Chapter 2

Organizing Data

Section 2.1

Variables and Data

Definition 2.1

Variables

Variable: A characteristic that varies from one person or thing to another.

Qualitative variable: A nonumerically valued variable.

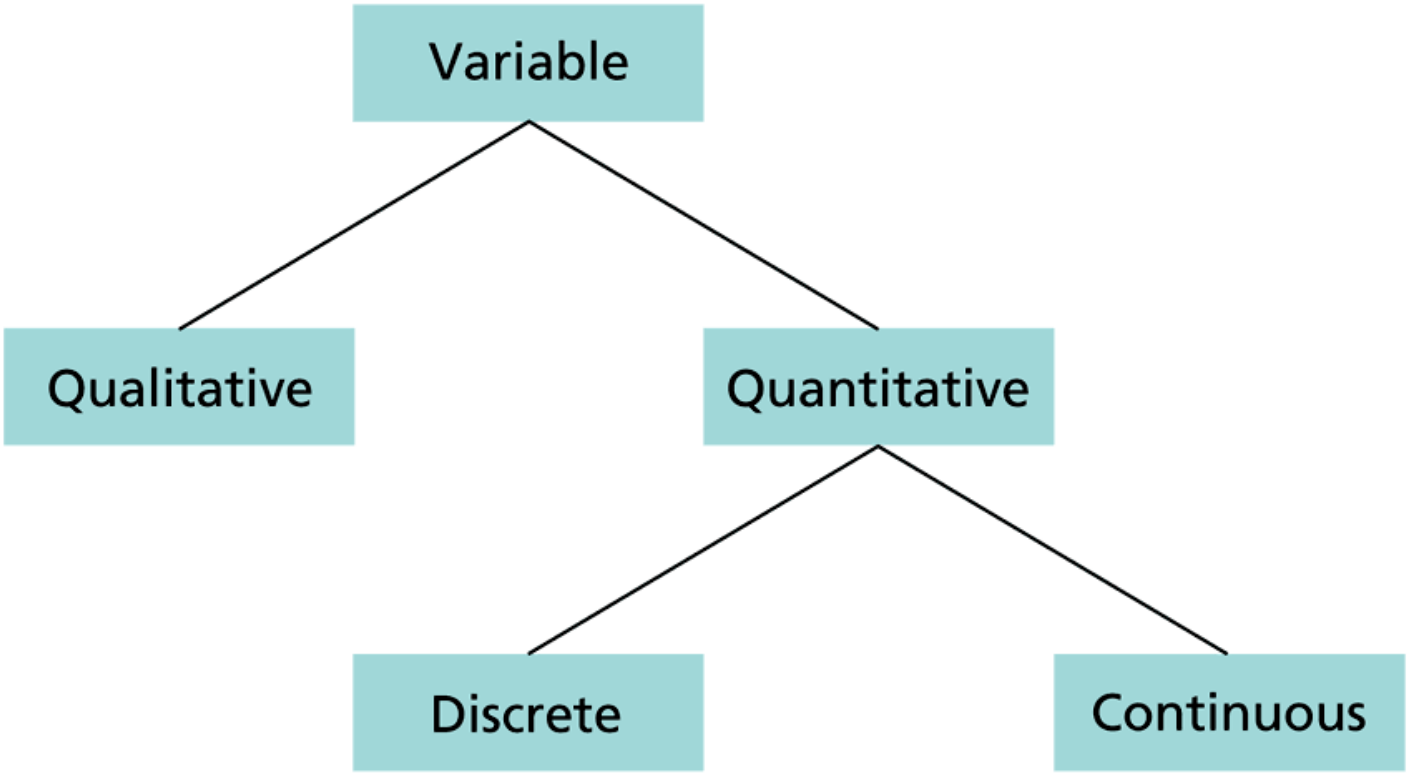
Quantitative variable: A numerically valued variable.

Discrete variable: A quantitative variable whose possible values can be listed. In particular, a quantitative variable with only a finite number of possible values is a discrete variable.

Continuous variable: A quantitative variable whose possible values form some interval of numbers.

Figure 2.1

Types of variables



Definition 2.2

Data

Data: Values of a variable.

Qualitative data: Values of a qualitative variable.

Quantitative data: Values of a quantitative variable.

Discrete data: Values of a discrete variable.

Continuous data: Values of a continuous variable.

Section 2.2

Organizing Qualitative Data

Definition 2.3

Frequency Distribution of Qualitative Data

A **frequency distribution** of qualitative data is a listing of the distinct values and their frequencies.

