52	44	51	45

SOLUTION

Step 1 Arrange the data in order:

02, 13, 14, 20, 23, 25, 31, 32, 32, 32,
32, 33, 36, 43, 44, 44, 45, 51, 52, 57

Note: Arranging the data in order is not essential and can be cumbersome when the data set is large; however, it is helpful in constructing a stem and leaf plot. The leaves in the final stem and leaf plot should be arranged in order.

Step 2 Separate the data according to the first digit, as shown.

02 13, 14 20, 23, 25 31, 32, 32, 32, 32, 33, 36
43, 44, 44, 45 51, 52, 57

Step 3 A display can be made by using the leading digit as the *stem* and the trailing digit as the *leaf*. For example, for the value 32, the leading digit, 3, is the stem and the trailing digit, 2, is the leaf. For the value 14, the 1 is the stem and the 4 is the leaf. Now a plot can be constructed as shown in Figure 2–17.

FIGURE 2–17
Stem and Leaf Plot for Example 2–14

Leading digit (stem)	Trailing digit (leaf)
0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7

Figure 2–17 shows that the distribution peaks in the center and that there are no gaps in the data. For 7 of the 20 days, the number of patients receiving cardiograms was between 31 and 36. The plot also shows that the testing center treated from a minimum of 2 patients to a maximum of 57 patients in any one day.

If there are no data values in a class, you should write the stem number and leave the leaf row blank. Do not put a zero in the leaf row.

SPEAKING OF STATISTICS

How Much Paper Money Is in Circulation Today?

The Federal Reserve estimated that during a recent year, there were 22 billion bills in circulation. About 35% of them were $1 bills, 3% were $2 bills, 8% were $5 bills, 7% were $10 bills, 23% were $20 bills, 5% were $50 bills, and 19% were $100 bills. It costs about 3¢ to print each bill.

The average life of a $1 bill is 22 months, a $10 bill 3 years, a $20 bill 4 years, a $50 bill 9 years, and a $100 bill 9 years. What type of graph would you use to represent the average lifetimes of the bills?

© Art Vandalay/Getty Images RF

EXAMPLE 2–15 Number of Car Thefts in a Large City

An insurance company researcher conducted a survey on the number of car thefts in a large city for a period of 30 days last summer. The raw data are shown. Construct a stem and leaf plot by using classes 50–54, 55–59, 60–64, 65–69, 70–74, and 75–79.

52	62	51	50	69
58	77	66	53	57
75	56	55	67	73
79	59	68	65	72
57	51	63	69	75
65	53	78	66	55

SOLUTION

Step 1 Arrange the data in order.

50, 51, 51, 52, 53, 53, 55, 55, 56, 57, 57, 58, 59, 62, 63, 65, 65, 66, 66, 67, 68, 69, 69, 72, 73, 75, 75, 77, 78, 79

Step 2 Separate the data according to the classes.

50, 51, 51, 52, 53, 53 55, 55, 56, 57, 57, 58, 59
62, 63 65, 65, 66, 66, 67, 68, 69, 69 72, 73
75, 75, 77, 78, 79

Step 3 Plot the data as shown here.

Leading digit (stem)	Trailing digit (leaf)
5	0 1 1 2 3 3
5	5 5 6 7 7 8 9
6	2 3
6	5 5 6 6 7 8 9 9
7	2 3
7	5 5 7 8 9

The graph for this plot is shown in Figure 2–18.

FIGURE 2–18
Stem and Leaf Plot for Example 2–15

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

When you analyze a stem and leaf plot, look for peaks and gaps in the distribution. See if the distribution is symmetric or skewed. Check the variability of the data by looking at the spread.

Related distributions can be compared by using a back-to-back stem and leaf plot. The back-to-back stem and leaf plot uses the same digits for the stems of both distributions, but the digits that are used for the leaves are arranged in order out from the stems on both sides. Example 2–16 shows a back-to-back stem and leaf plot.

EXAMPLE 2–16 Number of Stories in Tall Buildings

The number of stories in two selected samples of tall buildings in Atlanta and Philadelphia is shown. Construct a back-to-back stem and leaf plot, and compare the distributions.

Atlanta					Philadelphia				
55	70	44	36	40	61	40	38	32	30
63	40	44	34	38	58	40	40	25	30
60	47	52	32	32	54	40	36	30	30
50	53	32	28	31	53	39	36	34	33
52	32	34	32	50	50	38	36	39	32
26	29								

Source: *The World Almanac and Book of Facts.*

SOLUTION

Step 1 Arrange the data for both data sets in order.

Step 2 Construct a stem and leaf plot, using the same digits as stems. Place the digits for the leaves for Atlanta on the left side of the stem and the digits for the leaves for Philadelphia on the right side, as shown. See Figure 2–19.

FIGURE 2–19 Back-to-Back Stem and Leaf Plot for Example 2–16

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

Step 3 Compare the distributions. The buildings in Atlanta have a large variation in the number of stories per building. Although both distributions are peaked in the 30- to 39-story class, Philadelphia has more buildings in this class. Atlanta has more buildings that have 40 or more stories than Philadelphia does.

Stem and leaf plots are part of the techniques called *exploratory data analysis.* More information on this topic is presented in Chapter 3.

Misleading Graphs

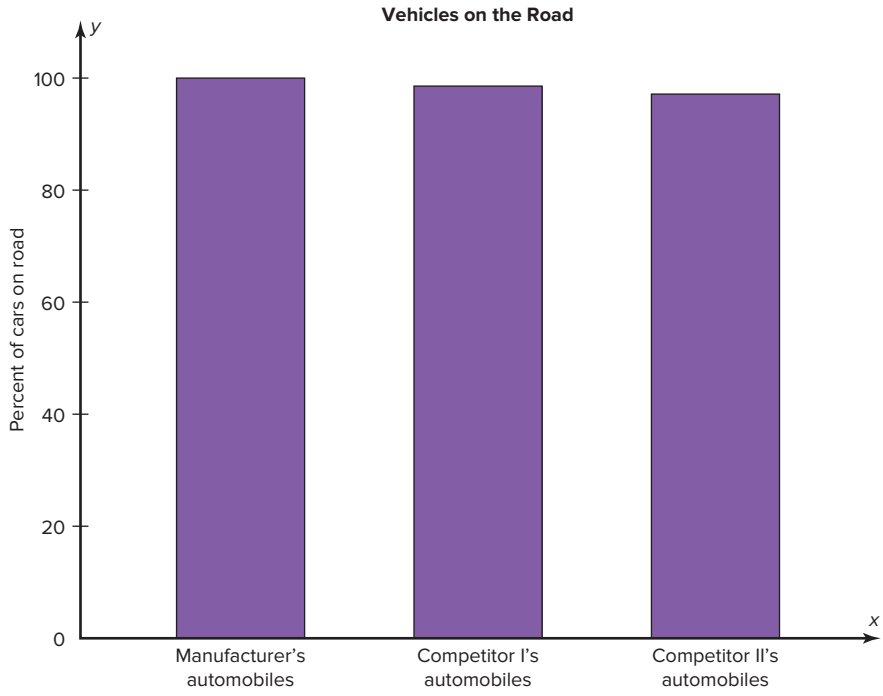
Graphs give a visual representation that enables readers to analyze and interpret data more easily than they could simply by looking at numbers. However, inappropriately drawn graphs can misrepresent the data and lead the reader to false conclusions. For example, a car manufacturer's ad stated that 98% of the vehicles it had sold in the past 10 years were still on the road. The ad then showed a graph similar to the one in Figure 2–20. The graph shows the percentage of the manufacturer's automobiles still on the road and the

percentage of its competitors' automobiles still on the road. Is there a large difference? Not necessarily.

Notice the scale on the vertical axis in Figure 2–20. It has been cut off (or truncated) and starts at 95%. When the graph is redrawn using a scale that goes from 0 to 100%, as in Figure 2–21, there is hardly a noticeable difference in the percentages. Thus, changing the units at the starting point on the *y* axis can convey a very different visual representation of the data.

FIGURE 2–20
Graph of Automaker's Claim Using a Scale from 95 to 100%

FIGURE 2–21
Graph in Figure 2–20 Redrawn Using a Scale from 0 to 100%



It is not wrong to truncate an axis of the graph; many times it is necessary to do so. However, the reader should be aware of this fact and interpret the graph accordingly. Do not be misled if an inappropriate impression is given.

Let us consider another example. The projected required fuel economy in miles per gallon for General Motors vehicles is shown. In this case, an increase from 21.9 to 23.2 miles per gallon is projected.

Year	2008	2009	2010	2011
MPG	21.9	22.6	22.9	23.2

Source: National Highway Traffic Safety Administration.

When you examine the graph shown in Figure 2–22(a), using a scale of 0 to 25 miles per gallon, the graph shows a slight increase. However, when the scale is changed to 21 to 24 miles per gallon, the graph shows a much larger increase even though the data remain the same. See Figure 2–22(b). Again, by changing the units or starting point on the *y* axis, one can change the visual representation.

