CHAPTER 2

Frequency Distributions and Graphs

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, and pie graphs.
4. Draw and interpret a stem and leaf plot.

Outline

Introduction

2–1 Organizing Data

2–2 Histograms, Frequency Polygons, and Ogives

2–3 Other Types of Graphs

Summary

Statistics Today

How Your Identity Can Be Stolen

Identity fraud is a big business today. The total amount of the fraud in 2006 was $56.6 billion. The average amount of the fraud for a victim is $6383, 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:

Lost or stolen wallet, checkbook, or credit card	38%
Friends, acquaintances	15
Corrupt business employees	15
Computer viruses and hackers	9
Stolen mail or fraudulent change of address	8
Online purchases or transactions	4
Other methods	11

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.

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.

Wealthy People

Suppose a researcher wished to do a study on the ages of the top 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.

49	57	38	73	81
74	59	76	65	69
54	56	69	68	78
65	85	49	69	61
48	81	68	37	43
78	82	43	64	67
52	56	81	77	79
85	40	85	59	80
60	71	57	61	69
61	83	90	87	74

Since little information can be obtained from looking at raw data, the researcher organizes the data into what is called a *frequency distribution.* A frequency distribution consists of *classes* and their corresponding *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
35–41	///	3
42–48	///	3
49–55	////	4
56–62	𝍸 𝍸	10
63–69	𝍸 𝍸	10
70–76	𝍸	5
77–83	𝍸 𝍸	10
84–90	𝍸	5
	Total	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 over 55 years old.

Unusual Stat

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.

The classes in this distribution are 35–41, 42–48, etc. These values are called *class limits.* The data values 35, 36, 37, 38, 39, 40, 41 can be tallied in the first class; 42, 43, 44, 45, 46, 47, 48 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.

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 number of hours that boat batteries lasted is the following.

Unusual Stat

Six percent of Americans say they find life dull.

Class limits	Class boundaries	Tally	Frequency
24–30	23.5–30.5	///	3
31–37	30.5–37.5	/	1
38–44	37.5–44.5	𝍸	5
45–51	44.5–51.5	𝍸 ////	9
52–58	51.5–58.5	𝍸 /	6
59–65	58.5–65.5	/	1
			25

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 24 and 30 of the first class are called *class limits.* The **lower class limit** is 24; it represents the smallest data value that can be included in the class. The **upper class limit** is 30; 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 30 and 31.

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 24, 32, and 18, the limits for a class might be 31–37, and the boundaries are 30.5–37.5. Find the boundaries by subtracting 0.5 from 31 (the lower class limit) and adding 0.5 to 37 (the upper class limit).

$$\text{Lower limit} - 0.5 = 31 - 0.5 = 30.5 = \text{lower boundary}$$
$$\text{Upper limit} + 0.5 = 37 + 0.5 = 37.5 = \text{upper boundary}$$

Unusual Stat

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.

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 duration of boat batteries is 7, found from $31 - 24 = 7$.

The class width can also be found by subtracting the lower boundary from the upper boundary for any given class. In this case, $30.5 - 23.5 = 7$.

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 the 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 boat batteries is

$$\frac{24 + 30}{2} = 27 \qquad \text{or} \qquad \frac{23.5 + 30.5}{2} = 27$$

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

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

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 class has no specific beginning value or no specific ending value. 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.

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.

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	𝍸 𝍸 𝍸 ///	18
115–119	114.5–119.5	𝍸 𝍸 ///	13
120–124	119.5–124.5	𝍸 //	7
125–129	124.5–129.5	/	1
130–134	129.5–134.5	/	1
			$n = \Sigma f = 50$

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 109.5 and 119.5°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, 28 of the total record high temperatures 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 MPGs for SUVs

The data shown here represent the number of miles per gallon (mpg) that 30 selected four-wheel-drive sports utility vehicles obtained in city driving. Construct a frequency distribution, and analyze the distribution.

12	17	12	14	16	18
16	18	12	16	17	15
15	16	12	15	16	16
12	14	15	12	15	15
19	13	16	18	16	14

Source: *Model Year Fuel Economy Guide.* United States Environmental Protection Agency.

Solution

Step 1 Determine the classes. Since the range of the data set is small (19 − 12 = 7), classes consisting of a single data value can be used. They are 12, 13, 14, 15, 16, 17, 18, 19.

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

The completed ungrouped frequency distribution is

Class limits	Class boundaries	Tally	Frequency
12	11.5–12.5	𝍸 /	6
13	12.5–13.5	/	1
14	13.5–14.5	///	3
15	14.5–15.5	𝍸 /	6
16	15.5–16.5	𝍸 ///	8
17	16.5–17.5	//	2
18	17.5–18.5	///	3
19	18.5–19.5	/	1

In this case, almost one-half (14) of the vehicles get 15 or 16 miles per gallon. The cumulative frequencies are

	Cumulative frequency
Less than 11.5	0
Less than 12.5	6
Less than 13.5	7
Less than 14.5	10
Less than 15.5	16
Less than 16.5	24
Less than 17.5	26
Less than 18.5	29
Less than 19.5	30

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.

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.

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.

Interesting Fact

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

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	50	47	55	55	54	42	51
56	55	54	51	60	62	43	55	56
61	52	69	64	46	54			

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. Identify any possible outliers.
7. Write a brief summary of the nature of the data as shown in the frequency distribution.

See page 101 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. Find the class boundaries, midpoints, and widths for each class.

a. 12–18
b. 56–74
c. 695–705
d. 13.6–14.7
e. 2.15–3.93

4. How many classes should frequency distributions have? Why should the class width be an odd number?

5. Shown here are four frequency distributions. Each is incorrectly constructed. State the reason why.

a.

Class	Frequency
27–32	1
33–38	0
39–44	6
45–49	4
50–55	2

b.

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

c.

Class	Frequency
123–127	3
128–132	7
138–142	2
143–147	19

d.

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

6. What are open-ended frequency distributions? Why are they necessary?

7. 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 everything 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

8. State Gasoline Tax The state gas tax in cents per gallon for 25 states is given below. Construct a grouped frequency distribution and a cumulative frequency distribution with 5 classes.

7.5	16	23.5	17	22
21.5	19	20	27.1	20
22	20.7	17	28	20
23	18.5	25.3	24	31
14.5	25.9	18	30	31.5

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

9. Weights of the NBA's Top 50 Players Listed are the weights of the NBA's top 50 players. Construct a grouped frequency distribution and a cumulative frequency distribution with 8 classes. Analyze the results in terms of peaks, extreme values, etc.

240	210	220	260	250	195	230	270	325	225
165	295	205	230	250	210	220	210	230	202
250	265	230	210	240	245	225	180	175	215
215	235	245	250	215	210	195	240	240	225
260	210	190	260	230	190	210	230	185	260

Source: www.msn.foxsports.com

10. Stories in the World's Tallest Buildings The number of stories in each of the world's 30 tallest buildings is listed below. 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.*

11. GRE Scores at Top-Ranked Engineering Schools The average quantitative GRE scores for the top 30 graduate schools of engineering are listed. Construct a grouped frequency distribution and a cumulative frequency distribution with 5 classes.

767	770	761	760	771	768	776	771	756	770
763	760	747	766	754	771	771	778	766	762
780	750	746	764	769	759	757	753	758	746

Source: *U.S. News & World Report Best Graduate Schools.*

12. Airline Passengers The number of passengers (in thousands) for the leading U.S. passenger airlines in 2004 is indicated below. Use the data to construct a grouped frequency distribution and a cumulative frequency distribution with a reasonable number of classes and comment on the shape of the distribution.

91,570	86,755	81,066	70,786	55,373	42,400
40,551	21,119	16,280	14,869	13,659	13,417
13,170	12,632	11,731	10,420	10,024	9,122
7,041	6,954	6,406	6,362	5,930	5,585
5,427					

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

13. 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. (The data for this exercise will be used for Exercise 5 in Section 2–2 and Exercise 23 in Section 3–1.)

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.*

14. Online Gambling Online computer gaming has become a popular leisure time activity. Fifty-six percent of the 117 million active gamers play games online. Below are listed the numbers of players playing a free online game at various times of the day. Construct a grouped frequency distribution and a cumulative frequency distribution with 6 classes.

3907	3629	3640	3839	3446	2313	2537	2037	3194
3739	3886	3698	3898	2101	1525	2311	3344	3647

Source: www.msn.tech.com

15. Presidential Vetoes The number of total vetoes exercised by the past 20 Presidents is listed below. Use the data to construct a grouped frequency distribution and a cumulative frequency distribution with 5 classes. What is challenging about this set of data?

44	39	37	21	31	170	44	632	30	78
42	6	250	43	44	82	50	181	66	37

Source: *World Almanac and Book of Facts.*

16. U.S. National Park Acreage The acreage of the 39 U.S. National Parks under 900,000 acres (in thousands of acres) is shown here. Construct a grouped frequency distribution and a cumulative frequency distribution for the data using 8 classes. (The data in this exercise will be used in Exercise 11 in Section 2–2.)

41	66	233	775	169
36	338	233	236	64
183	61	13	308	77
520	77	27	217	5
650	462	106	52	52
505	94	75	265	402
196	70	132	28	220
760	143	46	539	

Source: *The Universal Almanac.*

17. Heights of Alaskan Volcanoes The heights (in feet above sea level) of the major active volcanoes in Alaska are given here. Construct a grouped frequency distribution and a cumulative frequency distribution for the data using 10 classes. (The data in this exercise will be used in Exercise 9 in Section 3–1 and Exercise 17 in Section 3–2.)

4,265	3,545	4,025	7,050	11,413
3,490	5,370	4,885	5,030	6,830
4,450	5,775	3,945	7,545	8,450
3,995	10,140	6,050	10,265	6,965
150	8,185	7,295	2,015	5,055
5,315	2,945	6,720	3,465	1,980
2,560	4,450	2,759	9,430	
7,985	7,540	3,540	11,070	
5,710	885	8,960	7,015	

Source: *The Universal Almanac.*

18. Home Run Record Breakers During the 1998 baseball season, Mark McGwire and Sammy Sosa both broke Roger Maris's home run record of 61. The distances (in feet) for each home run follow. Construct a grouped frequency distribution and a cumulative frequency distribution for each player, using 8 classes. (The information in this exercise will be used for Exercise 12 in Section 2–2, Exercise 10 in Section 3–1, and Exercise 14 in Section 3–2.)

McGwire				Sosa			
306	370	370	430	371	350	430	420
420	340	460	410	430	434	370	420
440	410	380	360	440	410	420	460
350	527	380	550	400	430	410	370
478	420	390	420	370	410	380	340
425	370	480	390	350	420	410	415
430	388	423	410	430	380	380	366
360	410	450	350	500	380	390	400
450	430	461	430	364	430	450	440
470	440	400	390	365	420	350	420
510	430	450	452	400	380	380	400
420	380	470	398	370	420	360	368
409	385	369	460	430	433	388	440
390	510	500	450	414	482	364	370
470	430	458	380	400	405	433	390
430	341	385	410	480	480	434	344
420	380	400	440	410	420		
377	370						

Source: *USA TODAY.*

Extending the Concepts

19. 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?

Technology *Step by Step*

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		

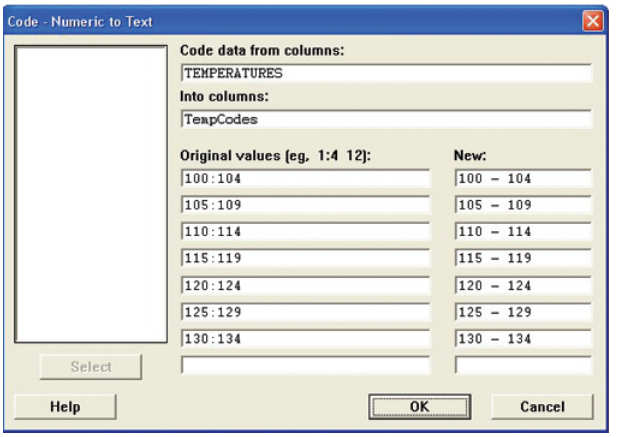
Make a Grouped Frequency Distribution (Quantitative Variable)

1. Select **File>New>New 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.

   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>Code>Numeric to Text.**
   a) The cursor should be blinking in Code data from 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 in the Into columns: then type the name of the new column, **TempCodes.**
   c) Press [Tab] to move to the next dialog box.
   d) Type in the first interval **100:104.**

      Use a colon to indicate the interval from 100 to 104 with no spaces before or after the colon.
   e) Press [Tab] to move to the New: column, and type the text category **100–104.**
   f) Continue to tab to each dialog box, typing the interval and then the category until the last category has been entered.

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 TempCodes 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: TempCodes

TempCodes	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		

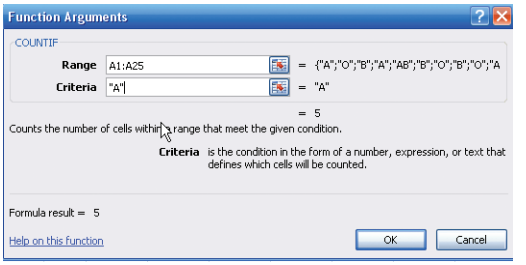
7. Click **File>Save Project As** . . . , and type the name of the project file, **Ch2-2.** This will save the two worksheets and the Session Window.

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.

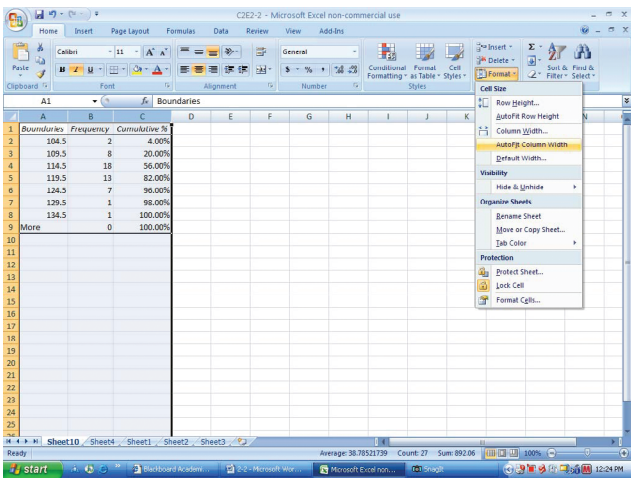
Making a Grouped Frequency Distribution (Quantitative Data)

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.

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).

An example of each type of graph is shown in Figure 2–1. The data for each graph are the distribution of the miles that 20 randomly selected runners ran during a given week.

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).

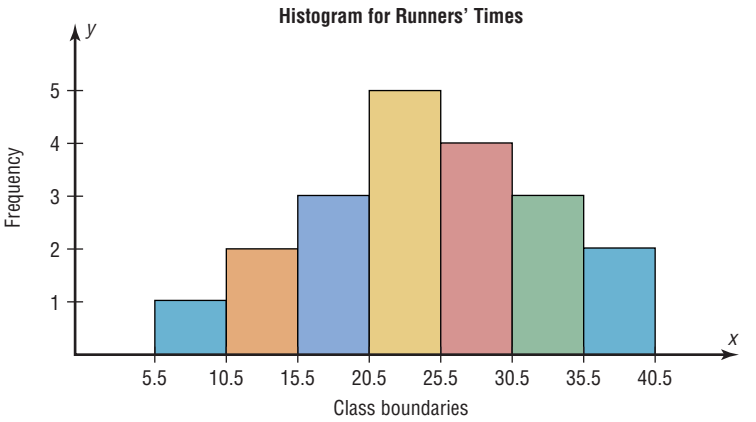
Class boundaries	Frequency
99.5–104.5	2
104.5–109.5	8
109.5–114.5	18
114.5–119.5	13
119.5–124.5	7
124.5–129.5	1
129.5–134.5	1

Solution

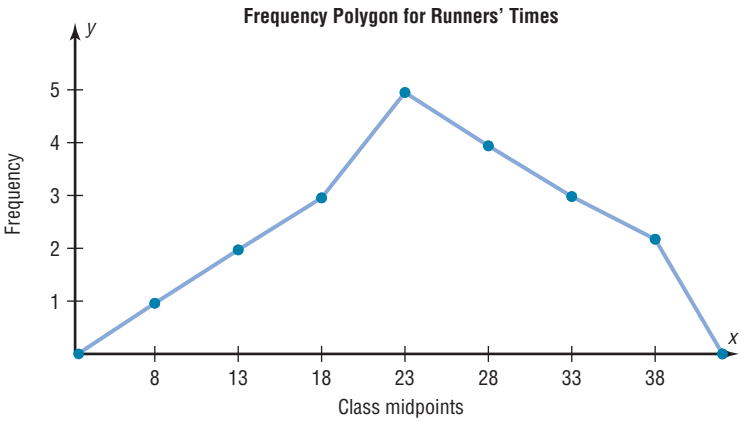
Step 1 Draw and label the x and y axes. The x axis is always the horizontal axis, and the y axis is always the vertical axis.