Procedure 2.1

To Construct a Frequency Distribution of Qualitative Data

Step 1 List the distinct values of the observations in the data set in the first column of a table.

Step 2 For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.

Step 3 Count the tallies for each distinct value and record the totals in the third column of the table.

Table 2.1

Political party affiliations of the students in introductory statistics

Democratic	Other	Democratic	Other	Democratic
Republican	Republican	Other	Other	Republican
Republican	Republican	Republican	Democratic	Republican
Republican	Democratic	Democratic	Other	Republican
Democratic	Democratic	Republican	Democratic	Democratic
Republican	Republican	Other	Other	Democratic
Republican	Democratic	Republican	Other	Other
Republican	Republican	Republican	Democratic	Republican

Table 2.2

Table for constructing a frequency distribution for the political party affiliation data in Table 2.1

Party	Tally	Frequency
Democratic	𝍸 𝍸 |||	13
Republican	𝍸 𝍸 𝍸 |||	18
Other	𝍸 ||||	9
		40

Definition 2.4

Relative-Frequency Distribution of Qualitative Data

A **relative-frequency distribution** of qualitative data is a listing of the distinct values and their relative frequencies.

Procedure 2.2

To Construct a Relative-Frequency Distribution of Qualitative Data

Step 1 Obtain a frequency distribution of the data.

Step 2 Divide each frequency by the total number of observations.

Table 2.3

Relative-frequency distribution for the political party affiliation data in Table 2.1

Party	Relative frequency	
Democratic	0.325	← *13/40*
Republican	0.450	← *18/40*
Other	0.225	← *9/40*
	1.000	

Definition 2.5

Pie Chart

A **pie chart** is a disk divided into wedge-shaped pieces proportional to the relative frequencies of the qualitative data.

Procedure 2.3

To Construct a Pie Chart

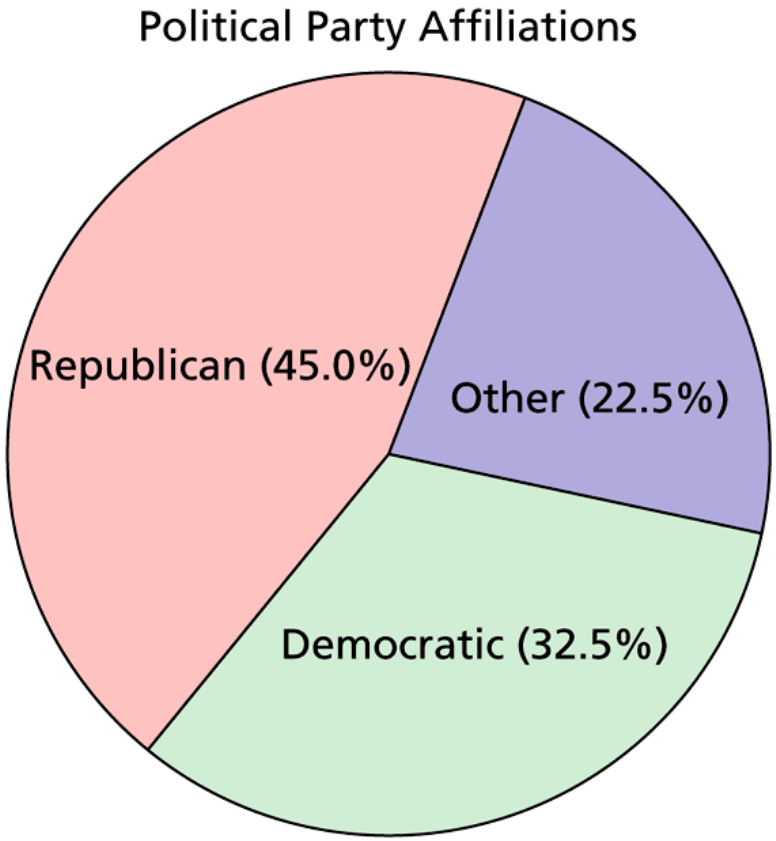
Step 1 Obtain a relative-frequency distribution of the data by applying Procedure 2.2.

Step 2 Divide a disk into wedge-shaped pieces proportional to the relative frequencies.

Step 3 Label the slices with the distinct values and their relative frequencies.