FIGURE 2–22
Projected Miles per Gallon

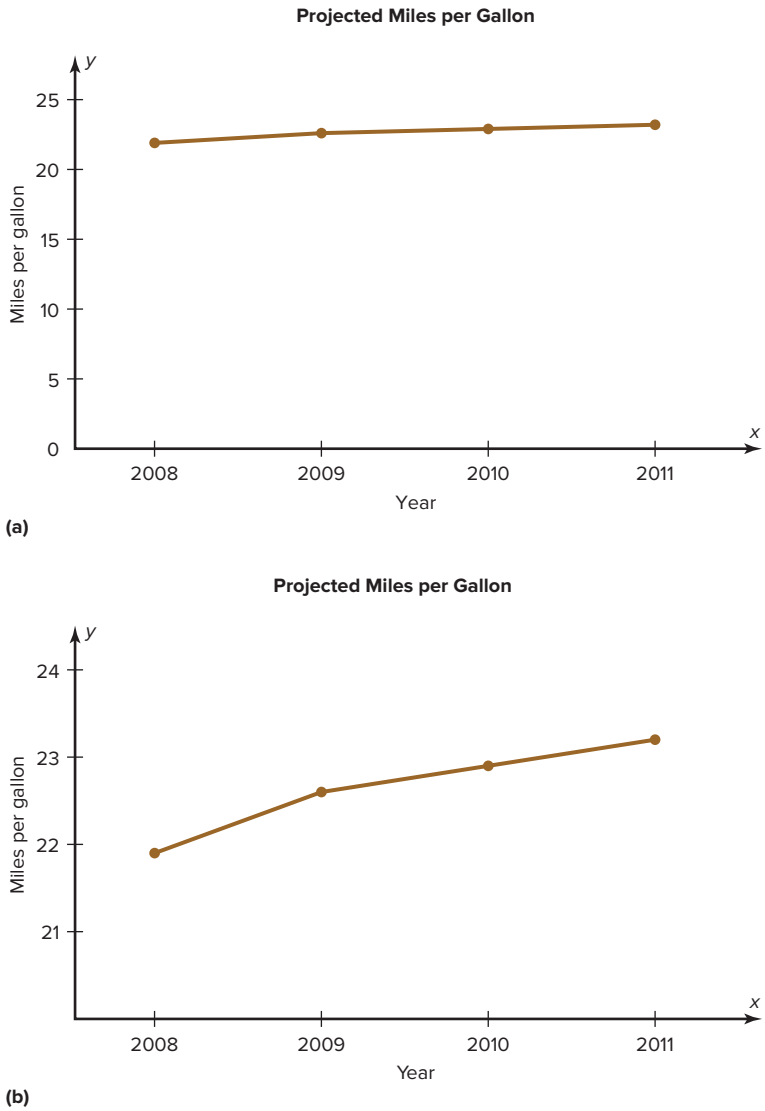
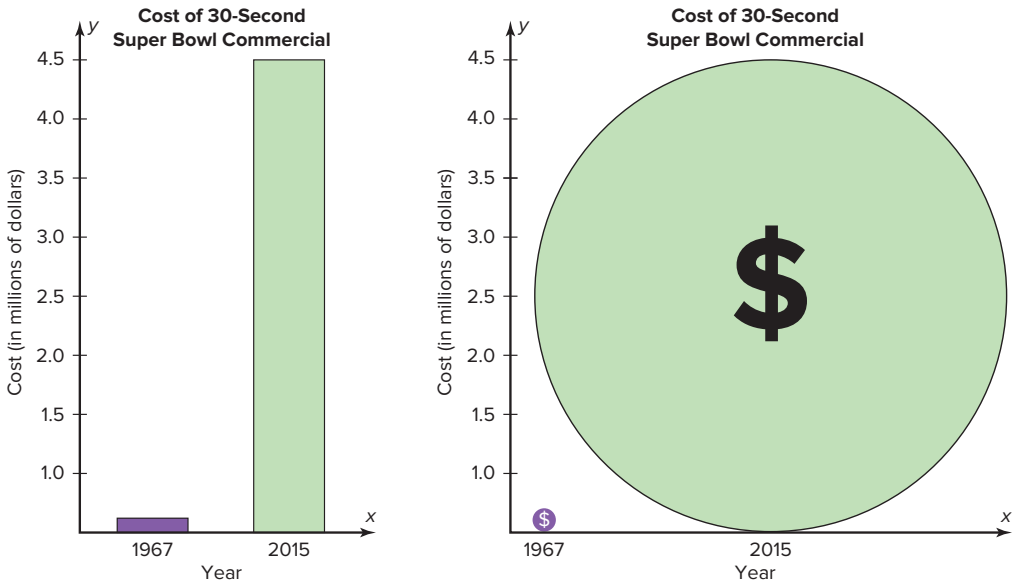


FIGURE 2–23
Comparison of Costs for a 30-Second Super Bowl Commercial



(a) Graph using bars

(b) Graph using circles

Another misleading graphing technique sometimes used involves exaggerating a one-dimensional increase by showing it in two dimensions. For example, the average cost of a 30-second Super Bowl commercial has increased from $42,000 in 1967 to $4.5 million in 2015 (*Source: USA TODAY*).

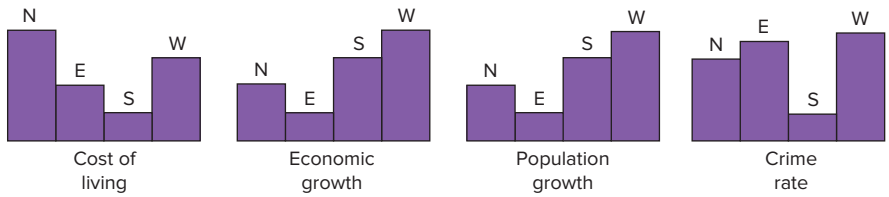
The increase shown by the graph in Figure 2–23(a) represents the change by a comparison of the heights of the two bars in one dimension. The same data are shown two-dimensionally with circles in Figure 2–23(b). Notice that the difference seems much larger because the eye is comparing the areas of the circles rather than the lengths of the diameters.

Note that it is not wrong to use the graphing techniques of truncating the scales or representing data by two-dimensional pictures. But when these techniques are used, the reader should be cautious of the conclusion drawn on the basis of the graphs.

Another way to misrepresent data on a graph is by omitting labels or units on the axes of the graph. The graph shown in Figure 2–24 compares the cost of living, economic growth, population growth, etc., of four main geographic areas in the United States. However, since there are no numbers on the *y* axis, very little information can be gained from this graph, except a crude ranking of each factor. There is no way to decide the actual magnitude of the differences.

Finally, all graphs should contain a source for the information presented. The inclusion of a source for the data will enable you to check the reliability of the organization presenting the data.

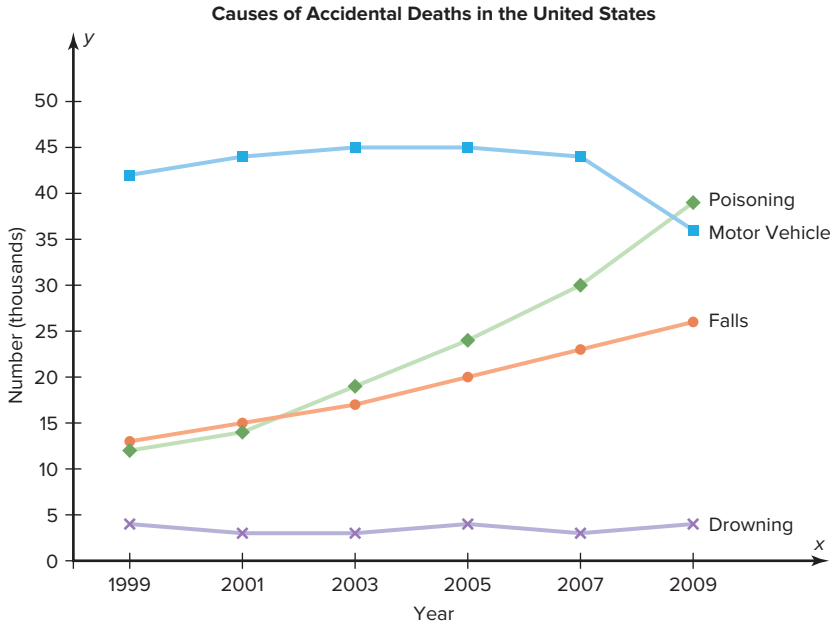
FIGURE 2–24
A Graph with No Units on the *y* Axis



Applying the Concepts 2–3

Causes of Accidental Deaths in the United States, 1999–2009

The graph shows the number of deaths in the United States due to accidents. Answer the following questions about the graph.



Source: National Safety Council.

1. Name the variables used in the graph.
2. Are the variables qualitative or quantitative?
3. What type of graph is used here?
4. Which variable shows a decrease in the number of deaths over the years?
5. Which variable or variables show an increase in the number of deaths over the years?
6. The number of deaths in which variable remains about the same over the years?
7. List the approximate number of deaths for each category for the year 2001.
8. In 1999, which variable accounted for the most deaths? In 2009, which variable accounted for the most deaths?
9. In what year were the numbers of deaths from poisoning and falls about the same?

See page 108 for the answers.

Exercises 2–3

1. Tech Company Employees Construct a vertical and horizontal bar graph for the number of employees (in thousands) of a sample of the largest tech companies as of 2014.

Company	Employees
IBM	380
Hewlett Packard	302
Xerox	147
Microsoft	128
Intel	107

Source: S & P Capital IQ

2. Worldwide Sales of Fast Foods The worldwide sales (in billions of dollars) for several fast-food franchises for a specific year are shown. Construct a vertical bar graph and a horizontal bar graph for the data.

Wendy's	$ 8.7
KFC	14.2
Pizza Hut	9.3
Burger King	12.7
Subway	10.0

Source: Franchise Times.

3. Gulf Coastlines Construct a Pareto chart for the sizes of Gulf coastlines in statute miles for each state.

State	Coastline
Alabama	53
Florida	770
Louisiana	397
Mississippi	44
Texas	367

4. Roller Coaster Mania The World Roller Coaster Census Report lists the following numbers of roller coasters on each continent. Represent the data graphically, using a Pareto chart.

Africa	17
Asia	315
Australia	22
Europe	413
North America	643
South America	45

Source: www.rcdb.com

5. Online Ad Spending The amount spent (in billions of dollars) for ads online is shown. (The numbers for 2016 through 2019 are projected numbers.) Draw a time series graph and comment on the trend.