Figure 2–1

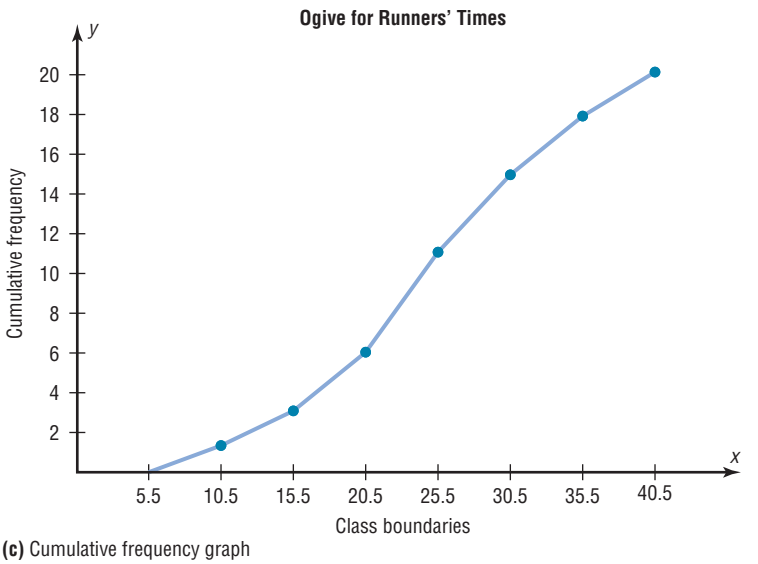
Examples of Commonly Used Graphs



(a) Histogram

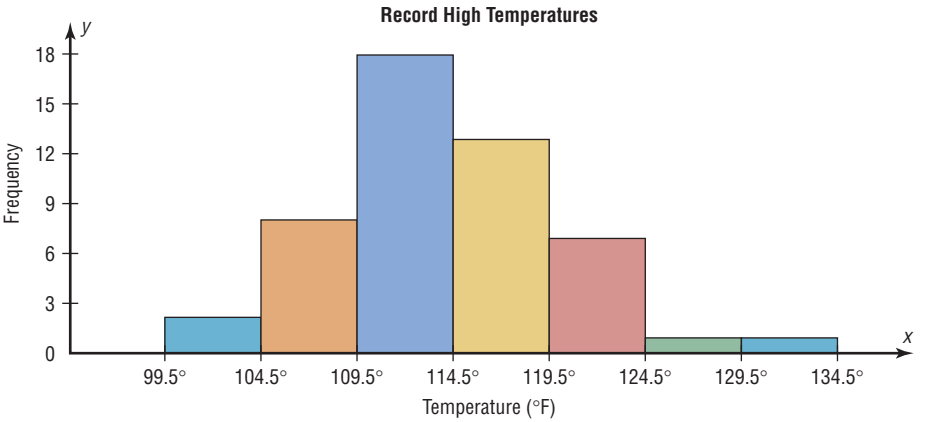


(b) Frequency polygon



(c) Cumulative frequency graph

Figure 2–2
Histogram for Example 2–4



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 (1748–1819), an engineer and drafter who used graphs to present economic data pictorially.

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 2–2.

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.

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.

Example 2–5

Record High Temperatures

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

Solution

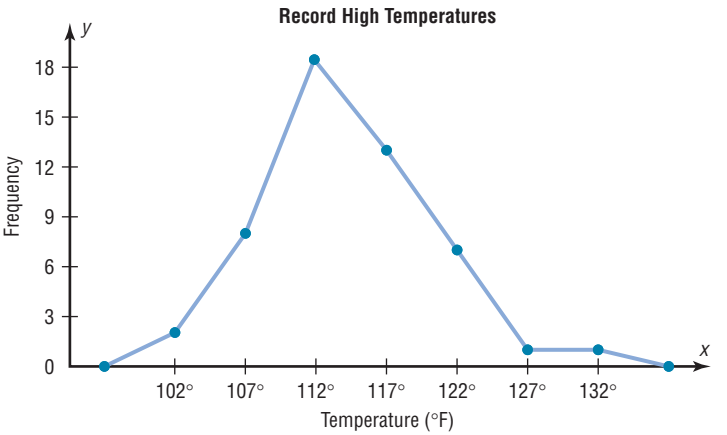
Step 1 Find the midpoints of each class. Recall that midpoints are found by adding the upper and lower boundaries and dividing by 2:

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

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

Figure 2–3
Frequency Polygon for Example 2–5



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 2–3.

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.

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

Figure 2–4
Plotting the Cumulative Frequency for Example 2–6

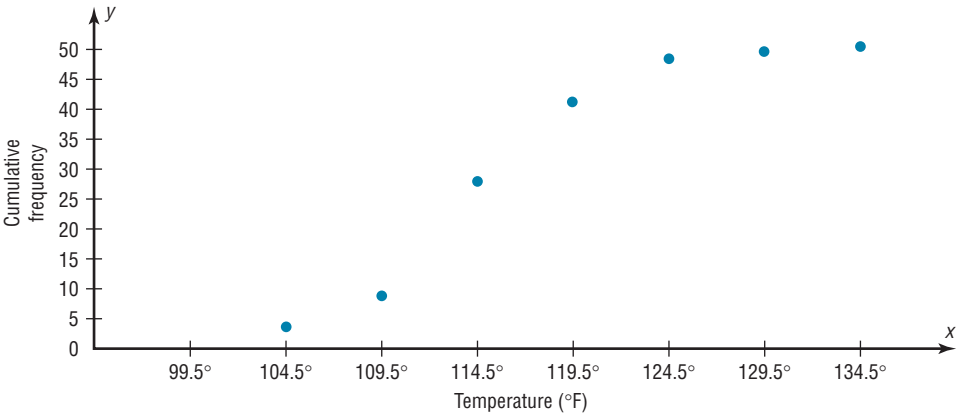
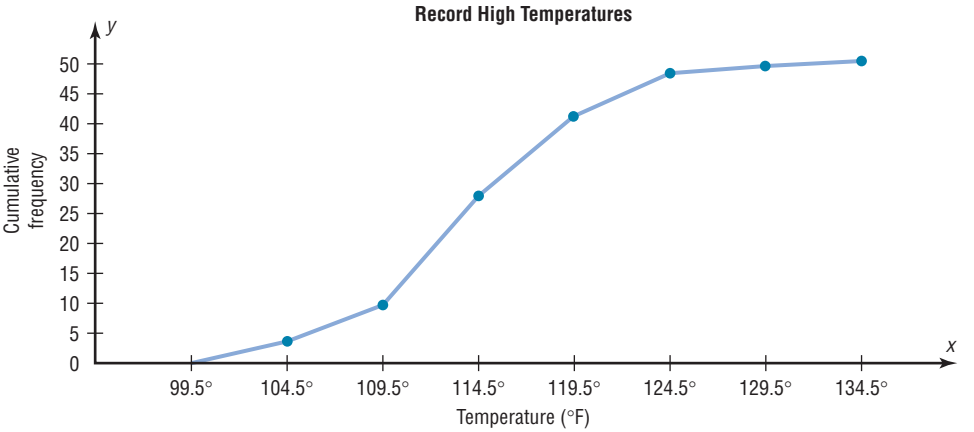


Figure 2–5
Ogive for Example 2–6



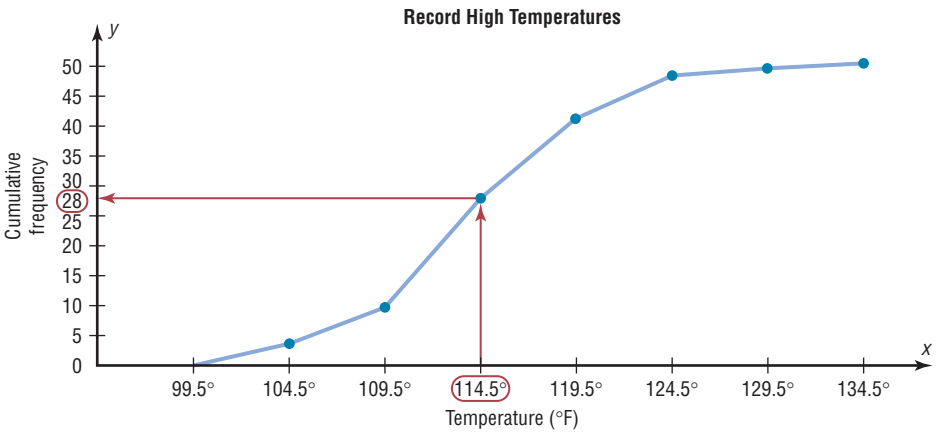
Step 2 Draw the *x* and *y* axes. Label the *x* axis with the class boundaries. Use an appropriate scale for the *y* axis to represent the cumulative frequencies. (Depending on the numbers in the cumulative frequency columns, scales such as 0, 1, 2, 3, . . . , or 5, 10, 15, 20, . . . , or 1000, 2000, 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 2–4. 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 2–5. Then extend the graph to the first lower class boundary, 99.5, on the *x* axis.

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 2–6.

Figure 2–6
Finding a Specific Cumulative Frequency



The steps for drawing these three types of graphs are shown in the following Procedure Table.

Unusual Stat
Twenty-two percent of Americans sleep 6 hours a day or fewer.

Procedure Table

Constructing Statistical Graphs

Step 1 Draw and label the *x* and *y* axes.

Step 2 Choose a suitable scale for the frequencies or cumulative frequencies, and label it on the *y* axis.

Step 3 Represent the class boundaries for the histogram or ogive, or the midpoint for the frequency polygon, on the *x* axis.

Step 4 Plot the points and then draw the bars or lines.

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 1,478,002 and the population of Erie is 105,270, 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 Miles Run per Week

Construct a histogram, frequency polygon, and ogive using relative frequencies for the distribution (shown here) of the miles that 20 randomly selected runners ran during a given week.

Class boundaries	Frequency
5.5–10.5	1
10.5–15.5	2
15.5–20.5	3
20.5–25.5	5
25.5–30.5	4
30.5–35.5	3
35.5–40.5	2
	20

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 class 5.5–10.5, the relative frequency is $\frac{1}{20} = 0.05$; for class 10.5–15.5, the relative frequency is $\frac{2}{20} = 0.10$; for class 15.5–20.5, the relative frequency is $\frac{3}{20} = 0.15$; and so on.

Place these values in the column labeled Relative frequency.

Class boundaries	Midpoints	Relative frequency
5.5–10.5	8	0.05
10.5–15.5	13	0.10
15.5–20.5	18	0.15
20.5–25.5	23	0.25
25.5–30.5	28	0.20
30.5–35.5	33	0.15
35.5–40.5	38	0.10
		1.00

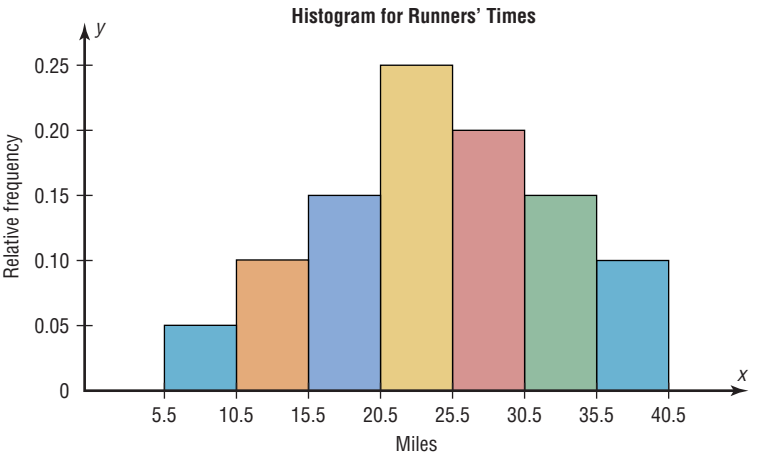
Step 2 Find the cumulative relative frequencies. To do this, add the frequency in each class to the total frequency of the preceding class. In this case, $0 + 0.05 = 0.05$, $0.05 + 0.10 = 0.15$, $0.15 + 0.15 = 0.30$, $0.30 + 0.25 = 0.55$, etc. Place these values in the column labeled Cumulative relative frequency.

An alternative method would be to find the cumulative frequencies and then convert each one to a relative frequency.

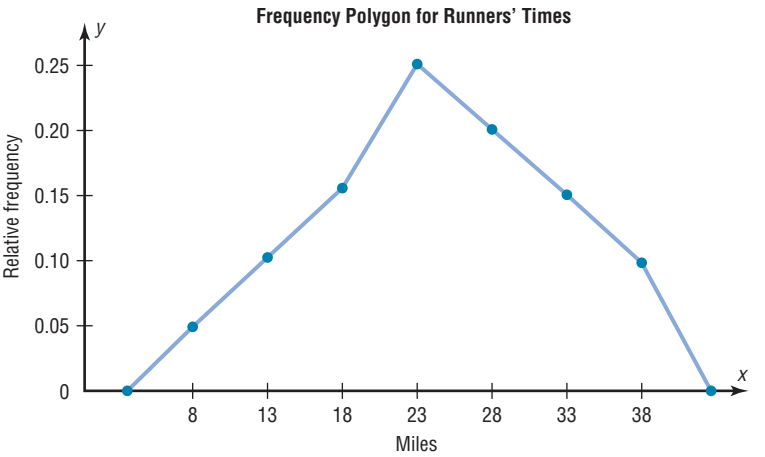
	Cumulative frequency	Cumulative relative frequency
Less than 5.5	0	0.00
Less than 10.5	1	0.05
Less than 15.5	3	0.15
Less than 20.5	6	0.30
Less than 25.5	11	0.55
Less than 30.5	15	0.75
Less than 35.5	18	0.90
Less than 40.5	20	1.00

Step 3 Draw each graph as shown in Figure 2–7. For the histogram and ogive, use the class boundaries along the *x* axis. For the frequency polygon, use the midpoints on the *x* axis. The scale on the *y* axis uses proportions.