Figure 2.2

Pie chart of the political party affiliation data in Table 2.1



Definition 2.6

Bar Chart

A **bar chart** displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percents) of those values on a vertical axis. The relative frequency of each distinct value is represented by a vertical bar whose height is equal to the relative frequency of that value. The bars should be positioned so that they do not touch each other.

Procedure 2.4

To Construct a Bar Chart

Step 1 Obtain a relative-frequency distribution of the data by applying Procedure 2.2.

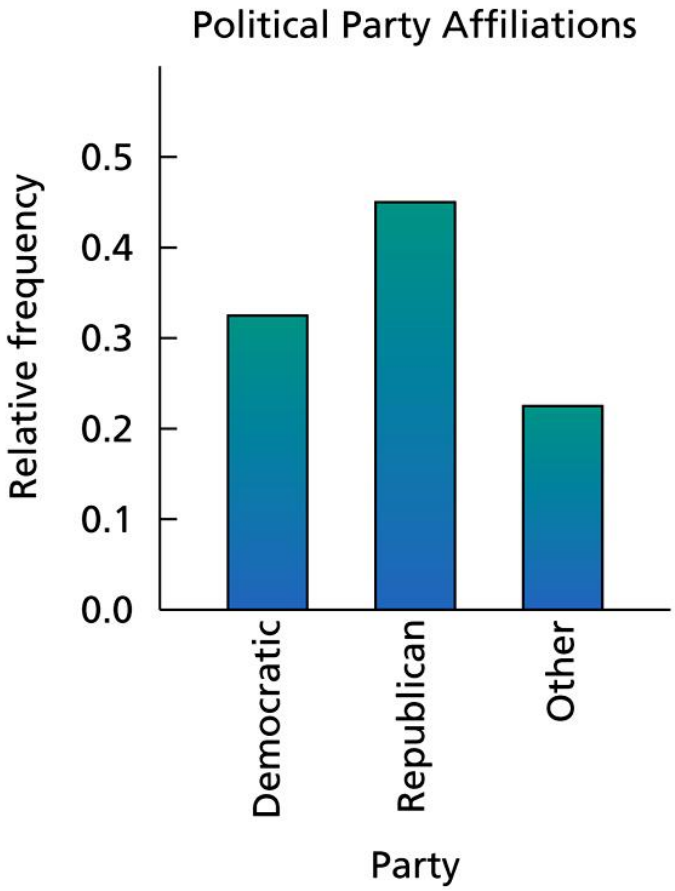
Step 2 Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.

Step 3 For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.

Step 4 Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with "Relative frequency."

Figure 2.3

Bar chart of the political party affiliation data in Table 2.1



Section 2.3

Organizing Quantitative Data

Table 2.4

Number of TV sets in each of 50 randomly selected households.

1	1	1	2	6	3	3	4	2	4
3	2	1	5	2	1	3	6	2	2
3	1	1	4	3	2	2	2	2	3
0	3	1	2	1	2	3	1	1	3
3	2	1	2	1	1	3	1	5	1

Table 2.5

Frequency and relative-frequency distributions, using single-value grouping, for the number-of-TVs data in Table 2.4

Number of TVs	Frequency	Relative frequency
0	1	0.02
1	16	0.32
2	14	0.28
3	12	0.24
4	3	0.06
5	2	0.04
6	2	0.04
	50	1.00

Table 2.6

Days to maturity for 40 short-term investments

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Table 2.7

Frequency and relative-frequency distributions, using limit grouping, for the days-to-maturity data in Table 2.6

Days to maturity	Tally	Frequency	Relative frequency
30–39	|||	3	0.075
40–49	|	1	0.025
50–59	𝍸 |||	8	0.200
60–69	𝍸 𝍸	10	0.250
70–79	𝍸 ||	7	0.175
80–89	𝍸 ||	7	0.175
90–99	||||	4	0.100
		40	1.000

Definition 2.7

Terms Used in Limit Grouping

Lower class limit: The smallest value that could go in a class.

Upper class limit: The largest value that could go in a class.

Class width: The difference between the lower limit of a class and the lower limit of the next-higher class.

Class midpoint: The average of the two class limits of a class.

Definition 2.9

Histogram

A **histogram** displays the classes of the quantitative data on a horizontal axis and the frequencies (relative frequencies, percents) of those classes on a vertical axis. The frequency (relative frequency, percent) of each class is represented by a vertical bar whose height is equal to the frequency (relative frequency, percent) of that class. The bars should be positioned so that they touch each other.