Year	2014	2015	2016	2017	2018	2019
Amount	$19.72	$31.53	$43.83	$53.29	$61.14	$69.04

Source: eMarketer.

6. Violent Crimes The number of all violent crimes (murder, nonnegligent homicide, manslaughter, forcible rape, robbery, and aggravated assault) in the United States for each of these years is listed below. Represent the data with a time series graph.

2000	1,425,486	2004	1,360,088	2008	1,394,461
2001	1,439,480	2005	1,390,745	2009	1,325,896
2002	1,423,677	2006	1,435,123	2010	1,246,248
2003	1,383,676	2007	1,422,970		

Source: World Almanac and Book of Facts.

7. U.S. Licensed Drivers 70 or Older Draw a time series graph for the number (in millions) of drivers in the United States 70 or older

Year	1982	1992	2002	2012
Number	10	15	20	23

Source: Federal Highway Administration

8. Valentine's Day Spending The data show the average amount of money spent by consumers on Valentine's Day. Draw a time series graph for the data and comment on the trend.

Year	2007	2008	2009	2010	2011	2012
Amount	$120	$123	$103	$103	$110	$126

Source: National Retail Federation.

9. Credit Cards Draw and analyze a pie graph for the number of credit cards a person has.

Number of cards	0	1	2 or 3	4 or more
Number	52	40	68	40

Source: Based on information from *AARP Bulletin* survey

10. Reasons We Travel The following data are based on a survey from American Travel Survey on why people travel. Construct a pie graph for the data and analyze the results.

Purpose	Number
Personal business	146
Visit friends or relatives	330
Work-related	225
Leisure	299

Source: USA TODAY.

11. Kids and Guns The following data show where children obtain guns for committing crimes. Draw and analyze a pie graph for the data.

Source	Friend	Family	Street	Gun or Pawn Shop	Other
Number	24	15	9	9	6

12. Colors of Automobiles The popular car colors are shown. Construct a pie graph for the data.

White	19%
Silver	18
Black	16
Red	13
Blue	12
Gray	12
Other	10

Source: Dupont Automotive Color Popularity Report.

13. Ages of Football Players The data show the ages of the players of the Super Bowl L Denver Bronco Champs in 2016. Construct a dotplot for the data, and comment on the distribution.

24	23	25	25	26	30
30	33	23	32	21	26
24	24	27	26	30	24
26	28	24	23	39	26
34	25	24	26	24	23
24	29	25	26	30	22
23	28	25	24	34	27
29	28	23	25	28	28
29	33	25	27	25	

Source: Fansided.com

14. Teacher Strikes In Pennsylvania the numbers of teacher strikes for the last 14 years are shown. Construct a dotplot for the data. Comment on the graph.

9	13	15	7	7	14	9
10	14	18	7	8	8	3

Source: School Leader News.

15. Years of Experience The data show the number of years of experience the players on the Pittsburgh Steelers football team have at the beginning of the season. Draw and analyze a dot plot for the data.

4	4	2	9	7	3	7	12	6
5	1	4	5	2	7	6	12	3
12	4	0	4	0	0	0	2	9
2	6	7	13	4	2	6	9	4
4	0	3	5	4	2	6	9	4
4	0	3	5	3	11	1	4	2
3	15	1	6	0	11	3	10	3

16. Commuting Times Fifty off-campus students were asked how long it takes them to get to school. The times (in minutes) are shown. Construct a dotplot and analyze the data.

23	22	29	19	12
18	17	30	11	27
11	18	26	25	20
25	15	24	21	31
29	14	22	25	29
24	12	30	27	21
27	25	21	14	28
17	17	24	20	26
13	20	27	26	17
18	25	21	33	29

17. 50 Home Run Club There are 43 Major League baseball players (as of 2015) that have hit 50 or more home runs in one season. Construct a stem and leaf plot and analyze the data.

50	51	52	54	59	51	53
54	50	58	51	54	53	
56	58	56	70	54	52	
58	54	64	52	73	57	
50	60	56	50	66	54	
52	51	58	63	57	52	
51	50	61	52	65	50	

Source: The World Almanac and Book of Facts.

18. Calories in Salad Dressings A listing of calories per 1 ounce of selected salad dressings (not fat-free) is given below. Construct a stem and leaf plot for the data.

100	130	130	130	110	110	120	130	140	100
140	170	160	130	160	120	150	100	145	145
145	115	120	100	120	160	140	120	180	100
160	120	140	150	190	150	180	160		

19. Length of Major Rivers The data show the lengths (in hundreds of miles) of major rivers in South America and Europe. Construct a back-to-back stem and leaf plot, and compare the distributions.

South America				Europe				
39	21	10	10	5	12	7	6	8
11	10	2	10	5	5	4	6	
10	14	10	12	18	5	13	9	
17	15	10		14	6	6	11	
15	25	16		8	6	3	4	

Source: The World Almanac and Book of Facts.

20. Math and Reading Achievement Scores The math and reading achievement scores from the National Assessment of Educational Progress for selected states are listed below. Construct a back-to-back stem and leaf plot with the data, and compare the distributions.

Math					Reading				
52	66	69	62	61	65	76	76	66	67
63	57	59	59	55	71	70	70	66	61
55	59	74	72	73	61	69	78	76	77
68	76	73			77	77	80		

Source: World Almanac.

21. State which type of graph (Pareto chart, time series graph, or pie graph) would most appropriately represent the data.

a. Situations that distract automobile drivers
b. Number of persons in an automobile used for getting to and from work each day
c. Amount of money spent for textbooks and supplies for one semester
d. Number of people killed by tornados in the United States each year for the last 10 years
e. The number of pets (dogs, cats, birds, fish, etc.) in the United States this year
f. The average amount of money that a person spent for his or her significant other for Christmas for the last 6 years

22. State which graph (Pareto chart, time series graph, or pie graph) would most appropriately represent the given situation.

a. The number of students enrolled at a local college for each year during the last 5 years
b. The budget for the student activities department at a certain college for a specific year
c. The means of transportation the students use to get to school
d. The percentage of votes each of the four candidates received in the last election
e. The record temperatures of a city for the last 30 years
f. The frequency of each type of crime committed in a city during the year

23. U.S. Health Dollar The U.S. health dollar is spent as indicated below. Construct two different types of graphs to represent the data.

Government administration	9.7%
Nursing home care	5.5
Prescription drugs	10.1
Physician and clinical services	20.3
Hospital care	30.5
Other (OTC drugs, dental, etc.)	23.9

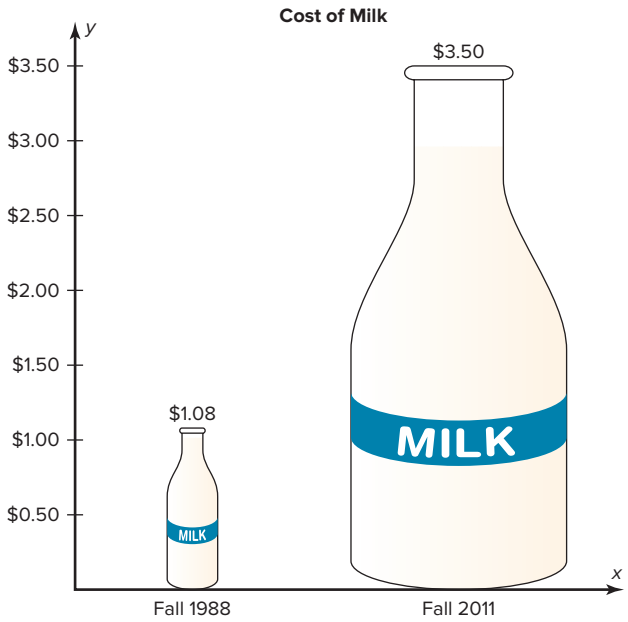
Source: Time Almanac.

24. Patents The U.S. Department of Commerce reports the following number of U.S. patents received by foreign countries and the United States in the year 2010. Illustrate the data with a bar graph and a pie graph. Which do you think better illustrates this data set?

Japan	44,814	United Kingdom	4,302
Germany	12,363	China	2,657
South Korea	11,671	Israel	1,819
Taiwan	8,238	Italy	1,796
Canada	4,852	United States	107,792

Source: World Almanac.

25. Cost of Milk The graph shows the increase in the price of a quart of milk. Why might the increase appear to be larger than it really is?



26. U.S. Population by Age The following information was found in a recent almanac. Use a pie graph to illustrate the information. Is there anything wrong with the data?

U.S. Population by Age in 2011

Under 20 years	27.0%
20 years and over	73.0
65 years and over	13.1

Source: Time Almanac.

27. Chicago Homicides Draw and compare two time series graphs for the number of homicides in the Chicago area.

Year	Homicides	As of June 29
2005	451	207
2007	448	204
2009	459	204
2011	435	187
2013	414	180

28. Trip Reimbursements The average amount requested for business trip reimbursement is itemized below. Illustrate the data with an appropriate graph. Do you have any questions regarding the data?

Flight	$440
Hotel stay	323
Entertainment	139
Phone usage	95
Transportation	65
Meal	38
Parking	34

Source: USA TODAY.