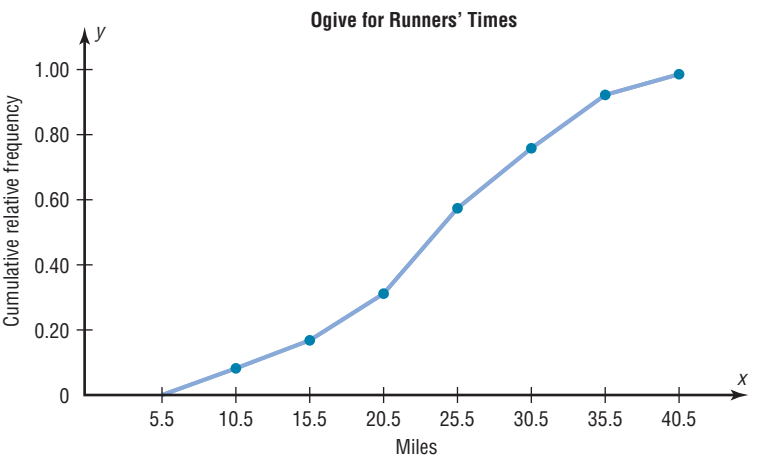
Figure 2–7
Graphs for Example 2–7



(a) Histogram



(b) Frequency polygon



(c) Ogive

Distribution Shapes

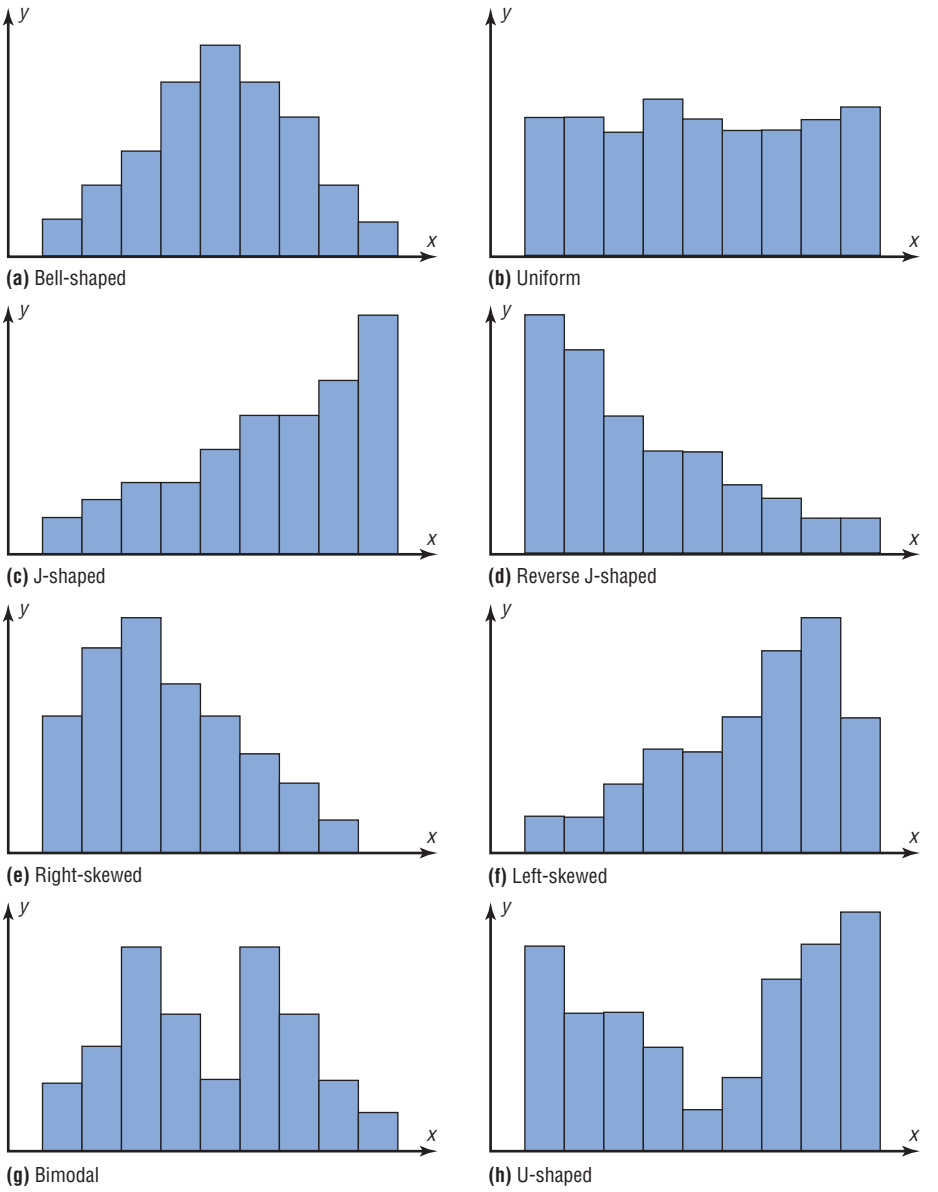
When one is describing data, it is important to be able to recognize the shapes of the distribution values. In later chapters you will see that the shape of a distribution also determines the appropriate statistical methods used to analyze the data.

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 2–8: *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 2–8(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.

Figure 2–8
Distribution Shapes



A *uniform distribution* is basically flat or rectangular. See Figure 2–8(b).

A *J-shaped distribution* is shown in Figure 2–8(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 2–8(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 2–8(e). When 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 2–8(f). Skewness will be explained in detail in Chapter 3. Distributions with one peak, such as those shown in Figure 2–8(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 2–8(g). Finally, the graph shown in Figure 2–8(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 shown in Figure 2–2 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 that 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 101 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

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. Number of College Faculty The number of faculty listed for a variety of private colleges which 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.*

3. Counties, Divisions, or Parishes for 50 States The number of counties, divisions, or parishes for each of the 50 states is given below. Use the data to construct a grouped frequency distribution with 6 classes, a histogram, a frequency polygon, and an ogive. Analyze the distribution.

67	27	15	75	58	64	8	67	159	5
102	44	92	99	105	120	64	16	23	14
83	87	82	114	56	93	16	10	21	33
62	100	53	88	77	36	67	5	46	66
95	254	29	14	95	39	55	72	23	3

Source: *World Almanac and Book of Facts.*

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

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

Source: NFL.com

Construct a histogram, a frequency polygon, and an ogive for the data; and comment on the shape of the distribution.

5. Automobile Fuel Efficiency Thirty automobiles were tested for fuel efficiency, in miles per gallon (mpg). The following frequency distribution was obtained. Construct a histogram, a frequency polygon, and an ogive for the data.

Class boundaries	Frequency
7.5–12.5	3
12.5–17.5	5
17.5–22.5	15
22.5–27.5	5
27.5–32.5	2

6. Construct a frequency histogram, a frequency polygon, and an ogive for the data in Exercise 9 in Section 2–1. Analyze the results.

7. Air Quality Standards The number of days that selected U.S. metropolitan areas failed to meet acceptable air quality standards is shown below for 1998 and 2003. Construct grouped frequency distributions and a histogram for each set of data, and compare your results.

1998						2003					
43	76	51	14	0	10	10	11	14	20	15	6
20	0	5	17	67	25	17	0	5	19	127	4
38	0	56	8	0	9	31	5	88	1	1	16
14	5	37	14	95	20	14	19	20	9	138	22
23	12	33	0	3	45	13	10	20	20	20	12

Source: *World Almanac.*

8. How Quick Are Dogs? In a study of reaction times of dogs to a specific stimulus, an animal trainer obtained the following data, given in seconds. Construct a histogram, a frequency polygon, and an ogive for the data; analyze the results. (The histogram in this exercise will be used for Exercise 18 in this section, Exercise 16 in Section 3–1, and Exercise 26 in Section 3–2.)

Class limits	Frequency
2.3–2.9	10
3.0–3.6	12
3.7–4.3	6
4.4–5.0	8
5.1–5.7	4
5.8–6.4	2

9. Quality of Health Care The scores of health care quality as calculated by a professional risk management company are listed on the next page for selected states.

Use the data to construct a frequency distribution, a histogram, a frequency polygon, and an ogive.

118.2 114.6 113.1 111.9 110.0 108.8 108.3 107.7 107.0 106.7
105.3 103.7 103.2 102.8 101.6 99.8 98.1 96.6 95.7 93.6
92.5 91.0 90.0 87.1 83.1

Source: *New York Times Almanac.*

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

Source: *National Center for Educational Statistics.*

11. Construct a histogram, a frequency polygon, and an ogive for the data in Exercise 16 in Section 2–1, and analyze the results.

12. For the data in Exercise 18 in Section 2–1, construct a histogram for the home run distances for each player and compare them. Are they basically the same, or are there any noticeable differences? Explain your answer.

13. For the data in Exercise 1 in this section, construct a histogram, a frequency polygon, and an ogive, using relative frequencies. What proportion of the applicants needs to enroll in the summer development program?

14. For the data for 2003 in Exercise 4 in this section, construct a histogram, a frequency polygon, and an ogive, using relative frequencies.

15. Cereal Calories The number of calories per serving for selected ready-to-eat cereals is listed here. Construct a frequency distribution using 7 classes. Draw a histogram, a frequency polygon, and an ogive for the data, using relative frequencies. Describe the shape of the histogram.

130 190 140 80 100 120 220 220 110 100
210 130 100 90 210 120 200 120 180 120
190 210 120 200 130 180 260 270 100 160
190 240 80 120 90 190 200 210 190 180
115 210 110 225 190 130

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

16. 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.*

17. For the data for year 2003 in Exercise 7 in this section, construct a histogram, a frequency polygon, and an ogive, using relative frequencies.

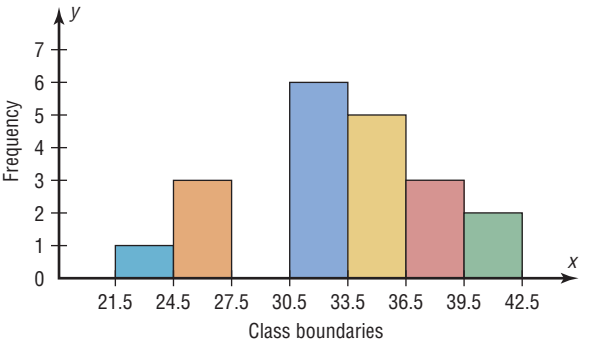
18. How Quick Are Older Dogs? The animal trainer in Exercise 8 in this section selected another group of dogs who were much older than the first group and measured their reaction times to the same stimulus. Construct a histogram, a frequency polygon, and an ogive for the data.

Class limits	Frequency
2.3–2.9	1
3.0–3.6	3
3.7–4.3	4
4.4–5.0	16
5.1–5.7	14
5.8–6.4	4

Analyze the results and compare the histogram for this group with the one obtained in Exercise 8 in this section. Are there any differences in the histograms? (The data in this exercise will be used for Exercise 16 in Section 3–1 and Exercise 26 in Section 3–2.)

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?

Technology *Step by Step*

MINITAB
Step by Step

Construct a Histogram

1. Enter the data from Example 2–2, the high temperatures for the 50 states.

2. Select **Graph>Histogram.**

3. Select [Simple], then click [OK].

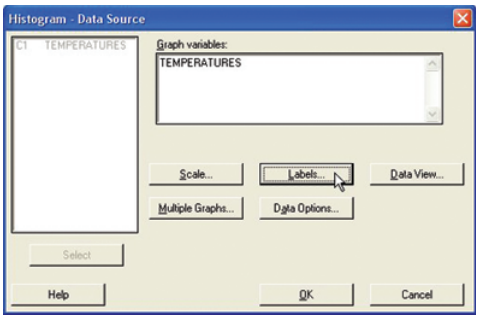
4. Click C1 TEMPERATURES in the Graph variables dialog box.

5. Click [Labels]. There are two tabs, Title/Footnote and Data Labels.

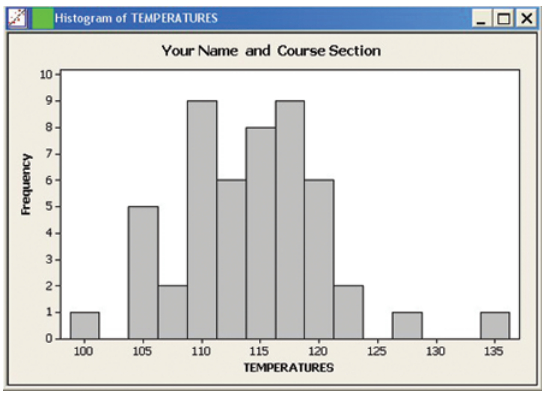
a) Click in the box for Title, and type in Your Name and Course Section.

b) Click [OK]. The Histogram dialog box is still open.

6. Click [OK]. A new graph window containing the histogram will open.



7. Click the **File** menu to print or save the graph.



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

TI-83 Plus or TI-84 Plus

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.

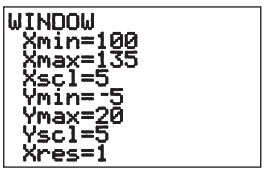
To change the values in the WINDOW:

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.

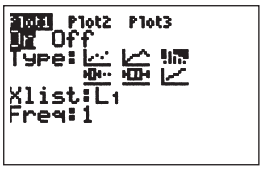
To plot the histogram from raw data:

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 on, if necessary.
5. Move cursor to the Histogram symbol and press **ENTER,** if necessary.
6. Make sure Xlist is L_1.
7. Make sure Freq is 1.
8. Press **GRAPH** to display the histogram.
9. To obtain the number of data values in each class, press the **TRACE** key, followed by ◄ or ► keys.

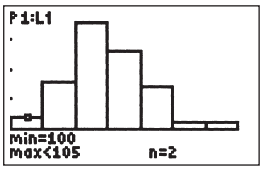
Input



Input



Output



Example TI2–1

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

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:

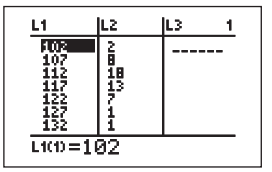
1. Enter the midpoints into L_1.
2. Enter the frequencies into L_2.
3. Make sure WINDOW values are appropriate for the histogram.
4. Press **[2nd] [STAT PLOT] ENTER.**
5. Press **ENTER** to turn the plot on, if necessary.
6. Move cursor to the histogram symbol, and press **ENTER,** if necessary.
7. Make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

Example TI2–2

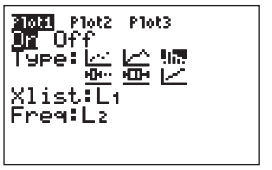
Plot a histogram for the data from Examples 2–4 and 2–5.

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

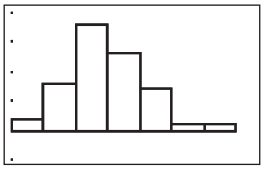
Input



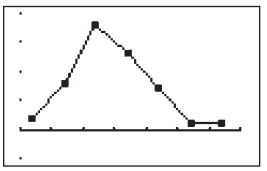
Input



Output

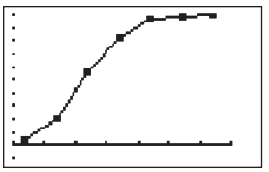


Output



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



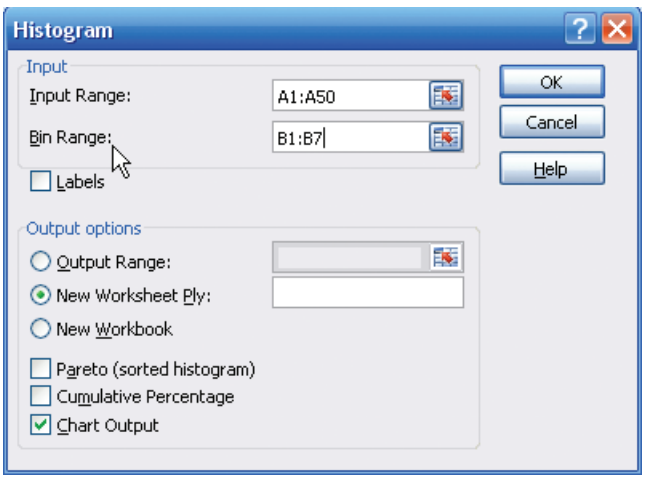
To graph an ogive from grouped data, modify the procedure for the histogram as follows:

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.

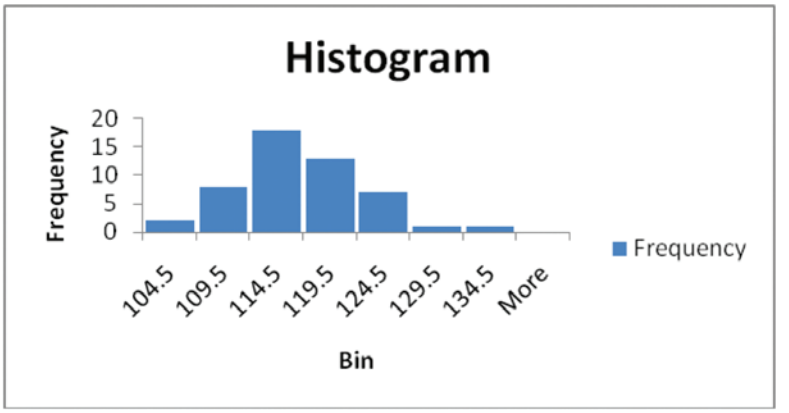
Excel
Step by Step

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].