- For single-value grouping, we use the distinct values of the observations to label the bars, with each such value centered under its bar.
- For limit grouping or cutpoint grouping, we use the lower class limits (or, equivalently, lower class cutpoints) to label the bars. Note: Some statisticians and technologies use class marks or class midpoints centered under the bars.

Procedure 2.5

To Construct a Histogram

Step 1 Obtain a frequency (relative-frequency, percent) distribution of the data.

Step 2 Draw a horizontal axis on which to place the bars and a vertical axis on which to display the frequencies (relative frequencies, percents).

Step 3 For each class, construct a vertical bar whose height equals the frequency (relative frequency, percent) of that class.

Step 4 Label the bars with the classes, as explained in Definition 2.9, the horizontal axis with the name of the variable, and the vertical axis with "Frequency" ("Relative frequency," "Percent").

Figure 2.4

Single-value grouping. Number of TVs per household:
(a) frequency histogram; (b) relative-frequency histogram

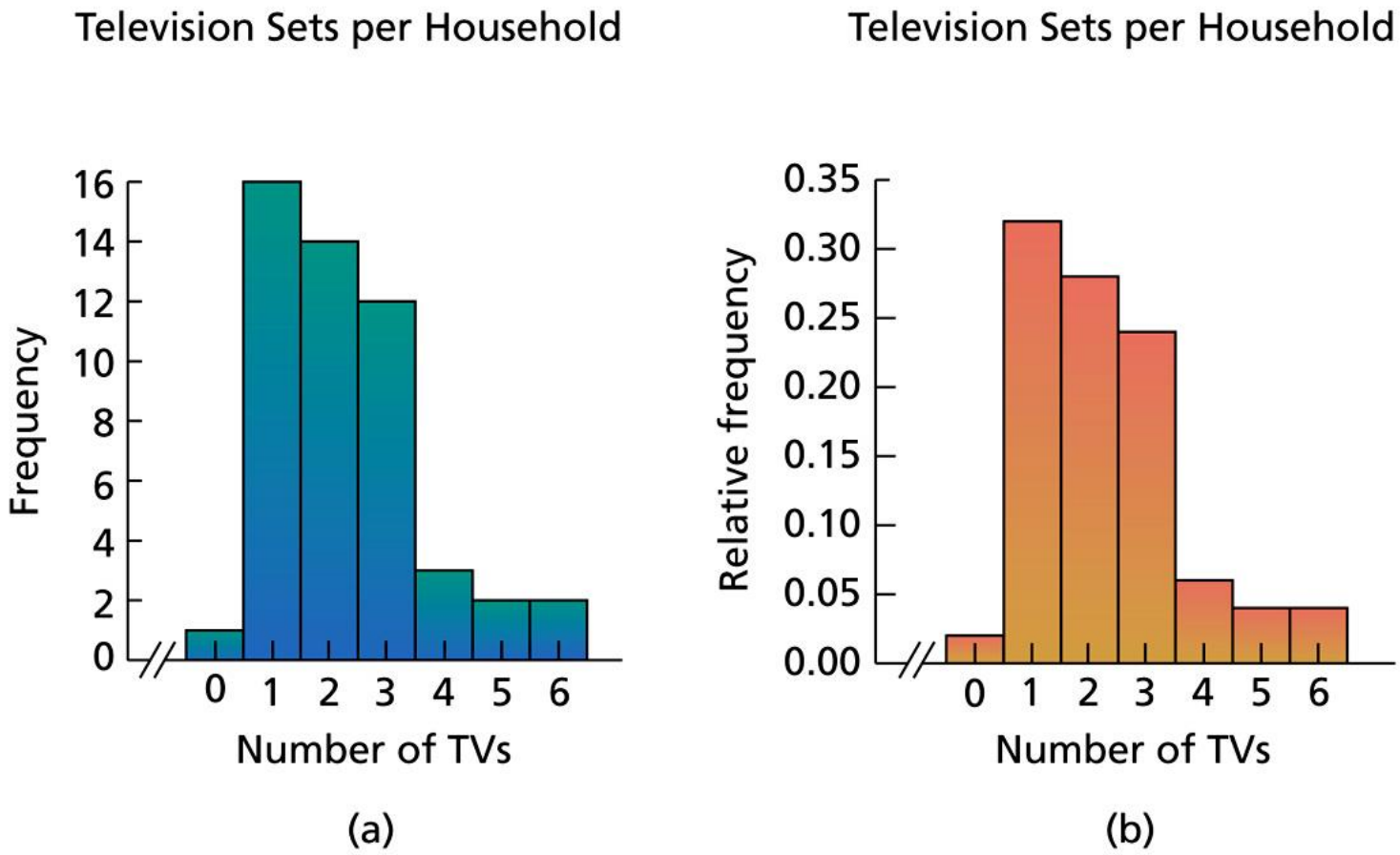
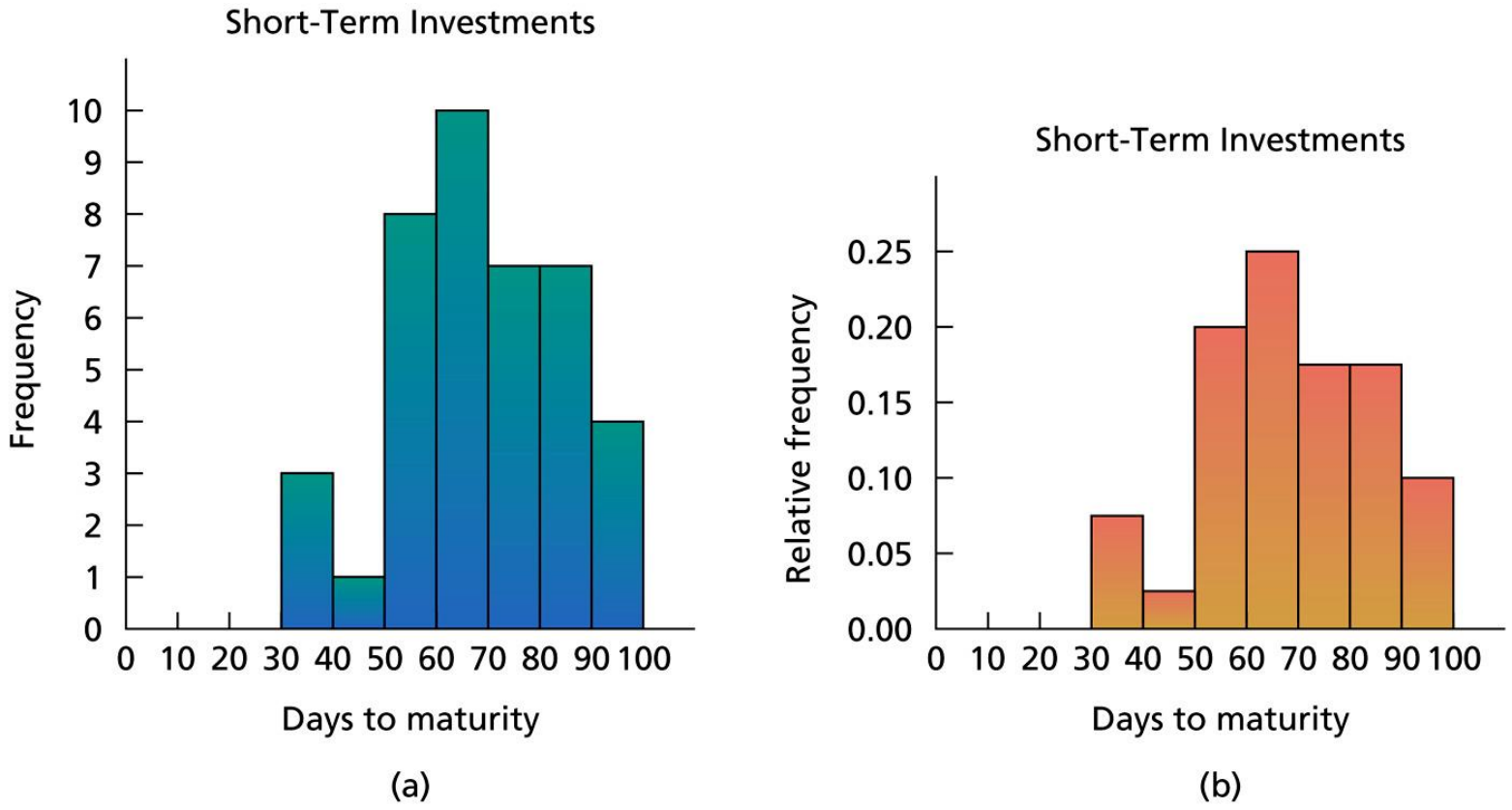


Figure 2.5

Limit grouping. Days to maturity: (a) frequency histogram; (b) relative-frequency histogram



Definition 2.10

Dotplot

A **dotplot** is a graph in which each observation is plotted as a dot at an appropriate place above a horizontal axis. Observations having equal values are stacked vertically.