Technology

TI-84 Plus

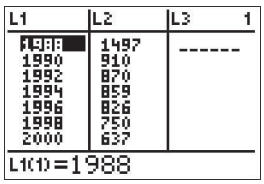
Step by Step

Step by Step

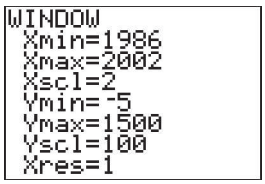
To graph a time series, follow the procedure for a frequency polygon from Section 2–2, using the following data for the number of outdoor drive-in theaters

Year	1988	1990	1992	1994	1996	1998	2000
Number	1497	910	870	859	826	750	637

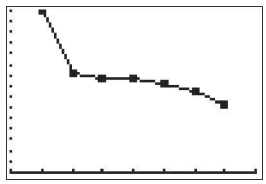
Input



Input



Output

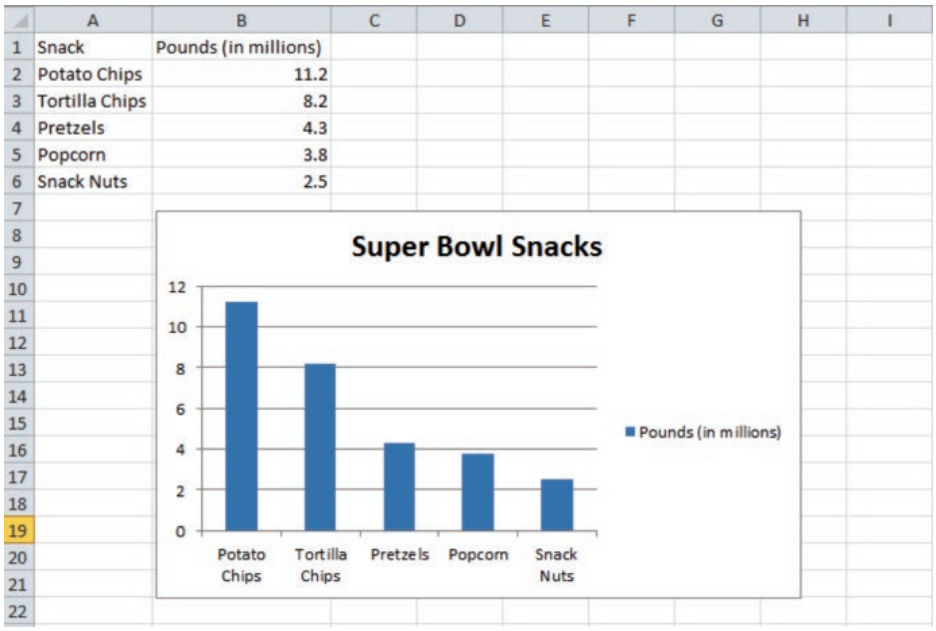


EXCEL
Step by Step

Constructing a Pareto Chart

To make a Pareto chart:

1. Enter the snack food categories from Example 2–11 into column A of a new worksheet.
2. Enter the corresponding frequencies in column B. The data should be entered in descending order according to frequency.
3. Highlight the data from columns A and B, and select the Insert tab from the toolbar.
4. Select the Column Chart type.
5. To change the title of the chart, click on the current title of the chart.
6. When the text box containing the title is highlighted, click the mouse in the text box and change the title.



Constructing a Time Series Chart

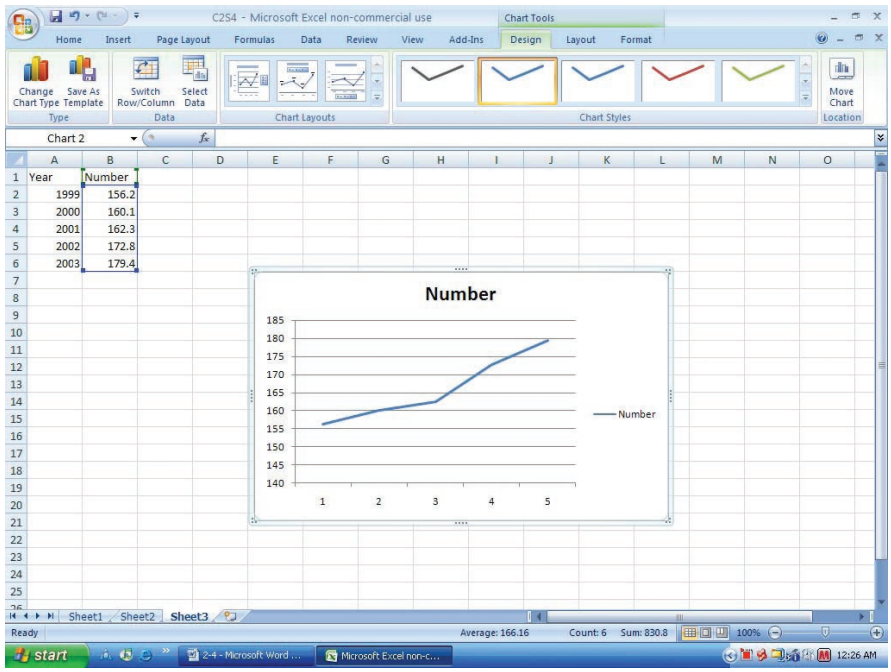
Example

Year	1999	2000	2001	2002	2003
Vehicles*	156.2	160.1	162.3	172.8	179.4

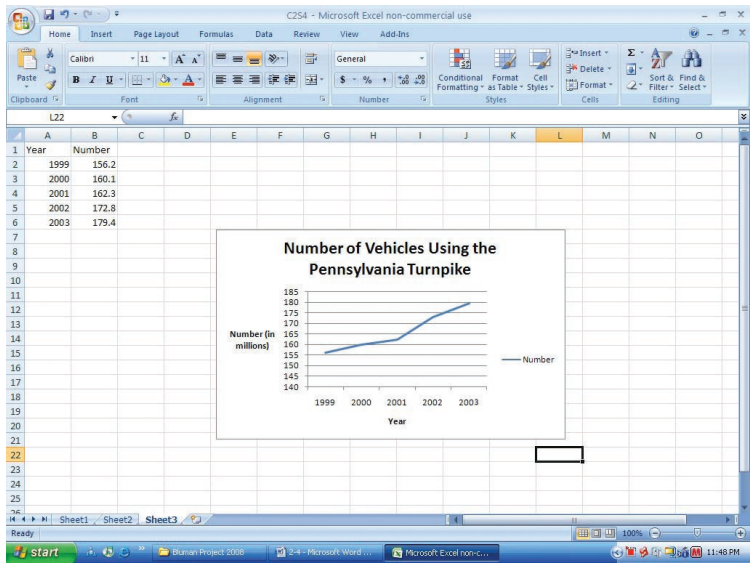
*Vehicles (in millions) that used the Pennsylvania Turnpike.
Source: Tribune Review.

To make a time series chart:

1. Enter the years 1999 through 2003 from the example in column A of a new worksheet.
2. Enter the corresponding frequencies in column B.
3. Highlight the data from column B and select the Insert tab from the toolbar.
4. Select the Line chart type.



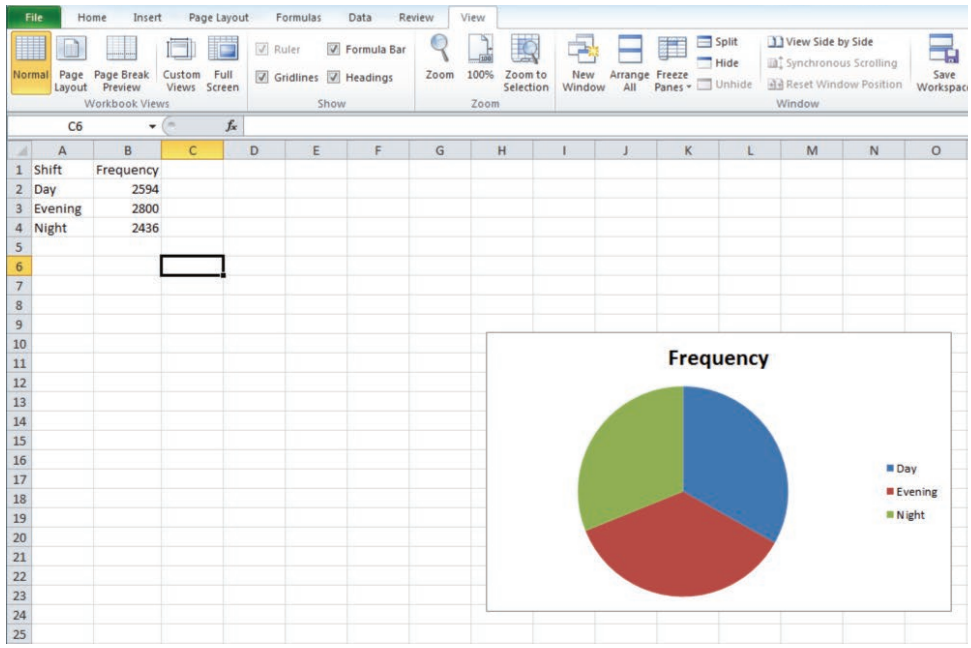
5. Right-click the mouse on any region of the graph.
6. Select the Select Data option.
7. Select Edit from the Horizontal Axis Labels and highlight the years from column A, then click [OK].
8. Click [OK] on the Select Data Source box.
9. Create a title for your chart, such as Number of Vehicles Using the Pennsylvania Turnpike Between 1999 and 2003. Right-click the mouse on any region of the chart. Select the Chart Tools tab from the toolbar, then Layout.
10. Select Chart Title and highlight the current title to change the title.
11. Select Axis Titles to change the horizontal and vertical axis labels.