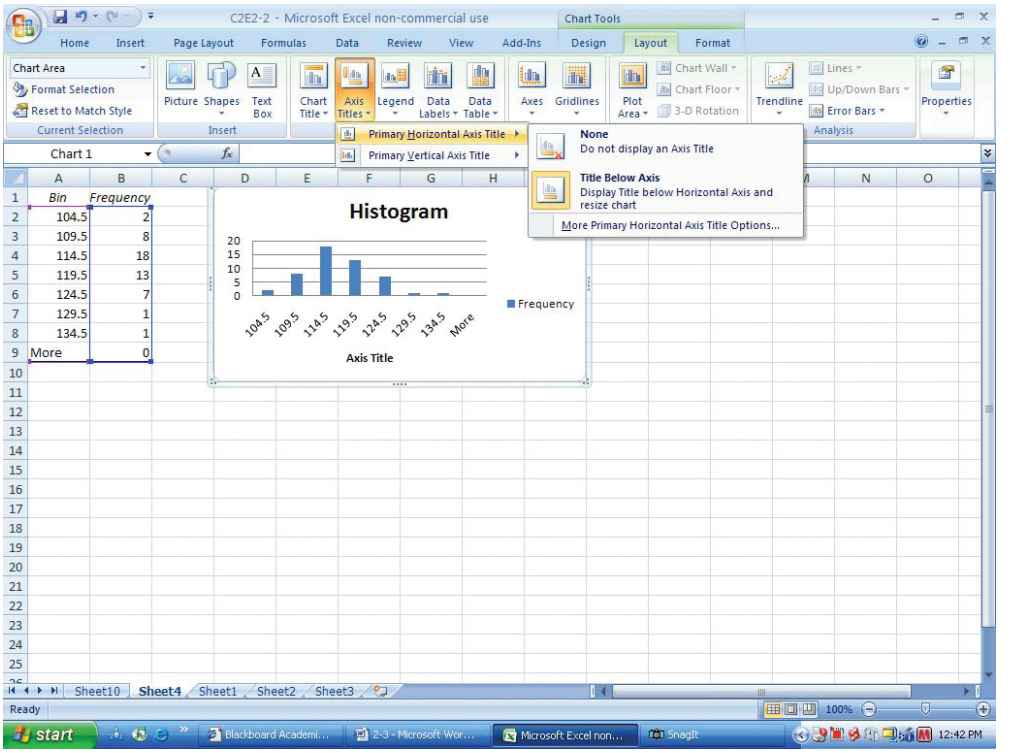
Editing the Histogram

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 bar to the left to narrow the distance between bars.

To change the label for the horizontal axis:

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



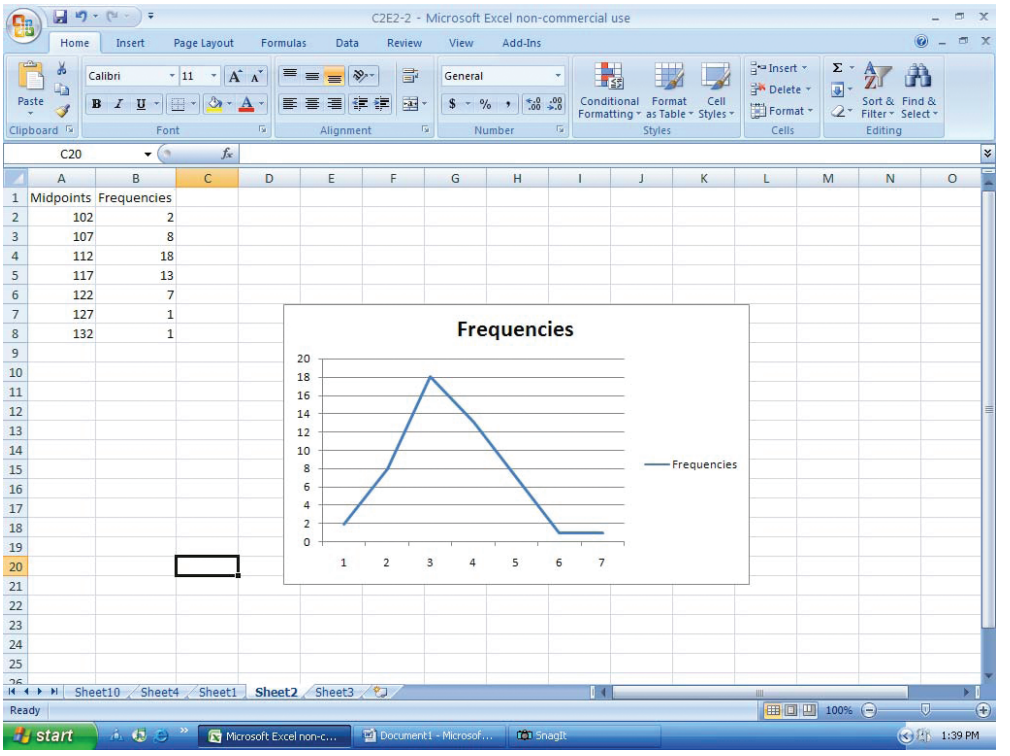
Once the Axis Title text box is selected, you can type in the name of the variable represented on the horizontal axis.

Constructing a Frequency Polygon

1. Press **[Ctrl]-N** for a new workbook.
2. Enter the midpoints of the data from Example 2–2 into column A. Enter the frequencies into column B.

	A	B
1	Midpoints	Frequencies
2	102	2
3	107	8
4	112	18
5	117	13
6	122	7
7	127	1
8	132	1

3. Highlight 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 and the frequencies are on the vertical axis.

1. Right-click the mouse on any region of the graph.
2. Select the Select Data option.

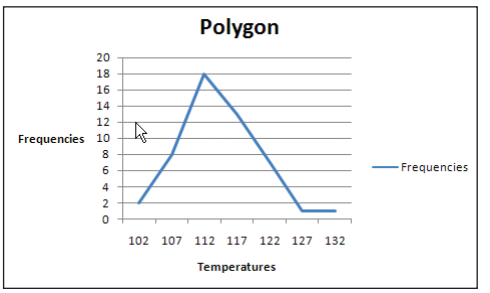
3. Select Edit from the Horizontal Axis Labels and highlight the midpoints from column A, then click [OK].
4. Click [OK] on the Select Data Source box.

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.

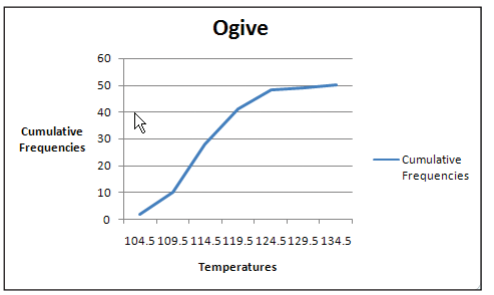


Constructing an Ogive

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

1. Type the upper class boundaries and cumulative frequencies into adjacent columns of an Excel worksheet.
2. Highlight the cumulative frequencies (including the label) and select the Insert tab from the toolbar.
3. Select Line Chart, then the 2-D Line option.

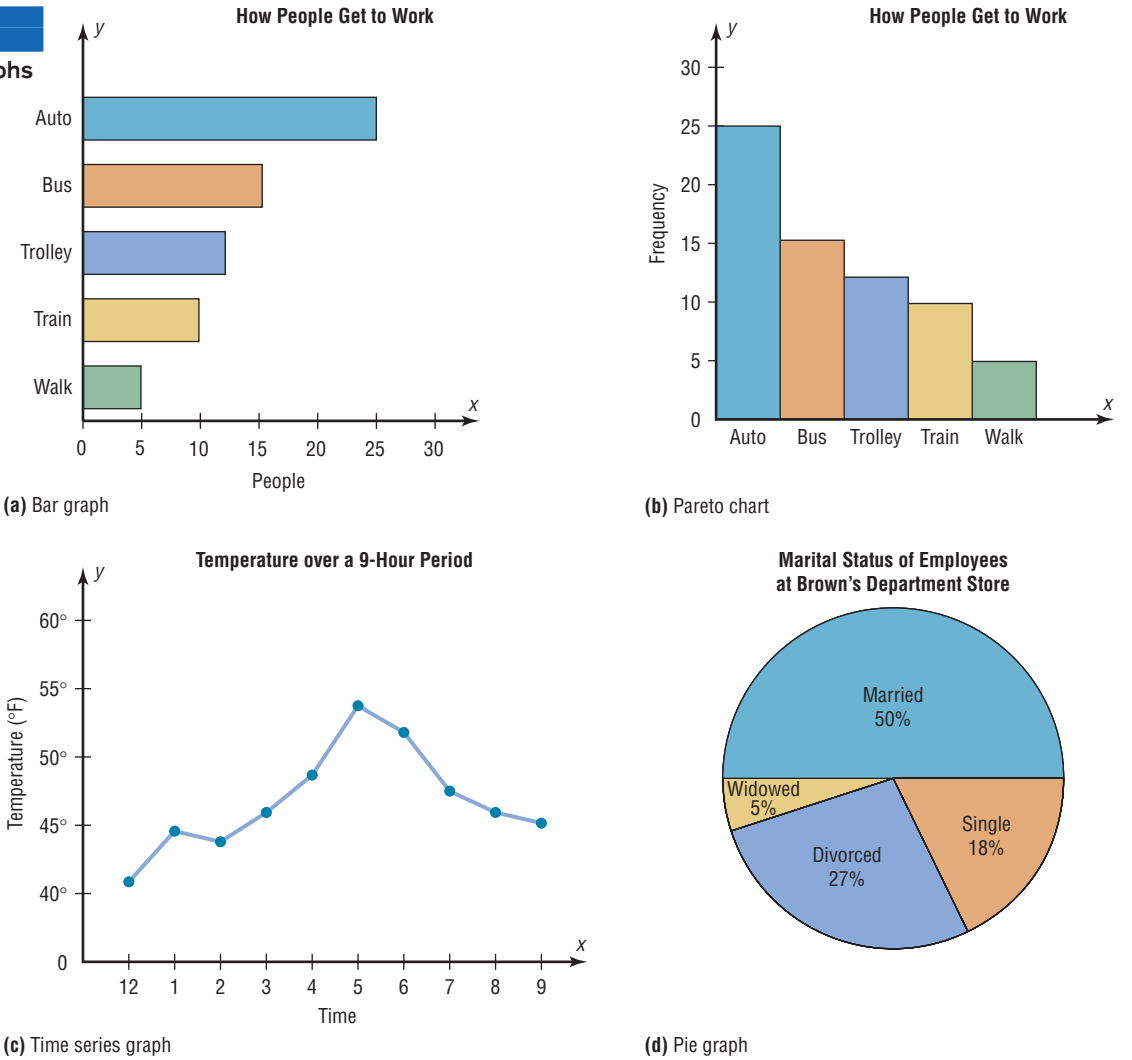
As with the frequency polygon, you can insert labels on the axes and a chart title for the ogive.



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, and pie graph. Figure 2–9 shows an example of each type of graph.

Figure 2–9
Other Types of Graphs Used in Statistics



(a) Bar graph

(b) Pareto chart

(c) Time series graph

(d) Pie graph

Objective 3

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

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.

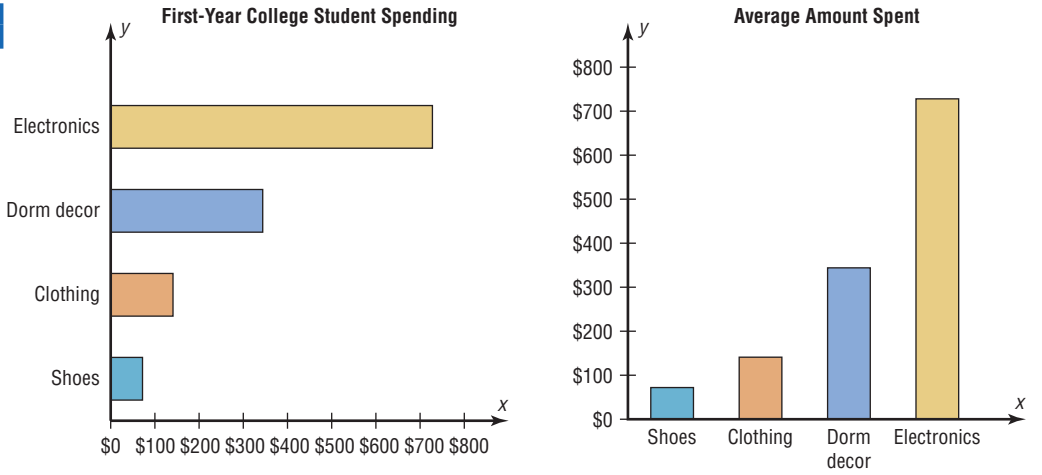
Electronics	$728
Dorm decor	344
Clothing	141
Shoes	72

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 2–10.

Figure 2–10
Bar Graphs for Example 2–8



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

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

Turnpike Costs

The table shown here is the average cost per mile for passenger vehicles on state turnpikes. Construct and analyze a Pareto chart for the data.

State	Number
Indiana	2.9¢
Oklahoma	4.3
Florida	6.0
Maine	3.8
Pennsylvania	5.8

Source: *Pittsburgh Tribune Review.*

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.

Solution

Step 1 Arrange the data from the largest to smallest according to frequency.

State	Number
Florida	6.0¢
Pennsylvania	5.8
Oklahoma	4.3
Maine	3.8
Indiana	2.9

Step 2 Draw and label the x and y axes.

Step 3 Draw the bars corresponding to the frequencies. See Figure 2–11. The Pareto chart shows that Florida has the highest cost per mile. The cost is more than twice as high as the cost for Indiana.

Suggestions for Drawing Pareto Charts

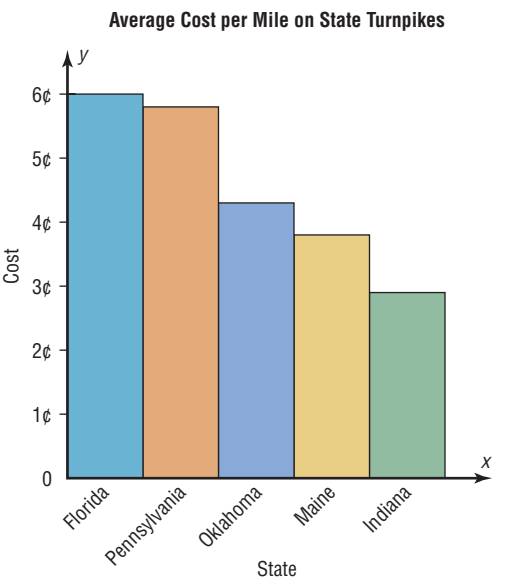
1. Make the bars the same width.
2. Arrange the data from largest to smallest according to frequency.
3. Make the units that are used for the frequency equal in size.

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.

Figure 2–11
Pareto Chart for Example 2–9



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

Arson Damage to Churches

The arson damage to churches for the years 2001 through 2005 is shown. Construct and analyze a time series graph for the data.

Year	Damage (in millions)
2001	$2.8
2002	3.3
2003	3.4
2004	5.0
2005	8.5

Source: U.S. Fire Administration.

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

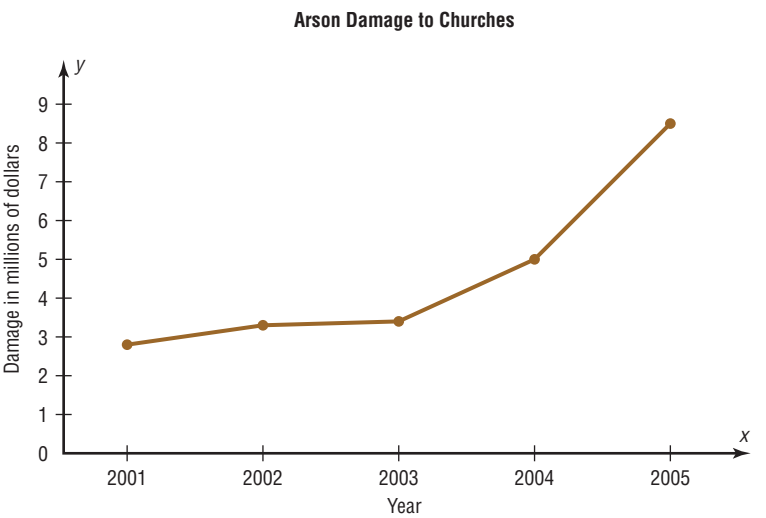
Step 1 Draw and label the x and y axes.