Procedure 2.6

To Construct a Dotplot

Step 1 Draw a horizontal axis that displays the possible values of the quantitative data.

Step 2 Record each observation by placing a dot over the appropriate value on the horizontal axis.

Step 3 Label the horizontal axis with the name of the variable.

Table 2.11 & Figure 2.7

Prices, in dollars, of 16 DVD players

210	219	214	197
224	219	199	199
208	209	215	199
212	212	219	210

Definition 2.11

Stem-and-Leaf Diagrams

In a **stem-and-leaf diagram** (or **stemplot**), each observation Is separated into two parts, namely, a **stem**—consisting of all but the rightmost digit– and a **leaf**, the rightmost digit.

Procedure 2.7

To Construct a Stem-and-Leaf Diagram

Step 1 Think of each observation as a stem—consisting of all but the rightmost digit—and a leaf, the rightmost digit.

Step 2 Write the stems from smallest to largest in a vertical column to the left of a vertical rule.

Step 3 Write each leaf to the right of the vertical rule in the row that contains the appropriate stem.

Step 4 Arrange the leaves in each row in ascending order.

Table 2.12 & Figure 2.8

Days to maturity for
40 short-term investments

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Constructing a stem-and-leaf diagram for the days-to-maturity data

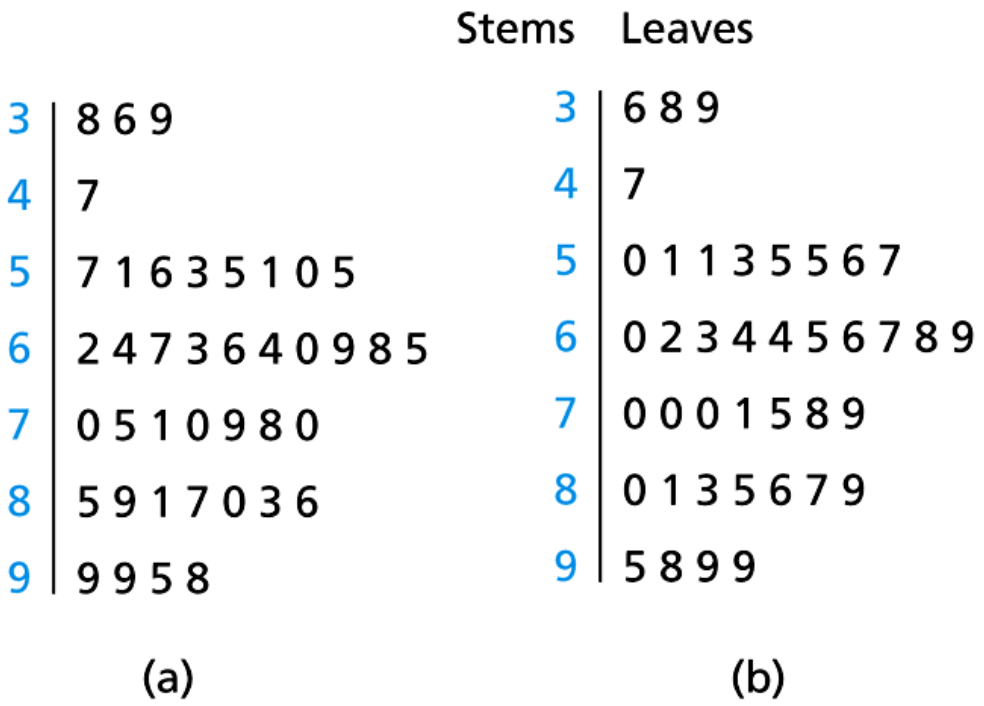


Table 2.13 & Figure 2.9

Cholesterol levels
for 20 high-level patients

210	209	212	208
217	207	210	203
208	210	210	199
215	221	213	218
202	218	200	214

Stem-and-leaf diagram for cholesterol levels:
(a) one line per stem; (b) two lines per stem

19	9
20	0 2 3 7 8 8 9
21	0 0 0 0 2 3 4 5 7 8 8
22	1

(a)

19	
19	9
20	0 2 3
20	7 8 8 9
21	0 0 0 0 2 3 4
21	5 7 8 8
22	1
22	

(b)

Section 2.4

Distribution Shapes

Definition 2.12

Distribution of a Data Set

The **distribution of a data set** is a table, graph, or formula that provides the values of the observations and how often they occur.