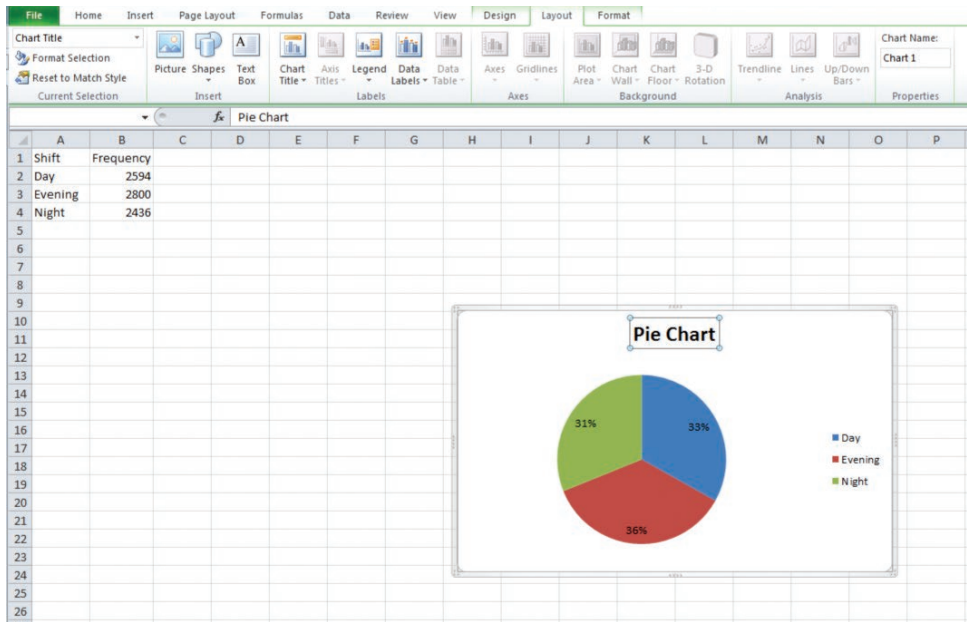
Constructing a Pie Chart

To make a pie chart:

1. Enter the shifts from Example 2–12 into column A of a new worksheet.
2. Enter the frequencies corresponding to each shift in column B.
3. Highlight the data in columns A and B and select Insert from the toolbar; then select the Pie chart type.



4. Click on any region of the chart. Then select Design from the Chart Tools tab on the toolbar.
5. Select Formulas from the chart Layouts tab on the toolbar.
6. To change the title of the chart, click on the current title of the chart.
7. When the text box containing the title is highlighted, click the mouse in the text box and change the title.

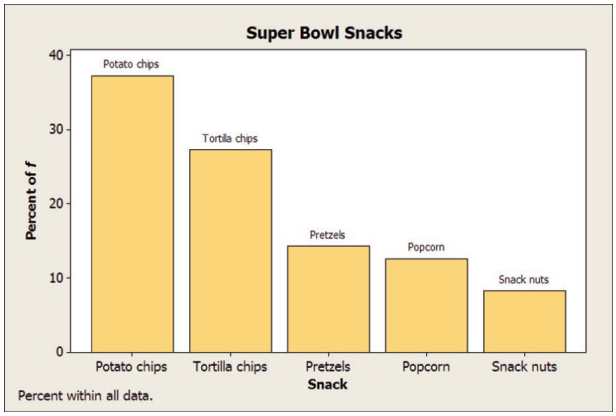


MINITAB
Step by Step

Construct a Bar Chart

The procedure for constructing a bar chart is similar to that for the pie chart.

1. Select **Graph>Bar Chart.**
   a) Click on the drop-down list in Bars Represent: and then select values from a table.
   b) Click on the Simple chart, then click [OK]. The dialog box will be similar to the Pie Chart Dialog Box.
2. Select the frequency column C2 f for Graph variables: and C1 Snack for the Categorical variable.



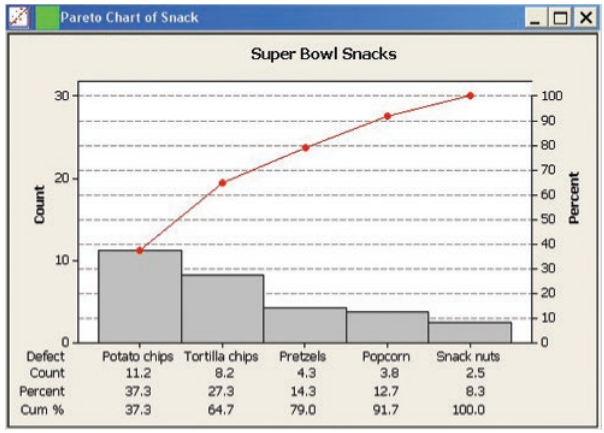
3. Click on [Labels], then type the title in the Titles/Footnote tab: **Super Bowl Snacks.**
4. Click the tab for Data Labels, then click the option to Use labels from column: and select C1 Snacks.
5. Click [OK] twice.

After the graph is made, right-click over any bar to change the appearance such as the color of the bars. To change the gap between them, right-click on the horizontal axis and then choose Edit X scale. In the Space Between Scale Categories select Gap between clusters then change the 1.5 to 0.2. Click [OK]. To change the *y* Scale to percents, right-click on the vertical axis and then choose Graph options and Show Y as a Percent.

Construct a Pareto Chart

Pareto charts are a quality control tool. They are similar to a bar chart with no gaps between the bars, and the bars are arranged by frequency.

1. Select **Stat>Quality Tools>Pareto.**
2. Click the option to Chart defects table.
3. Click in the box for the Labels in: and select C1 Snack.
4. Click on the frequencies column C2 f.



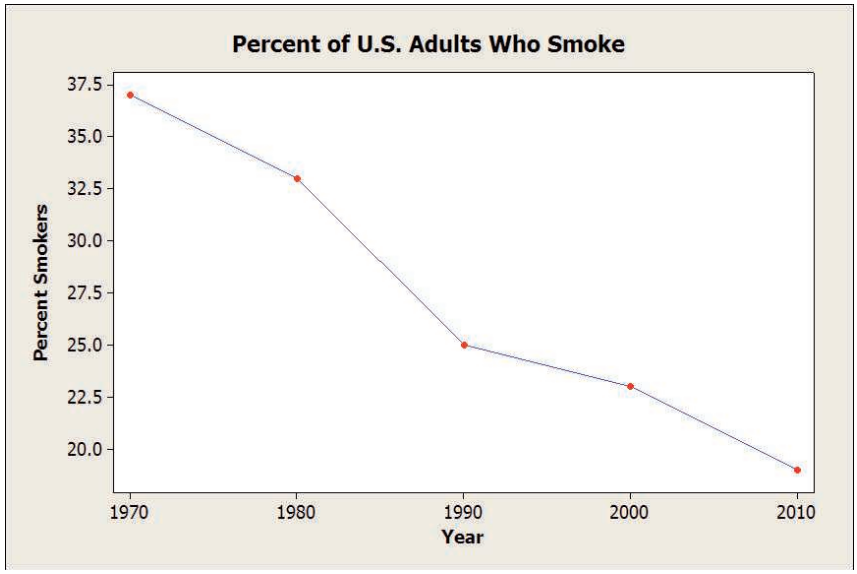
5. Click on [Options].
   a) Type Snack for the X axis label and Count for the Y axis label.
   b) Type in the title, **Super Bowl Snacks.**
6. Click [OK] twice. The chart is completed.

Construct a Time Series Plot

The data used are the percentage of U.S. adults who smoke (Example 2–10).

Year	1970	1980	1990	2000	2010
Number	37	33	25	23	19

1. Add a blank worksheet to the project by selecting **File>New>New-Minitab Worksheet.**
2. To enter the dates from 1970 to 2010 in C1, select **Calc>Make Patterned Data>Simple Set of Numbers.**
   a) Type **Year** in the text box for Store patterned data in.
   b) From First value: should be **1970.**
   c) To Last value: should be **2010.**
   d) In steps of should be **10** (for every 10-year increment). The last two boxes should be 1, the default value.
   e) Click [OK]. The sequence from 1970 to 2010 will be entered in C1 whose label will be Year.
3. Type **Percent Smokers** for the label row above row 1 in C2.
4. Type **37** for the first number, then press [Enter].
5. Continue entering each value in a row of C2.



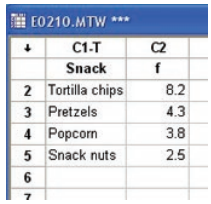
6. To make the graph, select **Graph>Time series plot,** then Simple, and press [OK].
   a) For Series select Percent Smokers; then click [Time/scale].
   b) Click the Stamp option and select Year for the Stamp column.
   c) Click the Gridlines tab and select all three boxes, Y major, Y minor, and X major.

d) Click [OK] twice. A new window will open that contains the graph.

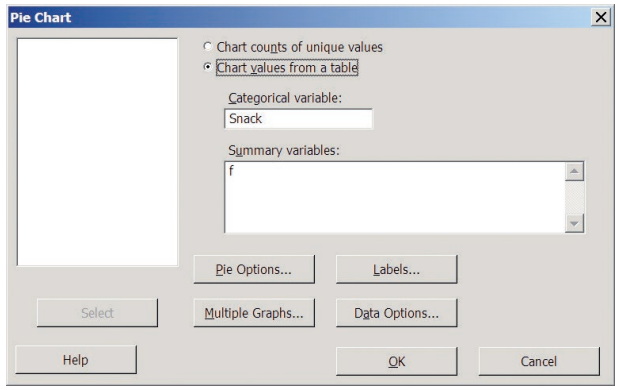
e) To change the title, double-click the title in the graph window. A dialog box will open, allowing you to change the text to Percent of U.S. Adults Who Smoke.