Step 2 Label the x axis for years and the y axis for the damage.

Step 3 Plot each point according to the table.

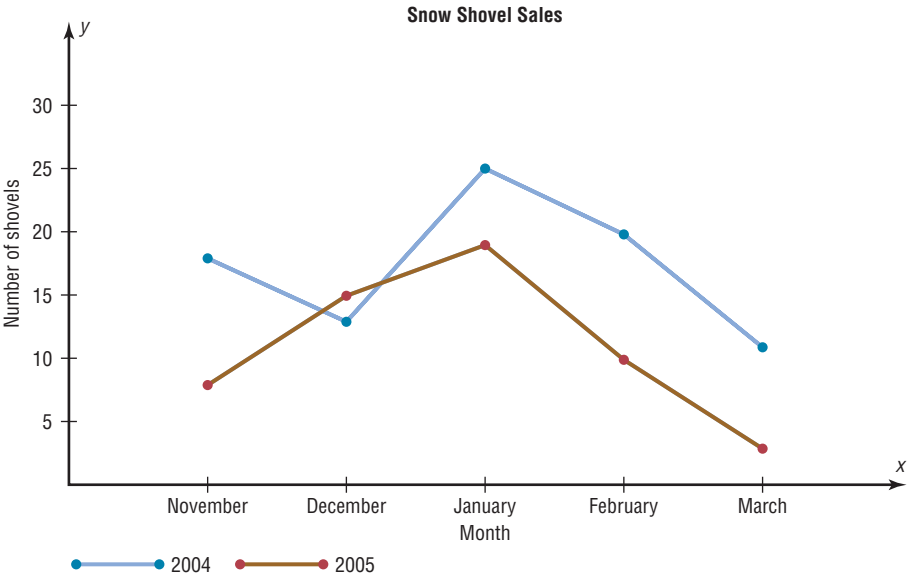
Step 4 Draw line segments connecting adjacent points. Do not try to fit a smooth curve through the data points. See Figure 2–12. The graph shows a steady increase over the 5-year period.

Figure 2–12
Time Series Graph 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.

Figure 2–13
Two Time Series Graphs for Comparison



Two data sets can be compared on the same graph (called a *compound time series graph*) if two lines are used, as shown in Figure 2–13. This graph shows the number of snow shovels sold at a store for two seasons.

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.

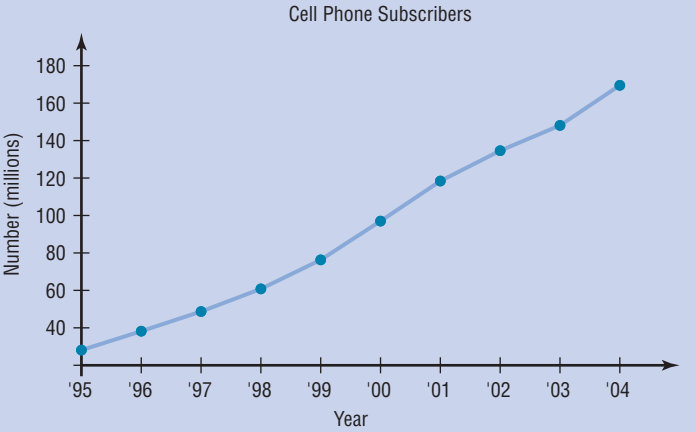
Snack	Pounds (frequency)
Potato chips	11.2 million
Tortilla chips	8.2 million
Pretzels	4.3 million
Popcorn	3.8 million
Snack nuts	2.5 million
Total	$n = 30.0$ million

Source: *USA TODAY Weekend.*

Speaking of Statistics

Cell Phone Usage

The graph shows the estimated number (in millions) of cell phone subscribers since 1995. How do you think the growth will affect our way of living? For example, emergencies can be handled faster since people are using their cell phones to call 911.



Source: Cellular Telecommunications and Internet Association.

Solution

Step 1 Since there are 360° in a circle, the frequency for each class must be converted into 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°.*

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	$\overline{360°}$

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%.†

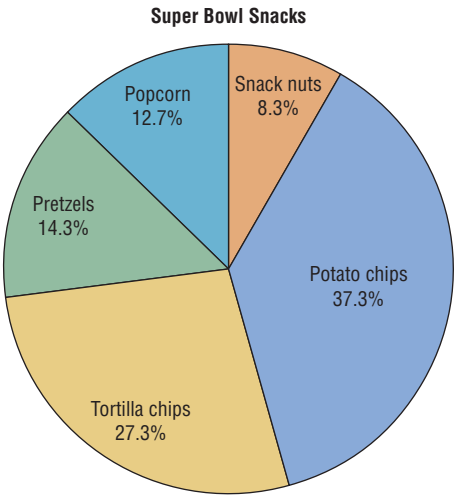
Potato chips	$\frac{11.2}{30} \cdot 100\% = 37.3\%$
Tortilla chips	$\frac{8.2}{30} \cdot 100\% = 27.3\%$

**Note:* The degrees column does not always sum to 360° due to rounding.
†*Note:* The percent column does not always sum to 100% due to rounding.

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%

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.

Figure 2–14
Pie Graph for Example 2–11



Example 2–12

Construct a pie graph showing the blood types of the army inductees described in Example 2–1. The frequency distribution is repeated here.

Class	Frequency	Percent
A	5	20
B	7	28
O	9	36
AB	4	16
	25	100

Solution

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

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

For each class, then, the following results are obtained.

A	$\frac{5}{25} \cdot 360^\circ = 72^\circ$
B	$\frac{7}{25} \cdot 360^\circ = 100.8^\circ$

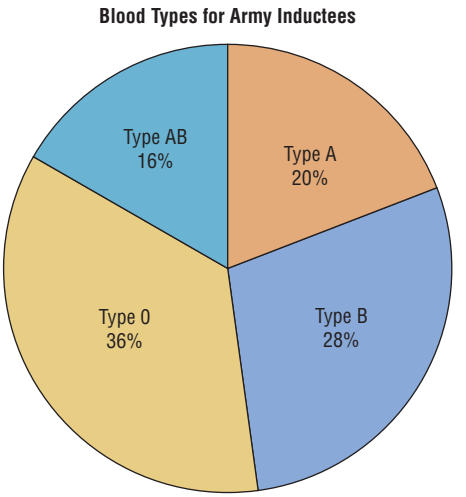
$$\text{O} \quad \frac{9}{25} \cdot 360^\circ = 129.6^\circ$$

$$\text{AB} \quad \frac{4}{25} \cdot 360^\circ = 57.6^\circ$$

Step 2 Find the percentages. (This was already done in Example 2–1.)

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

Figure 2–15
Pie Graph for Example 2–12



The graph in Figure 2–15 shows that in this case the most common blood type is type O.

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 among the inductees, type O blood is more prevalent than any other type. People who have type AB blood are in the minority. More than twice as many people have type O blood as type AB.

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 2–16. 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 2–16. 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 2–17, 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 2–16
Graph of Automaker's Claim Using a Scale from 95 to 100%

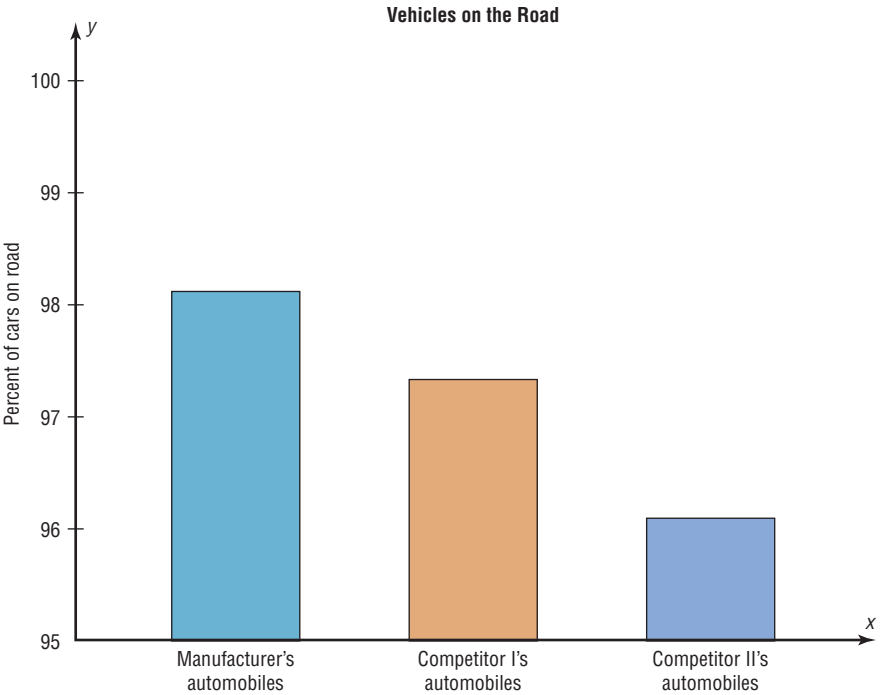
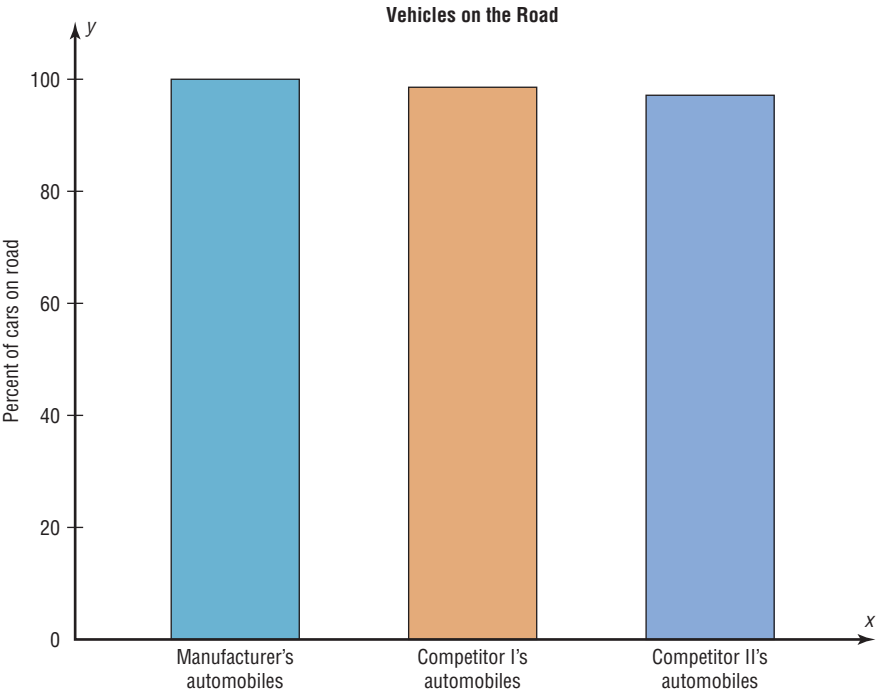


Figure 2–17
Graph in Figure 2–16 Redrawn Using a Scale from 0 to 100%



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.

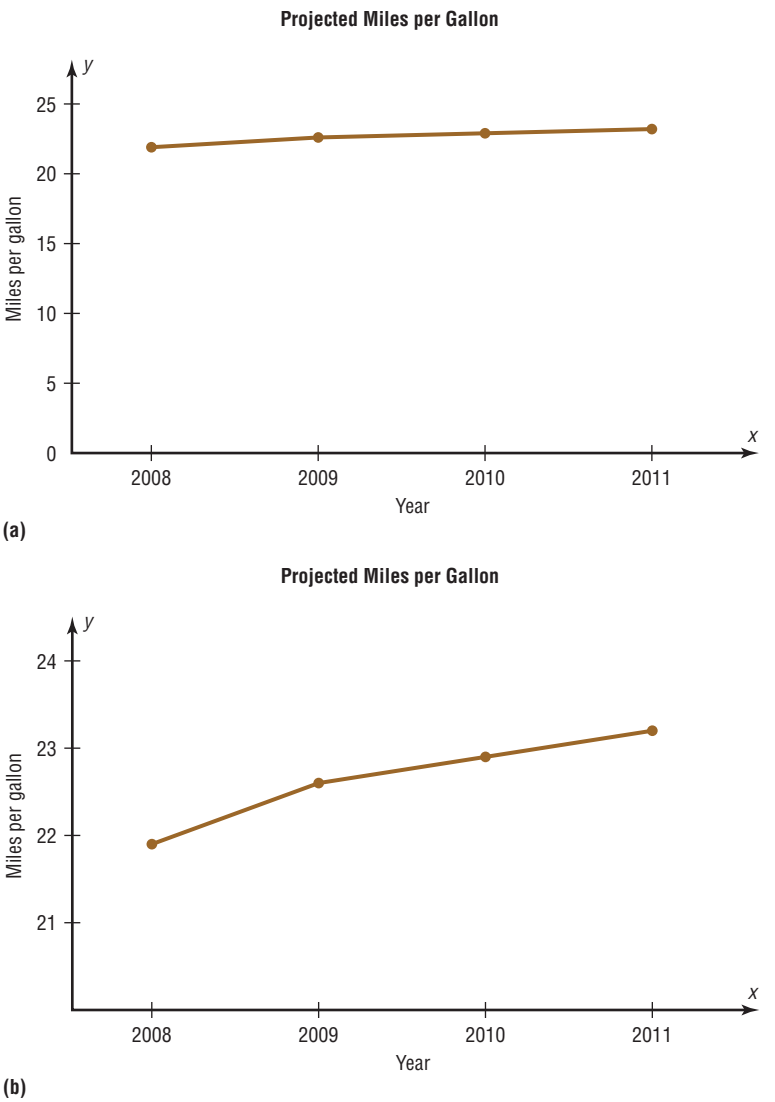
Let us consider another example. The projected required fuel economy in miles per gallon for General Motors vehicles is shown. In this case, an increase from 21.9 to 23.2 miles per gallon is projected.

Year	2008	2009	2010	2011
Projected MPG	21.9	22.6	22.9	23.2

Source: National Highway Traffic Safety Administration.

When you examine the graph shown in Figure 2–18(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

Figure 2–18
Projected Miles per Gallon



to 24 miles per gallon, the graph shows a much larger increase even though the data remain the same. See Figure 2–18(b). Again, by changing the units or starting point on the *y* axis, one can change the visual representation.

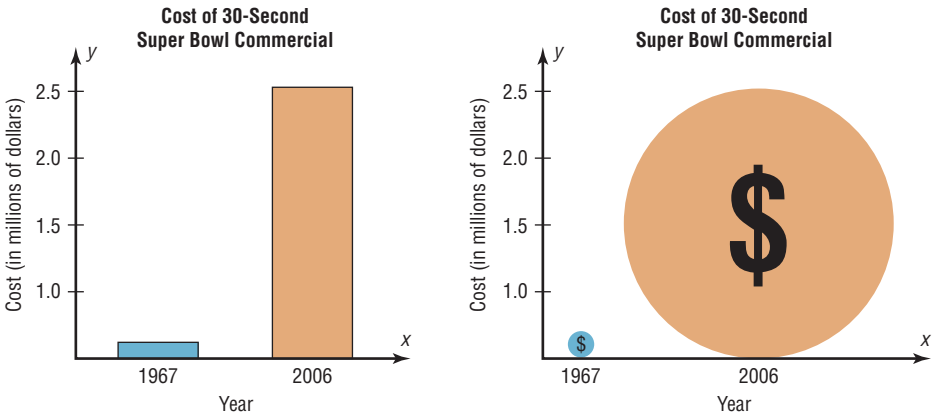
Another misleading graphing technique sometimes used involves exaggerating a one-dimensional increase by showing it in two dimensions. For example, the average cost of a 30-second Super Bowl commercial has increased from $42,000 in 1967 to $2.5 million in 2006 (Source: *USA TODAY*).