Figure 2.10

Relative-frequency histogram and approximating smooth curve for the distribution of heights

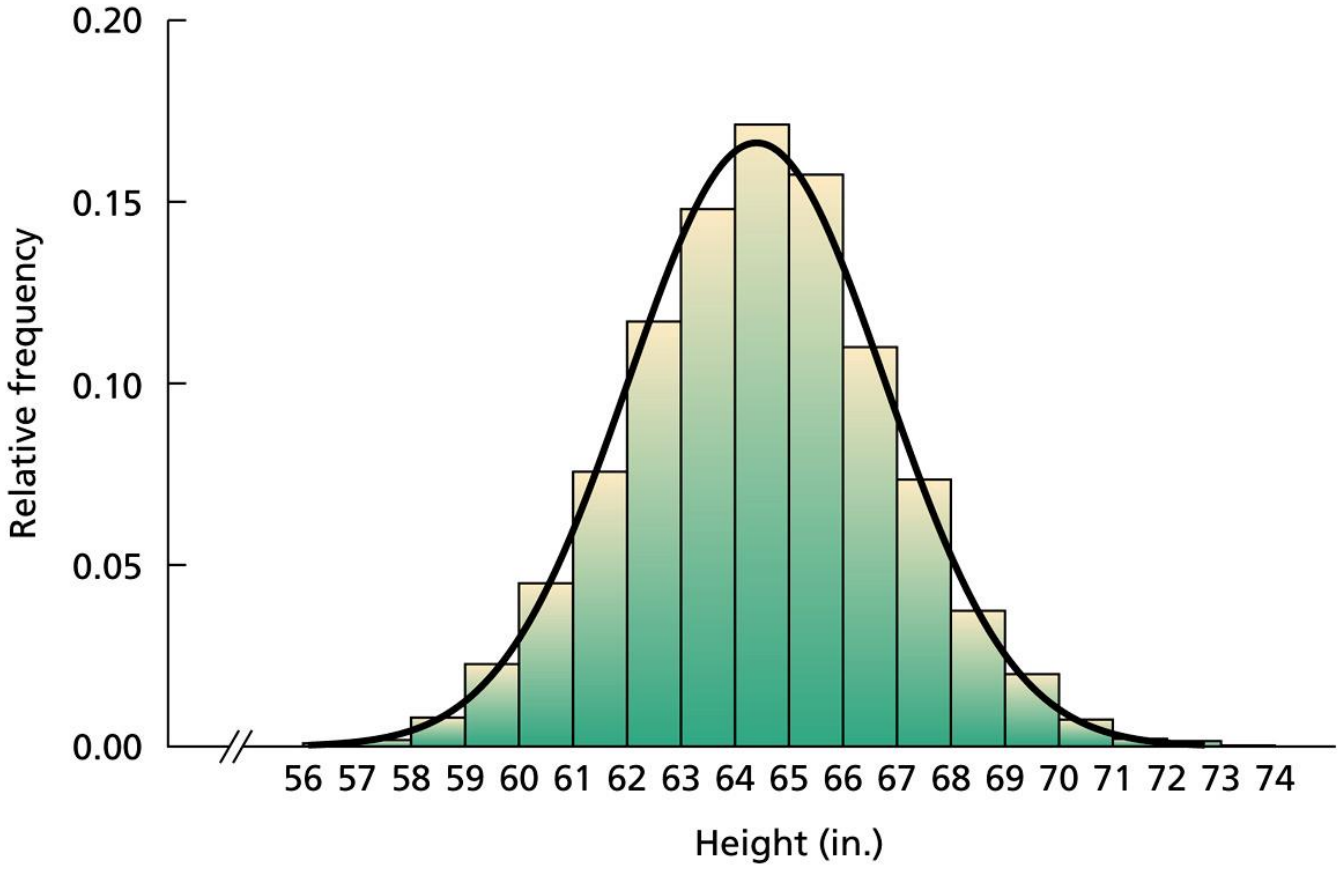
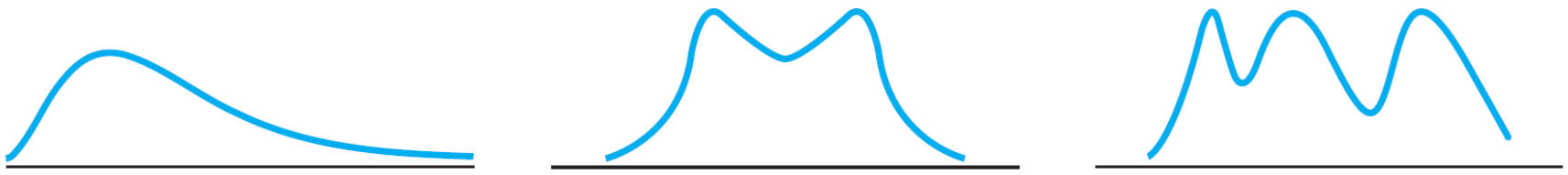