Construct a Pie Chart

1. Enter the summary data for snack foods and frequencies from Example 2–11 into C1 and C2.



↓	C1-T	C2
	Snack	f
2	Tortilla chips	8.2
3	Pretzels	4.3
4	Popcorn	3.8
5	Snack nuts	2.5
6		
7		



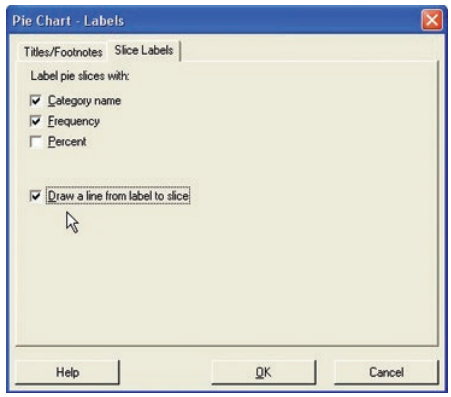
2. Name them **Snack** and **f.**

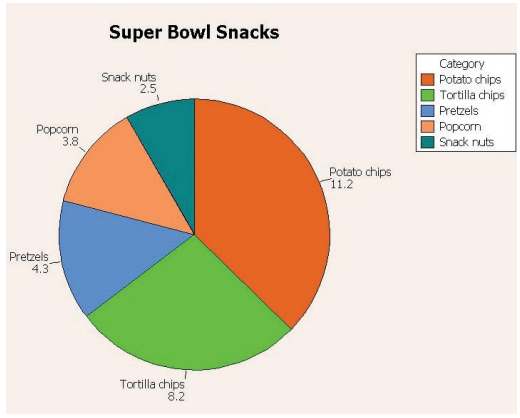
3. Select **Graph>Pie Chart.**

a) Click the option for Chart summarized data.

b) Press [Tab] to move to Categorical variable, then double-click C1 to select it.

c) Press [Tab] to move to Summary variables, and select the column with the frequencies f.





4. Click the [Labels] tab, then Titles/Footnotes.

a) Type in the title: **Super Bowl Snacks.**

b) Click the Slice Labels tab, then the options for Category name and Frequency.

c) Click the option to Draw a line from label to slice.

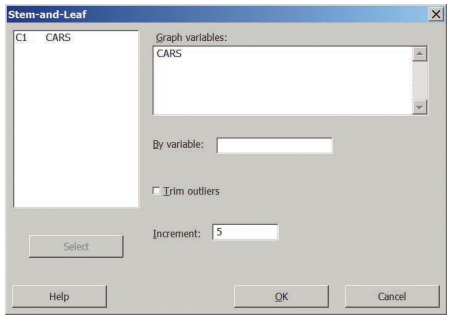
d) Click [OK] twice to create the chart.

Construct a Stem and Leaf Plot

1. Type in the data for Example 2–15. Label the column **CarThefts.**

2. Select **STAT>EDA>Stem-and-Leaf.** This is the same as **Graph>Stem-and-Leaf.**

3. Double-click on C1 CarThefts in the column list.
4. Click in the Increment text box, and enter the class width of 5.
5. Click [OK]. This character graph will be displayed in the session window.



Stem-and-Leaf Display: CarThefts

Stem-and-leaf of CarThefts N = 30
Leaf Unit = 1.0

6	5	011233
13	5	5567789
15	6	23
15	6	55667899
7	7	23
5	7	55789

Summary

- When data are collected, the values are called raw data. Since very little knowledge can be obtained from raw data, they must be organized in some meaningful way. A frequency distribution using classes is the common method that is used. (2–1)
- Once a frequency distribution is constructed, graphs can be drawn to give a visual representation of the data. The most commonly used graphs in statistics are the histogram, frequency polygon, and ogive. (2–2)
- Other graphs such as the bar graph, Pareto chart, time series graph, pie graph and dotplot can also be used. Some of these graphs are frequently seen in newspapers, magazines, and various statistical reports. (2–3)
- A stem and leaf plot uses part of the data values as stems and part of the data values as leaves. This graph has the advantage of a frequency distribution and a histogram. (2–3)
- Finally, graphs can be misleading if they are drawn improperly. For example, increases and decreases over time in time series graphs can be exaggerated by truncating the scale on the *y* axis. One-dimensional increases or decreases can be exaggerated by using two-dimensional figures. Finally, when labels or units are purposely omitted, there is no actual way to decide the magnitude of the differences between the categories. (2–3)

Important Terms

bar graph 75
categorical frequency distribution 43
class 42
class boundaries 45
class midpoint 45
class width 45
compound bar graphs 76
cumulative frequency 59
cumulative frequency distribution 48
dotplot 83
frequency 42
frequency distribution 42
frequency polygon 58
grouped frequency distribution 44
histogram 57
lower class limit 44
ogive 59
open-ended distribution 46
Pareto chart 77
pie graph 80
raw data 42
relative frequency graph 61
stem and leaf plot 84
time series graph 78
ungrouped frequency distribution 49
upper class limit 44

Important Formulas

Formula for the percentage of values in each class:

$$\% = \frac{f}{n} \cdot 100$$

where

f = **frequency of class**

n = **total number of values**

Formula for the range:

R = **highest value − lowest value**

Formula for the class width:

Class width = upper boundary − lower boundary

Formula for the class midpoint:

$$X_m = \frac{\text{lower boundary} + \text{upper boundary}}{2}$$

or

$$X_m = \frac{\text{lower limit} + \text{upper limit}}{2}$$

Formula for the degrees for each section of a pie graph:

$$\text{Degrees} = \frac{f}{n} \cdot 360°$$

Review Exercises

Section 2–1

1. How People Get Their News The Brunswick Research Organization surveyed 50 randomly selected individuals and asked them the primary way they received the daily news. Their choices were via newspaper (N), television (T), radio (R), or Internet (I). Construct a categorical frequency distribution for the data and interpret the results.

N	N	T	T	T	I	R	R	I	T
I	N	R	R	I	N	N	I	T	N
I	R	T	T	T	T	N	R	R	I
R	R	I	N	T	R	T	I	I	T
T	I	N	T	T	I	R	N	R	T

2. Men's World Hockey Champions The United States won the Men's World Hockey Championship in 1933 and 1960. Below are listed the world champions for the last 30 years. Use this information to construct a frequency distribution of the champions. What is the difficulty with these data?

1982	USSR	1999	Czech Republic
1983	USSR	2000	Czech Republic
1984	Not held	2001	Czech Republic
1985	Czechoslovakia	2002	Slovakia
1986	USSR	2003	Canada
1987	Sweden	2004	Canada
1988	Not held	2005	Czech Republic
1989	USSR	2006	Sweden
1990	Sweden	2007	Canada
1991	Sweden	2008	Russia
1992	Sweden	2009	Russia
1993	Russia	2010	Czech Republic
1994	Canada	2011	Finland
1995	Finland	2012	Russia
1996	Czech Republic	2013	Sweden
1997	Canada	2014	Russia
1998	Sweden	2015	Canada

Source: *Time Almanac.*

3. BUN Count The blood urea nitrogen (BUN) count of 20 randomly selected patients is given here in milligrams per deciliter (mg/dl). Construct an ungrouped frequency distribution for the data.

17	18	13	14
12	17	11	20
13	18	19	17
14	16	17	12
16	15	19	22

4. Wind Speed The data show the average wind speed for 36 days in a large city. Construct an ungrouped frequency distribution for the data.

8	15	9	8	9	10
8	10	14	9	8	8
12	9	8	8	14	9
9	13	13	10	12	9
13	8	11	11	9	8
9	13	9	8	8	10

5. Waterfall Heights The data show the heights (in feet) of notable waterfalls in North America. Organize the data into a grouped frequency distribution using 6 classes. This data will be used for Exercises 7, 9, and 11.

90	420	300	194
640	68	268	276
620	76	165	833
370	53	132	600
594	70	308	
574	215	109	
317	850	212	
300	256	187	

Source: National Geographic Society

6. Ages of the Vice Presidents at the Time of Their Death The ages at the time of death of those Vice Presidents of the United States who have passed away are listed below. Use the data to construct a frequency distribution.

Use 6 classes. The data for this exercise will be used for Exercises 8, 10, and 12.

90	83	80	73	70	51	68	79	70	71	
72	74	67	54	81	66	62	63	68	57	
66	96	78	55	60	66	57	71	60	85	
76	98	77	88	78	81	64	66	77	93	70

Source: World Almanac and Book of Facts.

Section 2–2

7. Find the relative frequency for the frequency distribution for the data in Exercise 5.

8. Find the relative frequency for the frequency distribution for the data in Exercise 6.

9. Construct a histogram, frequency polygon, and ogive for the data in Exercise 5.

10. Construct a histogram, frequency polygon, and ogive for the data in Exercise 6.

11. Construct a histogram, frequency polygon, and ogive, using relative frequencies for the data in Exercise 5.

12. Construct a histogram, frequency polygon, and ogive, using relative frequencies for the data in Exercise 6.

Section 2–3

13. Non-Alcoholic Beverages The data show the yearly consumption (in gallons) of popular non-alcoholic beverages. Draw a vertical and horizontal bar graph to represent the data.

Soft drinks	52
Water	34
Milk	26
Coffee	21

Source: U.S. Department of Agriculture

14. Calories of Nuts The data show the number of calories per ounce in selected types of nuts. Construct vertical and horizontal bar graphs for the data.

Types	Calories
Peanuts	160
Almonds	170
Macadamia	200
Pecans	190
Cashews	160

15. Crime The data show the percentage of the types of crimes commonly committed in the United States. Construct a Pareto chart for the data.

Theft	55%
Burglary	20%
Motor Vehicle Theft	11%
Assault	8%
Rape & Homicide	1%

Source: FBI

16. Pet Care The data (in billions of dollars) show the estimated amount of money spent on pet care in the United States. Construct a Pareto chart for the data.

Type of care	Amount spent
Veterinarian care	$14
Supplies and medicine	11
Grooming and boarding	4
Animal purchases	2

Source: American Pet Products Association.

17. Broadway Stage Engagements The data show the number of new Broadway productions for the seasons. Construct and analyze a time series graph for the data.

Season	New Productions
2004	39
2005	39
2006	35
2007	39
2008	43
2009	39
2010	42
2011	41
2012	46
2013	44

Source: The Broadway League

18. High School Dropout Rate The data show the high school dropout rate for students for the years 2003 to 2013. Construct a time series graph and analyze the graph.