The increase shown by the graph in Figure 2–19(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 2–19(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.

Figure 2–19

Comparison of Costs for a 30-Second Super Bowl Commercial



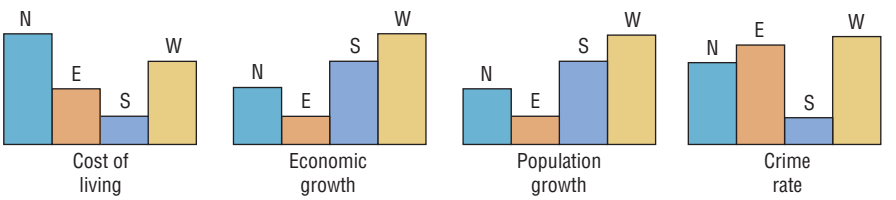
(a) Graph using bars

(b) Graph using circles

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 2–20 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.

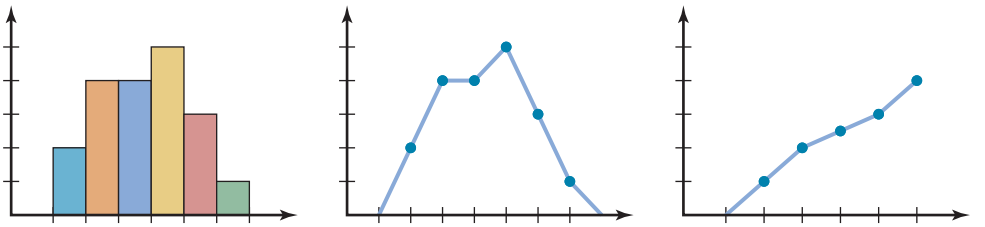
Figure 2–20

A Graph with No Units on the *y* Axis

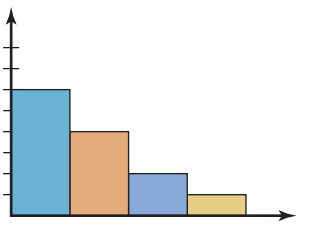


Finally, all graphs should contain a source for the information presented. The inclusion of a source for the data will enable you to check the reliability of the organization presenting the data. A summary of the types of graphs and their uses is shown in Figure 2–21.

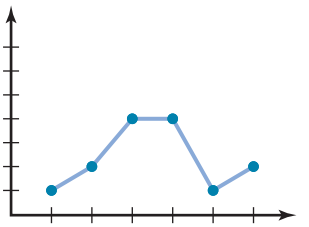
Figure 2–21
Summary of Graphs and Uses of Each



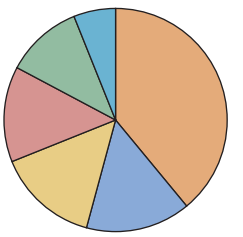
(a) Histogram; frequency polygon; ogive
Used when the data are contained in a grouped frequency distribution.



(b) Pareto chart
Used to show frequencies for nominal or qualitative variables.



(c) Time series graph
Used to show a pattern or trend that occurs over a period of time.



(d) Pie graph
Used to show the relationship between the parts and the whole. (Most often uses percentages.)

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.

Example 2–13 shows the procedure for constructing a stem and leaf plot.

Example 2–13

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

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 $1 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?

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 2–22.

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 2–22
Stem and Leaf Plot for Example 2–13

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 2–22 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.

Example 2–14

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

Figure 2–23
Stem and Leaf Plot for Example 2–14

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 2–23.

When the data values are in the hundreds, such as 325, the stem is 32 and the leaf is 5. For example, the stem and leaf plot for the data values 325, 327, 330, 332, 335, 341, 345, and 347 looks like this.

32	5 7
33	0 2 5
34	1 5 7

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.

Interesting Fact

The average number of pencils and index cards David Letterman tosses over his shoulder during one show is 4.

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 2–15 shows a back-to-back stem and leaf plot.

Example 2–15

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 2–24.

Figure 2–24
Back-to-Back Stem and Leaf Plot for Example 2–15

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	

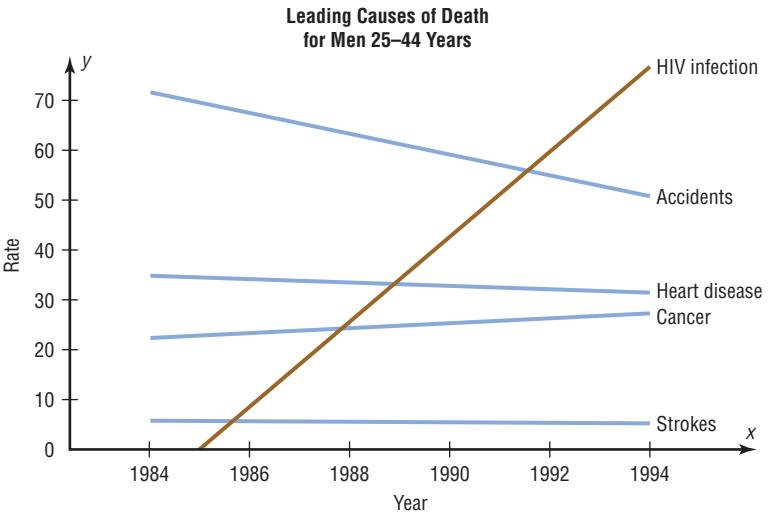
Step 3 Compare the distributions. The buildings in Atlanta have a large variation in the number of stories per building. Although both distributions are peaked in the 30- to 39-story class, Philadelphia has more buildings in this class. Atlanta has more buildings that have 40 or more stories than Philadelphia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

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

Applying the Concepts 2–3

Leading Cause of Death

The following shows approximations of the leading causes of death among men ages 25–44 years. The rates are per 100,000 men. Answer the following questions about the graph.



1. What are the variables in the graph?
2. Are the variables qualitative or quantitative?
3. Are the variables discrete or continuous?
4. What type of graph was used to display the data?
5. Could a Pareto chart be used to display the data?
6. Could a pie chart be used to display the data?
7. List some typical uses for the Pareto chart.
8. List some typical uses for the time series chart.

See page 101 for the answers.

Exercises 2–3

1. Women's Softball Champions The NCAA Women's Softball Division 1 Champions since 1982 are listed below. Use the data to construct a Pareto chart and a vertical bar graph.

'82 UCLA	'94 Arizona
'83 Texas A&M	'95 UCLA
'84 UCLA	'96 Arizona
'85 UCLA	'97 Arizona
'86 Cal St – Fullerton	'98 Fresno State
'87 Texas A&M	'99 UCLA
'88 UCLA	'00 Oklahoma
'89 UCLA	'01 Arizona
'90 UCLA	'02 California
'91 Arizona	'03 UCLA
'92 UCLA	'04 UCLA
'93 Arizona	'05 Michigan

Source: *New York Times Almanac.*

2. Delegates Who Signed the Declaration of Independence The state represented by each delegate who signed the Declaration of Independence is indicated. Organize the data in a Pareto chart and a vertical bar graph and comment on the results.

MA 5	PA 9	SC 4
NH 3	RI 2	CT 4
VA 7	NY 4	DE 3
MD 4	GA 3	
NJ 5	NC 3	

Source: *New York Times Almanac.*

3. Internet Connections The following data represent the estimated number (in millions) of computers connected to the Internet worldwide. Construct a Pareto chart and a horizontal bar graph for the data. Based on the data, suggest the best place to market appropriate Internet products.

Location	Number of computers
Homes	240
Small companies	102
Large companies	148
Government agencies	33
Schools	47

Source: IDC.

4. Roller Coaster Mania The World Roller Coaster Census Report lists the following number of roller

coasters on each continent. Represent the data graphically, using a Pareto chart and a horizontal bar graph.

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

Source: www.rcdb.com

5. World Energy Use The following percentages indicate the source of energy used worldwide. Construct a Pareto chart and a vertical bar graph for the energy used.

Petroleum	39.8%
Coal	23.2
Dry natural gas	22.4
Hydroelectric	7.0
Nuclear	6.4
Other (wind, solar, etc.)	1.2

Source: *New York Times Almanac.*

6. Airline Departures Draw a time series graph to represent the data for the number of airline departures (in millions) for the given years. Over the years, is the number of departures increasing, decreasing, or about the same?

Year	1996	1997	1998	1999	2000	2001	2002
Number of departures	7.9	9.9	10.5	10.9	11.0	9.8	10.1

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

7. Average Global Temperatures Represent these average global temperatures in a time series graph.

1900–09	56.5	1960–69	57.1
1910–19	56.6	1970–79	57.0
1920–29	56.7	1980–89	57.4
1930–39	57.0	1990–99	57.6
1940–49	57.1		
1950–59	57.1		

Source: *World Almanac.*

8. Nuclear Power Reactors Draw a time series graph for the data shown and comment on the trend. The data represent the number of active nuclear reactors.

Year	1992	1994	1996	1998	2000	2002
Number	109	109	109	104	104	104

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

9. Percentage of Voters in Presidential Elections Listed are the percentages of voters who voted in past Presidential elections since 1964. Illustrate the data with a time series graph. The day before the 2006 election, a website published a survey where 90% of the respondents said they voted in the 2004 election. Give possible reasons for the discrepancy.

1964	95.83	1980	76.53	1996	65.97
1968	89.65	1984	74.63	2000	67.50
1972	79.85	1988	72.48	2004	64.0
1976	77.64	1992	78.04		

Source: *New York Times Almanac.*

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. Characteristics of the Population 65 and Over Two characteristics of the population aged 65 and over are shown below for 2004. Illustrate each characteristic with a pie graph.

Marital status		**Educational attainment**	
Never married	3.9%	Less than ninth grade	13.9%
Married	57.2	Completed 9–12 but no diploma	13.0
Widowed	30.8	H.S. graduate	36.0
Divorced	8.1	Some college/ associates degree	18.4
		Bachelor's/advanced degree	18.7

Source: *New York Times Almanac.*

12. Components of the Earth's Crust The following elements comprise the earth's crust, the outermost solid layer. Illustrate the composition of the earth's crust with a pie graph.

Oxygen	45.6%
Silicon	27.3
Aluminum	8.4
Iron	6.2
Calcium	4.7
Other	7.8

Source: *New York Times Almanac.*

13. Workers Switch Jobs In a recent survey, 3 in 10 people indicated that they are likely to leave their jobs when the economy improves. Of those surveyed, 34% indicated that they would make a career change, 29% want a new job in the same industry, 21% are going to start a business, and 16% are going to retire. Make a pie chart and a Pareto chart for the data. Which chart do you think better represents the data?

Source: National Survey Institute.

14. 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 each year during the last 5 years.

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.

15. Presidents' Ages at Inauguration The age at inauguration for each U.S. President is shown. Construct a stem and leaf plot and analyze the data.

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

Source: *New York Times Almanac.*

16. Calories in Salad Dressings A listing of calories per one 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		

17. Twenty Days of Plant Growth The growth (in centimeters) of two varieties of plant after 20 days is shown in this table. Construct a back-to-back stem and leaf plot for the data, and compare the distributions.

Variety 1				Variety 2			
20	12	39	38	18	45	62	59
41	43	51	52	53	25	13	57
59	55	53	59	42	55	56	38
50	58	35	38	41	36	50	62
23	32	43	53	45	55		

18. 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.*

19. The sales of recorded music in 2004 by genre are listed below. Represent the data with an appropriate graph.

Rock	23.9	Jazz	2.7
Country	13.0	Classical	2.0
Rap/hip-hop	12.1	Oldies	1.4
R&B/urban	11.3	Soundtracks	1.1
Pop	10.0	New age	1.0
Religious	6.0	Other	8.9
Children's	2.8		

Source: *World Almanac.*

Extending the Concepts

20. Successful Space Launches The number of successful space launches by the United States and Japan for the years 1993–1997 is shown here. Construct a compound time series graph for the data. What comparison can be made regarding the launches?

Year	1993	1994	1995	1996	1997
United States	29	27	24	32	37
Japan	1	4	2	1	2

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

21. Meat Production Meat production for veal and lamb for the years 1960–2000 is shown here. (Data are in millions of pounds.) Construct a compound time series graph for the data. What comparison can be made regarding meat production?

Year	1960	1970	1980	1990	2000
Veal	1109	588	400	327	225
Lamb	769	551	318	358	234

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

22. Top 10 Airlines The top 10 airlines with the most aircraft are listed. Represent these data with an appropriate graph.

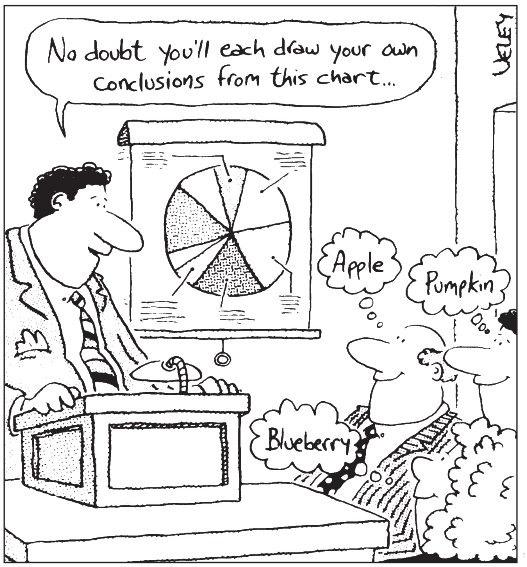
American	714	Continental	364
United	603	Southwest	327
Delta	600	British Airways	268
Northwest	424	American Eagle	245
U.S. Airways	384	Lufthansa (Ger.)	233

Source: *Top 10 of Everything.*

23. Nobel Prizes in Physiology or Medicine The top prize-winning countries for Nobel Prizes in Physiology or Medicine are listed here. Represent the data with an appropriate graph.

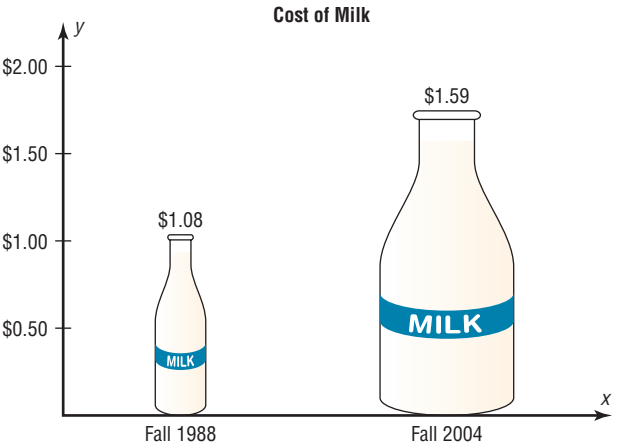
United States	80	Denmark	5
United Kingdom	24	Austria	4
Germany	16	Belgium	4
Sweden	8	Italy	3
France	7	Australia	3
Switzerland	6		

Source: *Top 10 of Everything.*

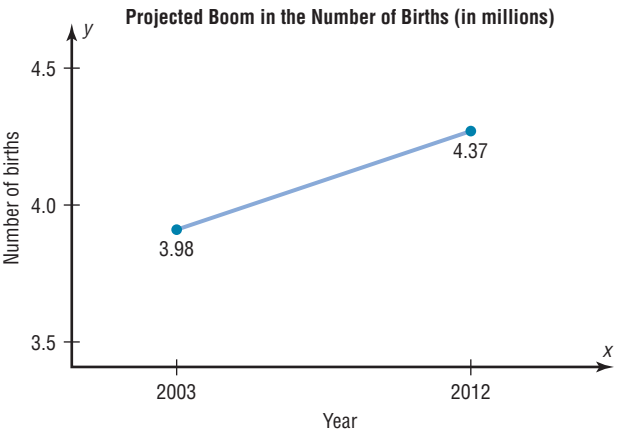


Source: Cartoon by Bradford Veley, Marquette, Michigan. Used with permission.