Year	Percent
2003	9.9
2004	10.3
2005	9.4
2006	9.3
2007	8.7
2008	8.0
2009	8.1
2010	7.4
2011	7.1
2012	6.6
2013	6.8

Source: U.S. Department of Commerce.

19. Spending of College Freshmen The average amounts spent by college freshmen for school items are shown. Construct a pie graph for the data.

Electronics/computers	$728
Dorm items	344
Clothing	141
Shoes	72

Source: National Retail Federation.

20. Smart Phone Insurance Construct and analyze a pie graph for the people who did or did not buy insurance for their smart phones at the time of purchase.

Response	Frequency
Yes	573
No	557
Don't Remember	166
Not offered any	211

Source: Based on information from Anderson Analytics

21. Peyton Manning's Colts Career Peyton Manning played for the Indianapolis Colts for 14 years. (He did not play in 2011.) The data show the number of touchdowns he scored for the years 1998–2010. Construct a dotplot for the data and comment on the graph.

26	33	27	49	31	27	33
26	26	29	28	31	33	

Source: NFL.com

22. Songs on CDs The data show the number of songs on each of 40 CDs from the author's collection. Construct a dotplot for the data and comment on the graph.

10	14	18	11
11	15	16	10
10	17	10	15
22	9	14	12
18	12	12	15
21	22	20	15
10	19	20	21
17	9	13	15
11	12	12	9
14	20	12	10

23. Weights of Football Players A local football team has 30 players; the weight of each player is shown. Construct a stem and leaf plot for the data. Use stems 20__, 21__, 22__, etc.

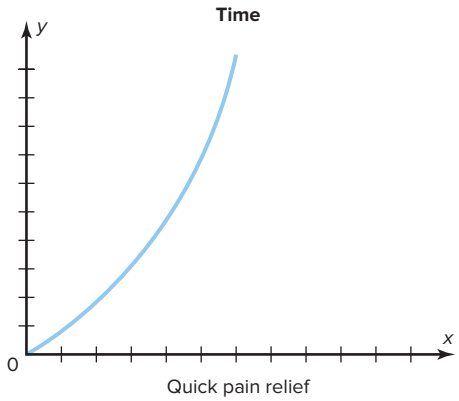
213	202	232	206	219
246	248	239	215	221
223	220	203	233	249
238	254	223	218	224
258	227	230	256	254
219	235	262	233	263

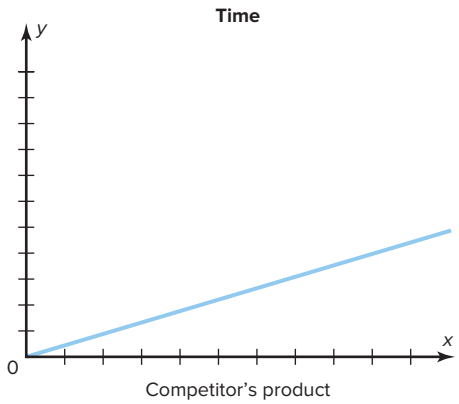
24. Public Libraries The numbers of public libraries in operation for selected states are listed below. Organize the data with a stem and leaf plot.

102	176	210	142	189	176	108	113	205
209	184	144	108	192	176			

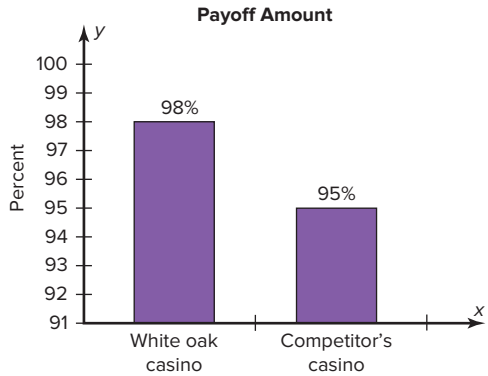
Source: World Almanac.

25. Pain Relief The graph below shows the time it takes Quick Pain Relief to relieve a person's pain. The graph below that shows the time a competitor's product takes to relieve pain. Why might these graphs be misleading?





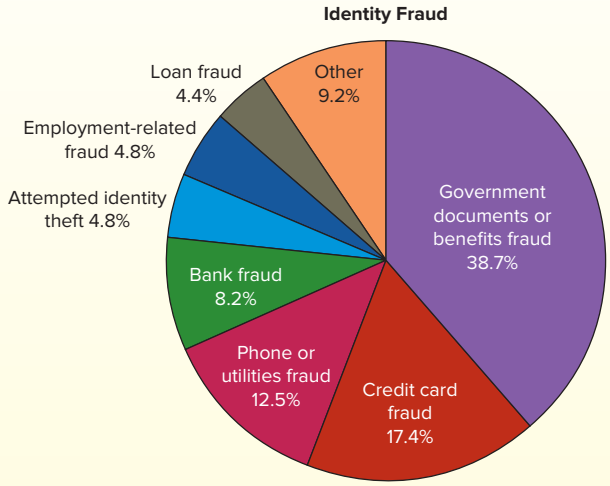
26. Casino Payoffs The graph shows the payoffs obtained from the White Oak Casino compared to the nearest competitor's casino. Why is this graph misleading?



STATISTICS TODAY

How Your Identity Can Be Stolen —Revisited

Data presented in numerical form do not convey an easy-to-interpret conclusion; however, when data are presented in graphical form, readers can see the visual impact of the numbers. In the case of identity fraud, the reader can see that most of the identity frauds are due to lost or stolen wallets, checkbooks, or credit cards, and very few identity frauds are caused by online purchases or transactions.



The Federal Trade Commission suggests some ways to protect your identity:

1. Shred all financial documents no longer needed.
2. Protect your Social Security number.
3. Don't give out personal information on the phone, through the mail, or over the Internet.
4. Never click on links sent in unsolicited emails.
5. Don't use an obvious password for your computer documents.
6. Keep your personal information in a secure place at home.

Data Analysis

A Data Bank is found in Appendix B, or on the World Wide Web by following links from www.mhhe.com/math/stat/bluman

1. From the Data Bank located in Appendix B, choose one of the following variables: age, weight, cholesterol level, systolic pressure, IQ, or sodium level. Select at least 30 values. For these values, construct a grouped frequency distribution. Draw a histogram, frequency polygon, and ogive for the distribution. Describe briefly the shape of the distribution.
2. From the Data Bank, choose one of the following variables: educational level, smoking status, or exercise. Select at least 20 values. Construct an ungrouped frequency distribution for the data. For the distribution, draw a Pareto chart and describe briefly the nature of the chart.
3. From the Data Bank, select at least 30 subjects and construct a categorical distribution for their marital status. Draw a pie graph and describe briefly the findings.
4. Using the data from Data Set IV in Appendix B, construct a frequency distribution and draw a histogram. Describe briefly the shape of the distribution of the tallest buildings in New York City.
5. Using the data from Data Set XI in Appendix B, construct a frequency distribution and draw a frequency polygon. Describe briefly the shape of the distribution for the number of pages in statistics books.

6. Using the data from Data Set IX in Appendix B, divide the United States into four regions, as follows:

Northeast	CT ME MA NH NJ NY PA RI VT
Midwest	IL IN IA KS MI MN MD MS NE ND OH SD WI
South	AL AR DE DC FL GA KY LA MD NC OK SC TN TX VA WV
West	AK AZ CA CO HI ID MT NV NM OR UT WA WY

Find the total population for each region, and draw a Pareto chart and a pie graph for the data. Analyze the results. Explain which chart might be a better representation for the data.

7. Using the data from Data Set I in Appendix B, make a stem and leaf plot for the record low temperatures in the United States. Describe the nature of the plot.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. In the construction of a frequency distribution, it is a good idea to have overlapping class limits, such as 10–20, 20–30, 30–40.

2. Bar graphs can be drawn by using vertical or horizontal bars.

3. It is not important to keep the width of each class the same in a frequency distribution.

4. Frequency distributions can aid the researcher in drawing charts and graphs.

5. The type of graph used to represent data is determined by the type of data collected and by the researcher's purpose.

6. In construction of a frequency polygon, the class limits are used for the *x* axis.

7. Data collected over a period of time can be graphed by using a pie graph.

Select the best answer.

8. What is another name for the ogive?
a. Histogram
b. Frequency polygon
c. Cumulative frequency graph
d. Pareto chart

9. What are the boundaries for 8.6–8.8?
a. 8–9
b. 8.5–8.9
c. 8.55–8.85
d. 8.65–8.75

10. What graph should be used to show the relationship between the parts and the whole?
a. Histogram
b. Pie graph
c. Pareto chart
d. Ogive

11. Except for rounding errors, relative frequencies should add up to what sum?
a. 0
b. 1
c. 50
d. 100

Complete these statements with the best answers.

12. The three types of frequency distributions are _____, _____, and _____.

13. In a frequency distribution, the number of classes should be between _____ and _____.

14. Data such as blood types (A, B, AB, O) can be organized into a(n) _____ frequency distribution.

15. Data collected over a period of time can be graphed using a(n) _____ graph.

16. A statistical device used in exploratory data analysis that is a combination of a frequency distribution and a histogram is called a(n) _____.

17. On a Pareto chart, the frequencies should be represented on the _____ axis.

18. Housing Arrangements A questionnaire on housing arrangements showed this information obtained from 25 respondents. Construct a frequency distribution for the data (H = house, A = apartment, M = mobile home, C = condominium). These data will be used in Exercise 19.

H	C	H	M	H	A	C	A	M
C	M	C	A	M	A	C	C	M
C	C	H	A	H	H	M		

19. Construct a pie graph for the data in Exercise 18.

20. Items Purchased at a Convenience Store When 30 randomly selected customers left a convenience store, each was asked the number of items he or she purchased. Construct an ungrouped frequency

distribution for the data. These data will be used in Exercise 21.