24. 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?



25. Boom in Number of Births The graph shows the projected boom (in millions) in the number of births. Cite several reasons why the graph might be misleading.

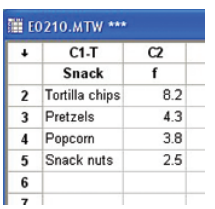


Technology *Step by Step*

MINITAB Step by Step

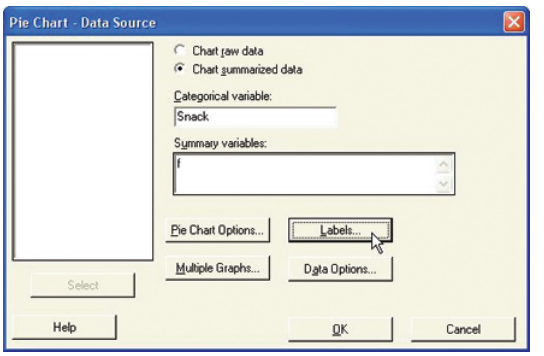
Construct a Pie Chart

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

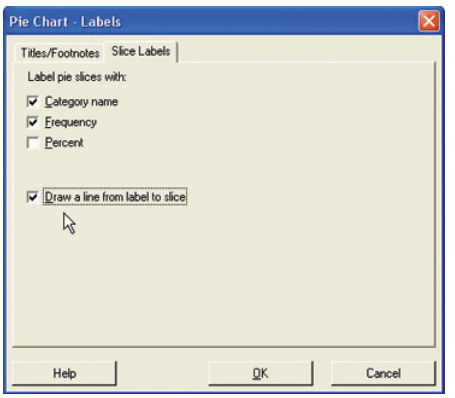


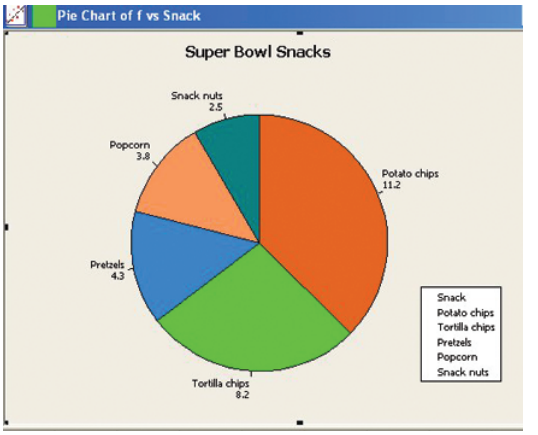
EO210.MTW ***

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



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.



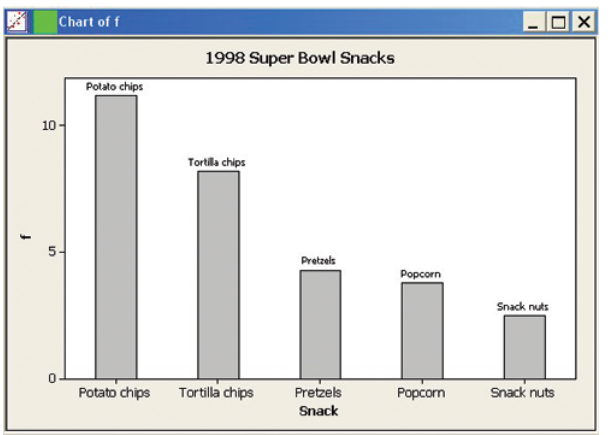


4. Click the [Labels] tab, then Titles/Footnotes.
   a) Type in the title: **Super Bowl Snacks.**
   b) Click the Slice Labels tab, then the options for Category name and Frequency.
   c) Click the option to Draw a line from label to slice.
   d) Click [OK] twice to create the chart.

Construct a 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: 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 Snack for the Categorical variable.

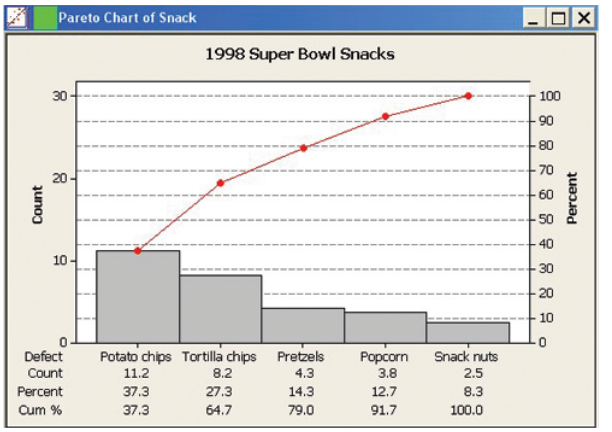


3. Click on [Labels], then type the title in the Titles/Footnote tab: **1998 Super Bowl Snacks.**
4. Click the tab for Data Labels, then click the option to Use labels from column: and select C1 Snacks.
5. Click [OK] twice.

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 Snack.
4. Click on the frequencies column C2 f.



5. Click on [Options].
   a) Check the box for Cumulative percents.
   b) Type in the title, **1998 Super Bowl Snacks.**
6. Click [OK] twice. The chart is completed.

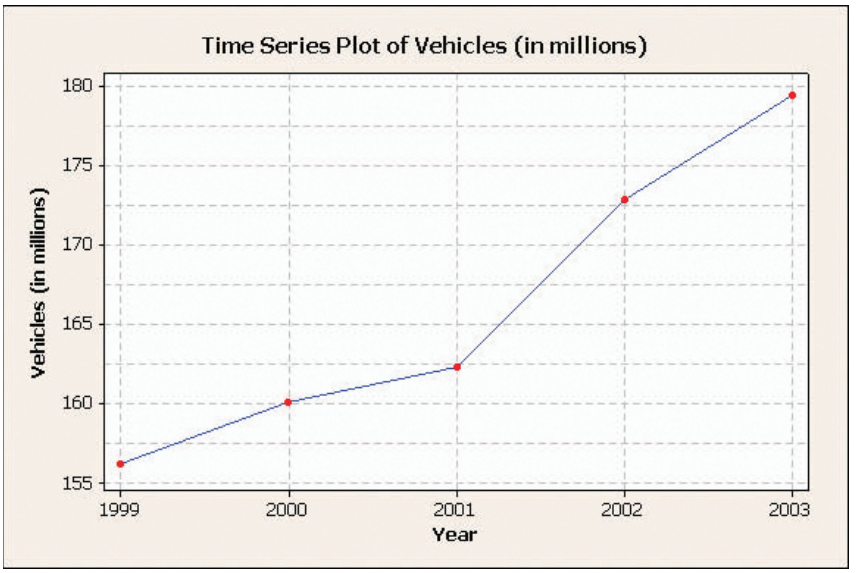
Construct a Time Series Plot

The data used are for the number of vehicles that used the Pennsylvania Turnpike.

Year	1999	2000	2001	2002	2003
Number	156.2	160.1	162.3	172.8	179.4

1. Add a blank worksheet to the project by selecting **File>New>New Worksheet.**
2. To enter the dates from 1999 to 2003 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 **1999.**
   c) To Last value: should be **2003.**
   d) In steps of should be **1** (for every other year). The last two boxes should be 1, the default value.
   e) Click [OK]. The sequence from 1999 to 2003 will be entered in C1 whose label will be Year.
3. Type **Vehicles (in millions)** for the label row above row 1 in C2.

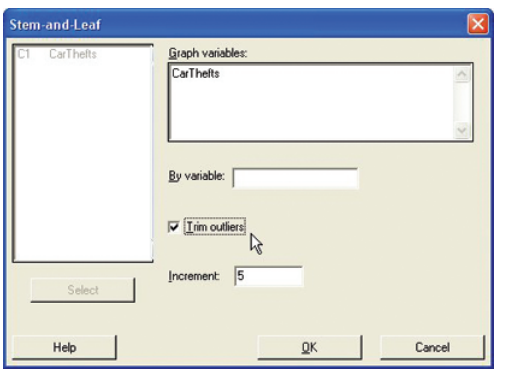
4. Type **156.2** for the first number, then press [Enter]. Never enter the commas for large numbers!
5. Continue entering the value in each row of C2.



6. To make the graph, select **Graph>Time series plot,** then Simple, and press [OK].
   a) For Series select Vehicles (in millions), 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 edit the text.

Construct a Stem and Leaf Plot

1. Type in the data for Example 2–14. 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

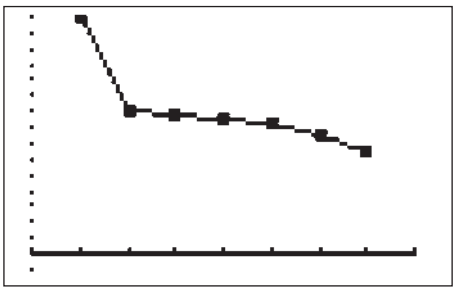
6	5	011233
13	5	5567789
15	6	23
15	6	55667899
7	7	23
5	7	55789

TI-83 Plus or TI-84 Plus 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

Output

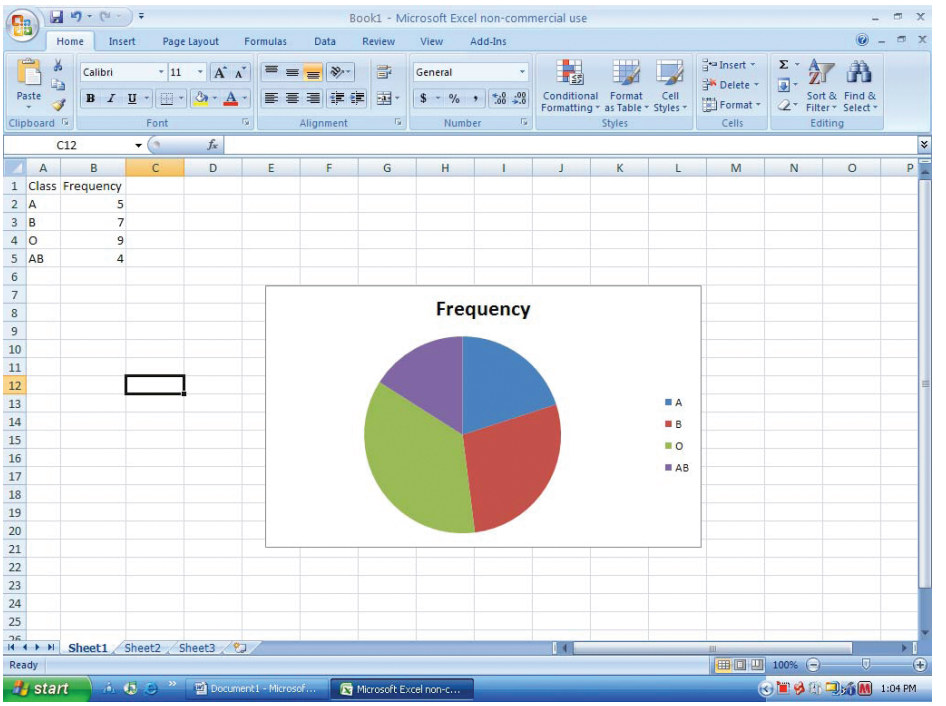


Excel Step by Step

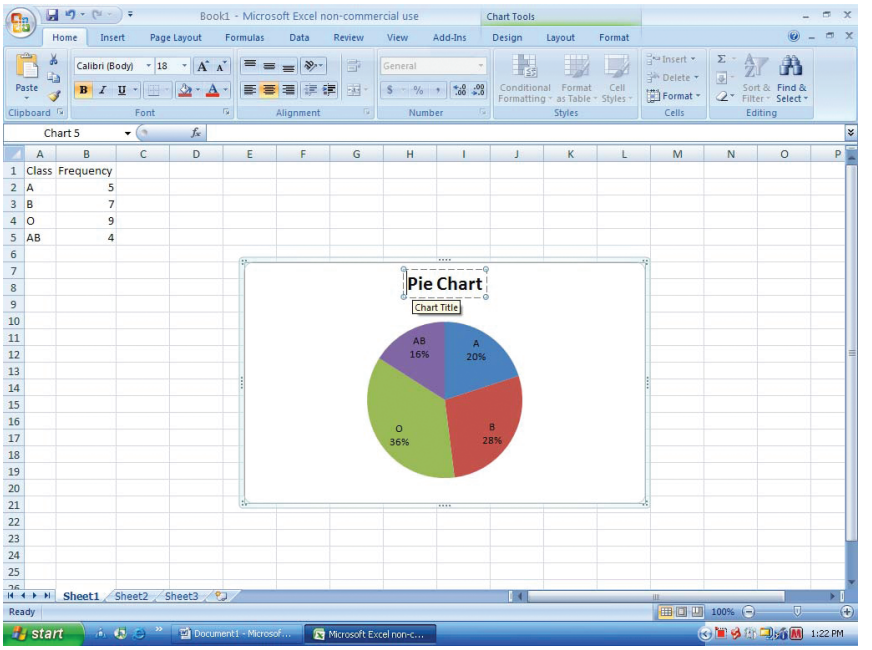
Constructing a Pie Chart

To make a pie chart:

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



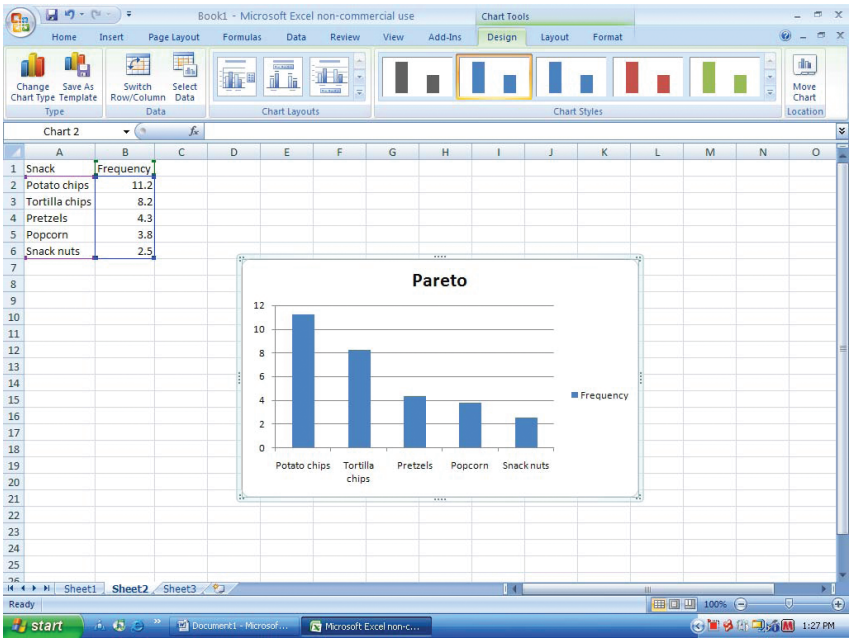
4. Click on any region of the chart. Then select Design from the Chart Tools tab on the toolbar.
5. Select Formulas from the chart Layouts tab on the toolbar.
6. To change the title of the chart, click on the current title of the chart.
7. When the text box containing the title is highlighted, click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



Constructing a Pareto Chart

To make a Pareto chart:

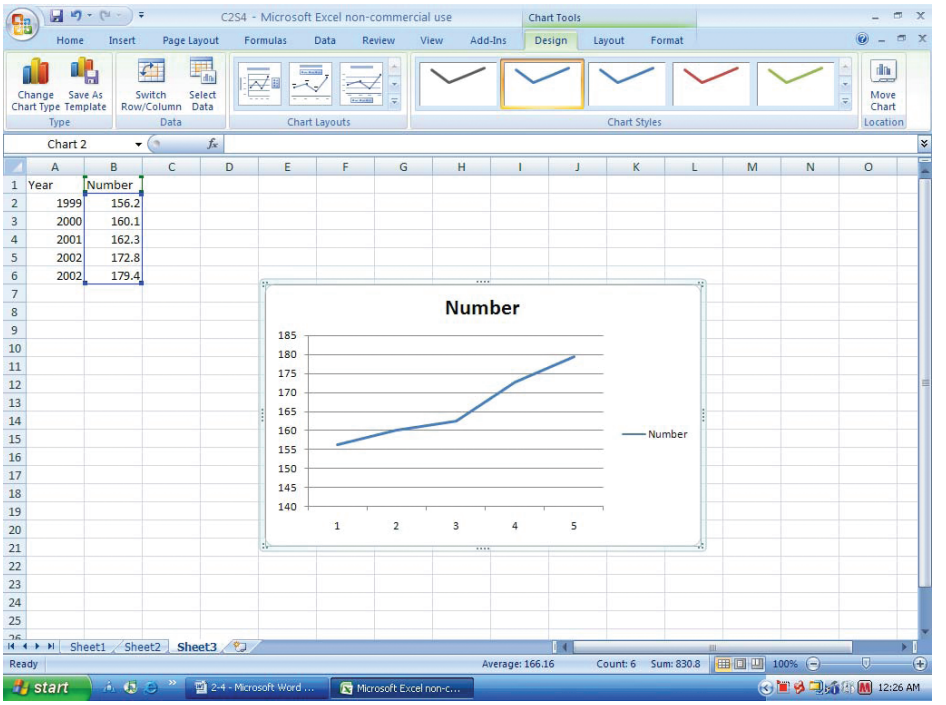
1. Enter the snack food categories from Example 2–11 into column A of a new worksheet.
2. Enter the corresponding frequencies in column B. The data should be entered in descending order according to frequency.
3. Highlight the data from columns A and B and select the Insert tab from the toolbar.
4. Select 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.



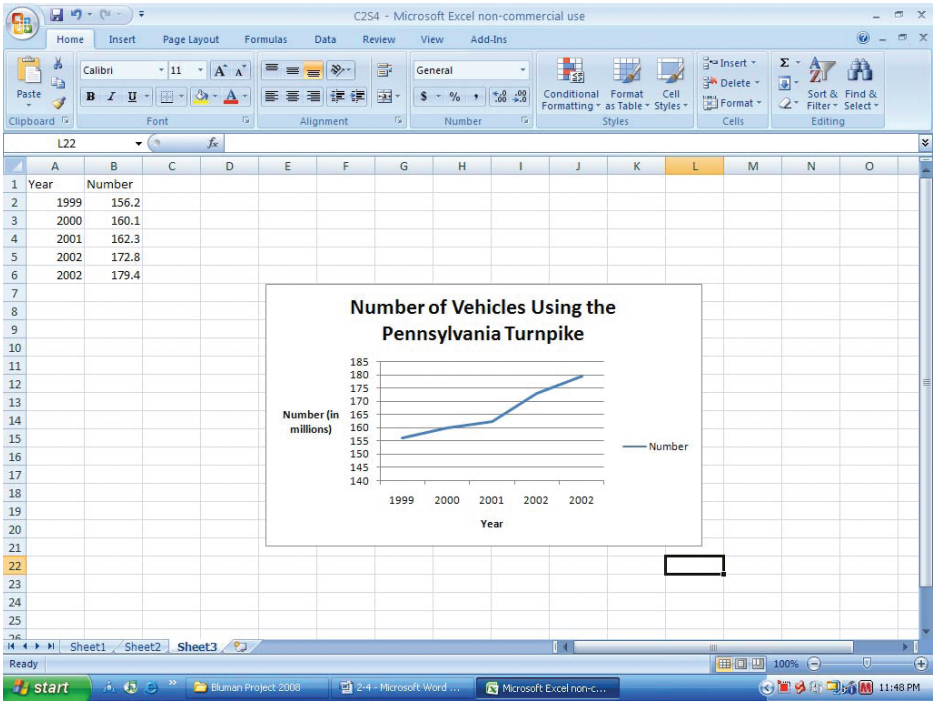
Constructing a Time Series Plot

To make a time series chart:

1. Enter the years 1999 through 2003 from Example 2–10 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.



Summary

When data are collected, they 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 solution. Once a frequency distribution is constructed, the representation of the data by graphs is a simple task. The most commonly used graphs in research statistics are the histogram, frequency polygon, and ogive. Other graphs, such as the bar graph, Pareto chart, time series graph, and pie graph, can also be used. Some of these graphs are seen frequently in newspapers, magazines, and various statistical reports.

Finally, 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 advantages of a frequency distribution and a histogram.

Important Terms

bar graph 69
categorical frequency distribution 38
class 37
class boundaries 39
class midpoint 40
class width 39
cumulative frequency 54
cumulative frequency distribution 42
frequency 37
frequency distribution 37
frequency polygon 53
grouped frequency distribution 39
histogram 51
lower class limit 39
ogive 54
open-ended distribution 41
Pareto chart 70
pie graph 73
raw data 37
relative frequency graph 56
stem and leaf plot 80
time series graph 72
ungrouped frequency distribution 43
upper class limit 39

Important Formulas

Formula for the percentage of values in each class:

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

where

f = **frequency of the 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:

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

or

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

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

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

Review Exercises

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. The data in this exercise will be used for Exercise 2 in this section.

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. Construct a pie graph for the data in Exercise 1, and analyze the results.

3. **Ball Sales** A sporting goods store kept a record of sales of five items for one randomly selected hour during a recent sale. Construct a frequency distribution for the data (B = baseballs, G = golf balls, T = tennis balls, S = soccer balls, F = footballs). (The data for this exercise will be used for Exercise 4 in this section.)

F	B	B	B	G	T	F
G	G	F	S	G	T	
F	T	T	T	S	T	
F	S	S	G	S	B	

4. Draw a pie graph for the data in Exercise 3 showing the sales of each item, and analyze the results.

5. **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. (The data for this exercise will be used for Exercise 6.)

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

6. Construct a histogram, a frequency polygon, and an ogive for the data in Exercise 5 in this section, and analyze the results.

7. The percentage (rounded to the nearest whole percent) of persons from each state completing 4 years or more of college is listed below. Organize the data into a grouped frequency distribution with 5 classes.

Percentage of persons completing 4 years of college

23	25	24	34	22	24	27	37	33	24
26	23	38	24	24	17	28	23	30	25
30	22	33	24	28	36	24	19	25	31
34	31	27	24	29	28	21	25	26	15
26	22	27	21	25	28	24	21	25	26

Source: *New York Times Almanac.*

8. Using the data in Exercise 7, construct a histogram, a frequency polygon, and an ogive.

9. **NFL Franchise Values** The data shown (in millions of dollars) are the values of the 30 National Football League franchises. Construct a frequency distribution for the data using 8 classes. (The data for

this exercise will be used for Exercises 10 and 12 in this section.)

170	191	171	235	173	187	181	191
200	218	243	200	182	320	184	239
186	199	186	210	209	240	204	193
211	186	197	204	188	242		

Source: *Pittsburgh Post-Gazette.*

10. Construct a histogram, a frequency polygon, and an ogive for the data in Exercise 9 in this section, and analyze the results.

11. Ages of the Vice Presidents at the Time of Their Death The ages of the Vice Presidents of the United States at the time of their death are listed below. Use the data to construct a frequency distribution, histogram, frequency polygon, and ogive, using relative frequencies. Use 6 classes.

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	70

Source: *New York Times Almanac.*

12. Construct a histogram, frequency polygon, and ogive by using relative frequencies for the data in Exercise 9 in this section.

13. NBA Champions The NBA Champions from 1985 on are listed below. Use the data to construct a Pareto chart and a vertical bar graph.

1985 Los Angeles	1996 Chicago
1986 Boston	1997 Chicago
1987 Los Angeles	1998 Chicago
1988 Detroit	1999 San Antonio
1989 Detroit	2000 Los Angeles
1990 Detroit	2001 Los Angeles
1991 Chicago	2002 Los Angeles
1992 Chicago	2003 San Antonio
1993 Chicago	2004 Detroit
1994 Houston	2005 San Antonio
1995 Houston	

Source: *World Almanac.*

14. Trial-Ready Cases Construct a Pareto chart and a horizontal bar graph for the number of trial-ready civil action and equity cases decided in less than 6 months for the selected counties in southwestern Pennsylvania.

County	Number of cases
Westmoreland	427
Washington	298
Green	151
Fayette	106
Somerset	87

Source: *Pittsburgh Tribune-Review.*

15. Minimum Wage The given data represent the federal minimum hourly wage in the years shown. Draw a time series graph to represent the data and analyze the results.

Year	Wage
1960	$1.00
1965	1.25
1970	1.60
1975	2.10
1980	3.10
1985	3.35
1990	3.80
1995	4.25
2000	5.15
2005	5.15

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

16. Farm Data Construct a time series graph for each set of data and analyze the results.

Year	No. of farms (millions)	Avg. size (acres)
1940	6.35	174
1950	5.65	213
1960	3.96	297
1970	2.95	374
1980	2.44	426
1990	2.15	460
2000	2.17	436

Source: *World Almanac.*

17. Presidential Debates The data show the number (in millions) of viewers who watched the first and second Presidential debates. Construct two time series graphs and compare the results.

Year	1992	1996	2000	2004
First debate	62.4	36.1	46.6	62.5
Second debate	69.9	36.3	37.6	46.7

Source: *Nielson Media Research.*

18. Working Women In a study of 100 women, the numbers shown here indicate the major reason why each woman surveyed worked outside the home. Construct a pie graph for the data and analyze the results.

Reason	Number of women
To support self/family	62
For extra money	18
For something different to do	12
Other	8

19. Career Changes A survey asked if people would like to spend the rest of their careers with their present employers. The results are shown. Construct a pie graph for the data and analyze the results.

Answer	Number of people
Yes	660
No	260
Undecided	80

20. Museum Visitors The number of visitors to the Railroad Museum during 24 randomly selected hours is shown here. Construct a stem and leaf plot for the data.

67	62	38	73	34	43	72	35
53	55	58	63	47	42	51	62
32	29	47	62	29	38	36	41

21. 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.*

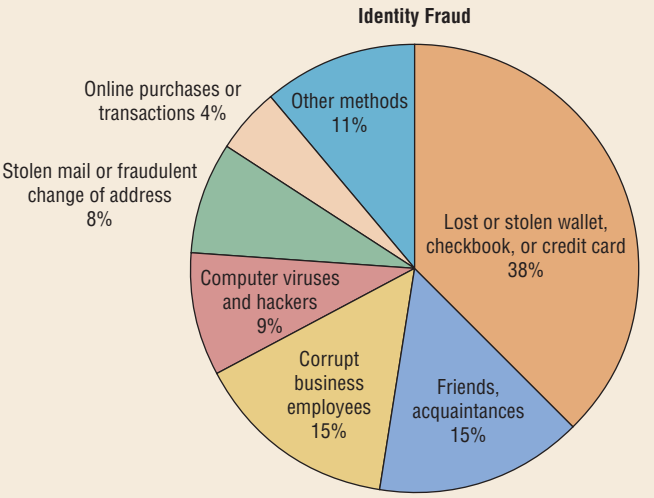
22. Job Aptitude Test A special aptitude test is given to job applicants. The data shown here represent the scores of 30 applicants. Construct a stem and leaf plot for the data and summarize the results.

204	210	227	218	254
256	238	242	253	227
251	243	233	251	241
237	247	211	222	231
218	212	217	227	209
260	230	228	242	200

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.



Data Analysis

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

1. From the Data Bank located in Appendix D, 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 D, 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 D, 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 D, 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 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 D, 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. Histograms 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).

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 Problem 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.

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 Problem 20.

22. Murders in Selected Cities For a recent year, the number of murders in 25 selected cities is shown. Construct a frequency distribution using 9 classes, and analyze the nature of the data in terms of shape, extreme values, etc. (The information in this exercise will be used for Exercise 23 in this section.)

248	348	74	514	597
270	71	226	41	39
366	73	241	46	34
149	68	73	63	65
109	598	278	69	27

Source: *Pittsburgh Tribune Review.*

23. Construct a histogram, frequency polygon, and ogive for the data in Problem 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. Trespasser Fatalities The data show the number of fatal trespasser casualties on railroad property in the United States. Draw a time series graph and explain any trend.

Year	1998	1999	2000	2001
Number	536	463	511	540

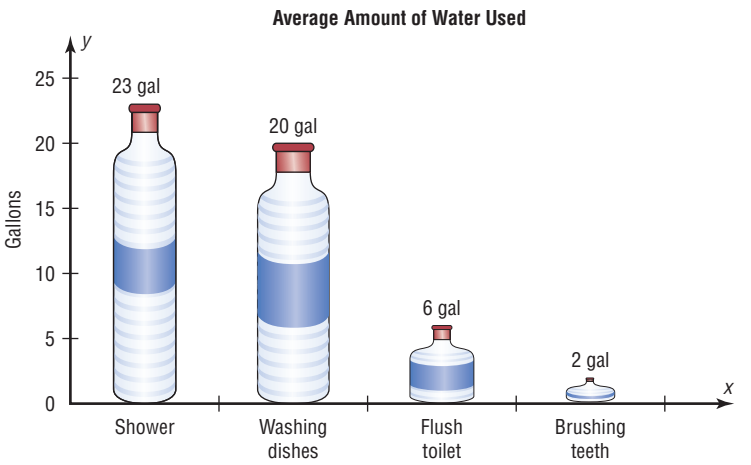
Source: Federal Railroad Administration.

26. 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

Critical Thinking Challenges

1. 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?



2. 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.*

3. 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 1992 to 2004. 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(s) did the smallest number of teacher strikes occur? How many were there?

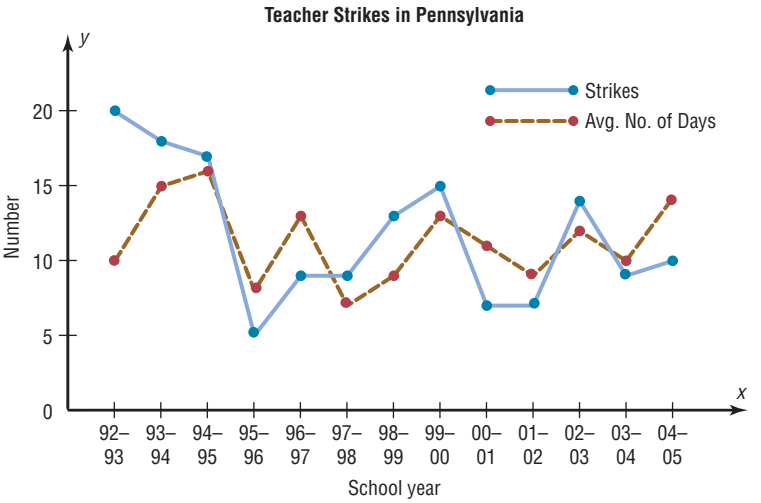
c. In what year was the average duration of the strikes the longest? What was it?

d. In what year was the average duration of the strikes the shortest? What 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 1992–1993 and 2004–2005.

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



Source: Pennsylvania School Boards Associations.

Data Projects

Where appropriate, use MINITAB, the TI-83 Plus, the TI-84 Plus, Excel, 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 their earnings per share. Randomly select 30 stocks traded on the NASDAQ. For each, find their 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 one 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	6
50–53	7
54–57	16
58–61	5
62–65	4
66–69	2

5. Answers will vary. For the frequency distribution given in Exercise 4, there is a peak for the 54–57 bin.
6. Answers will vary. This frequency distribution shows no outliers. However, if we had split our frequency into 14 bins 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 Leading Cause of Death

1. The variables in the graph are the year, cause of death, and rate of death per 100,000 men.
2. The cause of death is qualitative, while the year and death rates are quantitative.
3. Year is a discrete variable, and death rate is continuous. Since cause of death is qualitative, it is neither discrete nor continuous.
4. A line graph was used to display the data.
5. No, a Pareto chart could not be used to display the data, since we can only have one quantitative variable and one categorical variable in a Pareto chart.
6. We cannot use a pie chart for the same reasons as given for the Pareto chart.
7. A Pareto chart is typically used to show a categorical variable listed from the highest-frequency category to the category with the lowest frequency.
8. A time series chart is used to see trends in the data. It can also be used for forecasting and predicting.