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

21. Construct a histogram, a frequency polygon, and an ogive for the data in Exercise 20.

22. Coal Consumption The following data represent the energy consumption of coal (in billions of Btu) by each of the 50 states and the District of Columbia. Use the data to construct a frequency distribution and a relative frequency distribution with 7 classes.

631	723	267	60	372	15	19	92	306	38
413	8	736	156	478	264	1015	329	679	1498
52	1365	142	423	365	350	445	776	1267	0
26	356	173	373	335	34	937	250	33	84
0	253	84	1224	743	582	2	33	0	426
474									

Source: Time Almanac.

23. Construct a histogram, frequency polygon, and ogive for the data in Exercise 22. Analyze the histogram.

24. Recycled Trash Construct a Pareto chart and a horizontal bar graph for the number of tons (in millions) of trash recycled per year by Americans based on an Environmental Protection Agency study.

Type	Amount
Paper	320.0
Iron/steel	292.0
Aluminum	276.0
Yard waste	242.4
Glass	196.0
Plastics	41.6

Source: USA TODAY.

25. Identity Thefts The results of a survey of 84 people whose identities were stolen using various methods are shown. Draw a pie chart for the information.

Lost or stolen wallet, checkbook, or credit card	38
Retail purchases or telephone transactions	15
Stolen mail	9
Computer viruses or hackers	8
Phishing	4
Other	10
	84

Source: Javelin Strategy and Research.

26. Needless Deaths of Children *The New England Journal of Medicine* predicted the number of needless deaths due to childhood obesity. Draw a time series graph for the data.

Year	2020	2025	2030	2035
Deaths	130	550	1500	3700

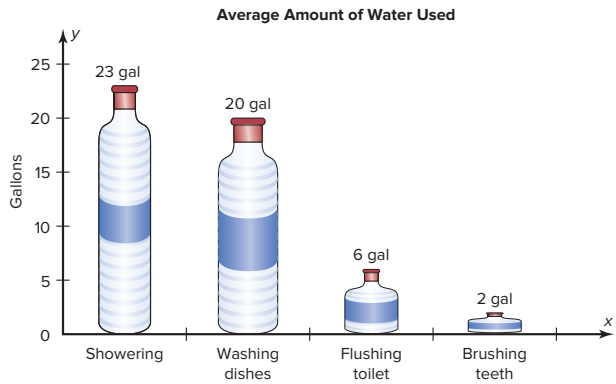
27. Museum Visitors The number of visitors to the Historic Museum for 25 randomly selected hours is shown. Construct a stem and leaf plot for the data.

15	53	48	19	38
86	63	98	79	38
62	89	67	39	26
28	35	54	88	76
31	47	53	41	68

28. Parking Meter Revenue In a small city the number of quarters collected from the parking meters is shown. Construct a dotplot for the data.

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

29. Water Usage The graph shows the average number of gallons of water a person uses for various activities. Can you see anything misleading about the way the graph is drawn?



Critical Thinking Challenges

1. The Great Lakes Shown are various statistics about the Great Lakes. Using appropriate graphs (your choice) and summary statements, write a report analyzing the data.

	Superior	Michigan	Huron	Erie	Ontario
Length (miles)	350	307	206	241	193
Breadth (miles)	160	118	183	57	53
Depth (feet)	1,330	923	750	210	802
Volume (cubic miles)	2,900	1,180	850	116	393
Area (square miles)	31,700	22,300	23,000	9,910	7,550
Shoreline (U.S., miles)	863	1,400	580	431	300

Source: The World Almanac and Book of Facts.

2. Teacher Strikes In Pennsylvania there were more teacher strikes in 2004 than there were in all other states combined. Because of the disruptions, state legislators want to pass a bill outlawing teacher strikes and submitting contract disputes to binding arbitration. The graph shows the number of teacher strikes in Pennsylvania for the school years 1997 to 2011. Use the graph to answer these questions.

a. In what year did the largest number of strikes occur? How many were there?

b. In what year did the smallest number of teacher strikes occur? How many were there?

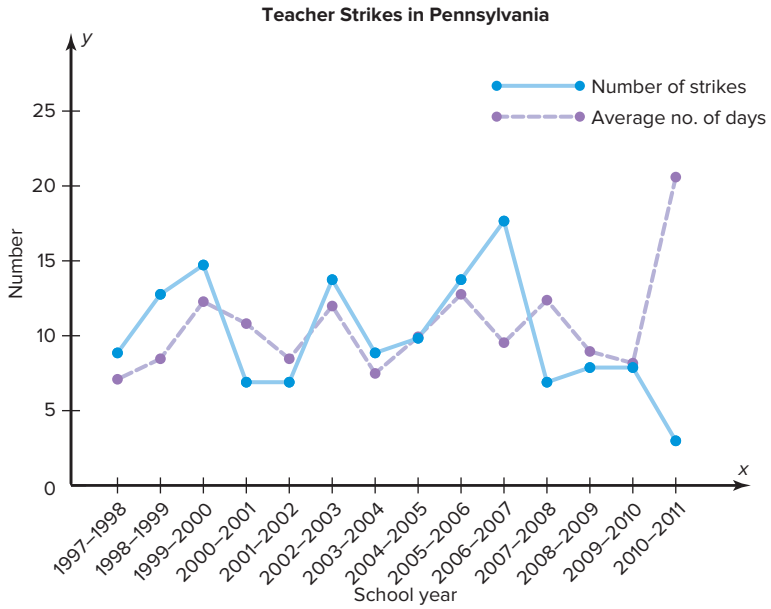
c. In what year was the average duration of the strikes the longest? How long was it?

d. In what year was the average duration of the strikes the shortest? How long was it?

e. In what year was the number of teacher strikes the same as the average duration of the strikes?

f. Find the difference in the number of strikes for the school years 1997–1998 and 2010–2011.

g. Do you think teacher strikes should be outlawed? Justify your conclusions.



Source: Pennsylvania School Boards Association.

Data Projects

Where appropriate, use the TI-84 Plus, Excel, MINITAB, or a computer program of your choice to complete the following exercises.

1. Business and Finance Consider the 30 stocks listed as the Dow Jones Industrials. For each, find its earnings per share. Randomly select 30 stocks traded on the NASDAQ. For each, find its earnings per share. Create a frequency table with 5 categories for each data set. Sketch a histogram for each. How do the two data sets compare?

2. **Sports and Leisure** Use systematic sampling to create a sample of 25 National League and 25 American League baseball players from the most recently completed season. Find the number of home runs for each player. Create a frequency table with 5 categories for each data set. Sketch a histogram for each. How do the two leagues compare?

3. **Technology** Randomly select 50 songs from your music player or music organization program. Find the length (in seconds) for each song. Use these data to create a frequency table with 6 categories. Sketch a frequency polygon for the frequency table. Is the shape of the distribution of times uniform, skewed, or bell-shaped? Also note the genre of each song. Create a Pareto chart showing the frequencies of the various categories. Finally, note the year each song was released. Create a pie chart organized by decade to show the percentage of songs from various time periods.

4. **Health and Wellness** Use information from the Red Cross to create a pie chart depicting the percentages of Americans with various blood types. Also find information about blood donations and the percentage of each type donated. How do the charts compare? Why is the collection of type O blood so important?

5. **Politics and Economics** Consider the U.S. Electoral College System. For each of the 50 states, determine the number of delegates received. Create a frequency table with 8 classes. Is this distribution uniform, skewed, or bell-shaped?

6. **Your Class** Have each person in class take his or her pulse and determine the heart rate (beats in 1 minute). Use the data to create a frequency table with 6 classes. Then have everyone in the class do 25 jumping jacks and immediately take the pulse again after the activity. Create a frequency table for those data as well. Compare the two results. Are they similarly distributed? How does the range of scores compare?

Answers to Applying the Concepts

Section 2–1 Ages of Presidents at Inauguration

1. The data were obtained from the population of all Presidents at the time this text was written.
2. The oldest inauguration age was 69 years old.
3. The youngest inauguration age was 42 years old.
4. Answers will vary. One possible answer is

Age at inauguration	Frequency
42–45	2
46–49	7
50–53	8
54–57	16
58–61	5
62–65	4
66–69	2

5. Answers will vary. For the frequency distribution given in question 4, there is a peak for the 54–57 class.
6. Answers will vary. This frequency distribution shows no outliers. However, if we had split our frequency into 14 classes instead of 7, then the ages 42, 43, 68, and 69 might appear as outliers.
7. Answers will vary. The data appear to be unimodal and fairly symmetric, centering on 55 years of age.

Section 2–2 Selling Real Estate

1. A histogram of the data gives price ranges and the counts of homes in each price range. We can also talk about how the data are distributed by looking at a histogram.
2. A frequency polygon shows increases or decreases in the number of home prices around values.
3. A cumulative frequency polygon shows the number of homes sold at or below a given price.
4. The house that sold for $321,550 is an extreme value in this data set.
5. Answers will vary. One possible answer is that the histogram displays the outlier well since there is a gap in the prices of the homes sold.
6. The distribution of the data is skewed to the right.

Section 2–3 Causes of Accidental Deaths in the United States

1. The variables in the graph are the year, cause of death, and number of deaths in thousands.
2. The cause of death is qualitative, while the year and number of deaths are quantitative.
3. A time series graph is used here.
4. The motor vehicle accidents showed a slight increase from 1999 to 2007, and then a decrease.
5. The number of deaths due to poisoning and falls is increasing.
6. The number of deaths due to drowning remains about the same over the years.
7. For 2001, about 44,000 people died in motor vehicle accidents, about 14,000 people died from falls, about 15,000 people died from poisoning, and about 3000 people died from drowning.
8. In 1999, motor vehicle accidents claimed the most lives, while in 2009, poisoning claimed the most lives.
9. Around 2002, the number of deaths from falls and poisoning were about the same.