Figure 2.11

Examples of (a) unimodal, (b) bimodal, and (c) multimodal distributions



(a) Unimodal (b) Bimodal (c) Multimodal

Shapes of Distributions - Symmetric

Symmetric Distribution

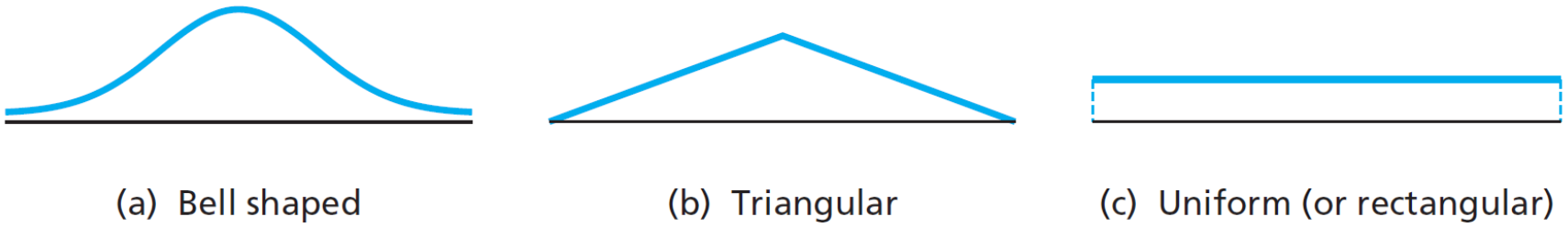
- A vertical line can be drawn through the middle of a graph of the distribution and the resulting halves are approximately mirror images.

Uniform Distribution (rectangular)

- All entries or classes in the distribution have equal or approximately equal frequencies.
- Symmetric

Figure 2.12

Examples of symmetric distributions: (a) bell shaped, (b) triangular, and (c) uniform



(a) Bell shaped (b) Triangular (c) Uniform (or rectangular)

Shapes of Distributions - Skewed

Skewed Left Distribution (negatively skewed)

- The "tail" of the graph elongates more to the left.

Skewed Right Distribution (positively skewed)

- The "tail" of the graph elongates more to the right.

Figure 2.13

Generic skewed distributions: (a) right skewed (b) left skewed

(a) Right skewed

(b) Left skewed

Figure 2.14

Reverse-J-shaped distribution

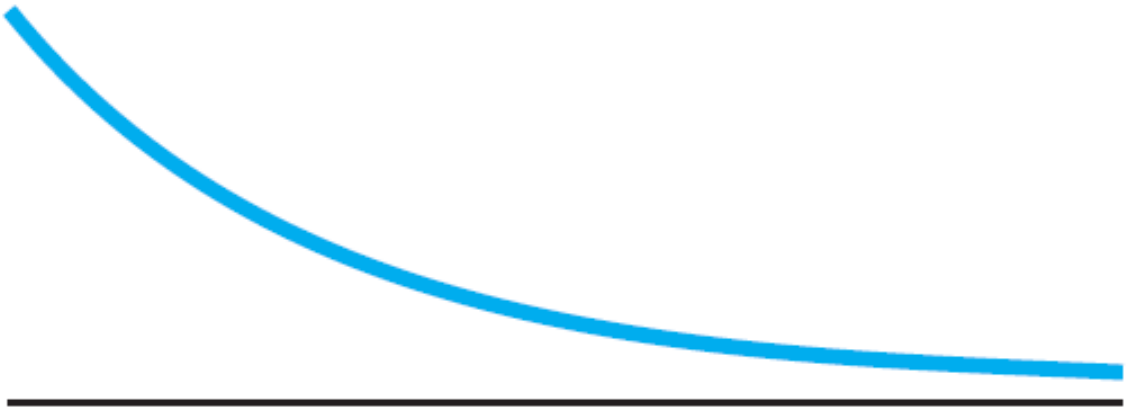
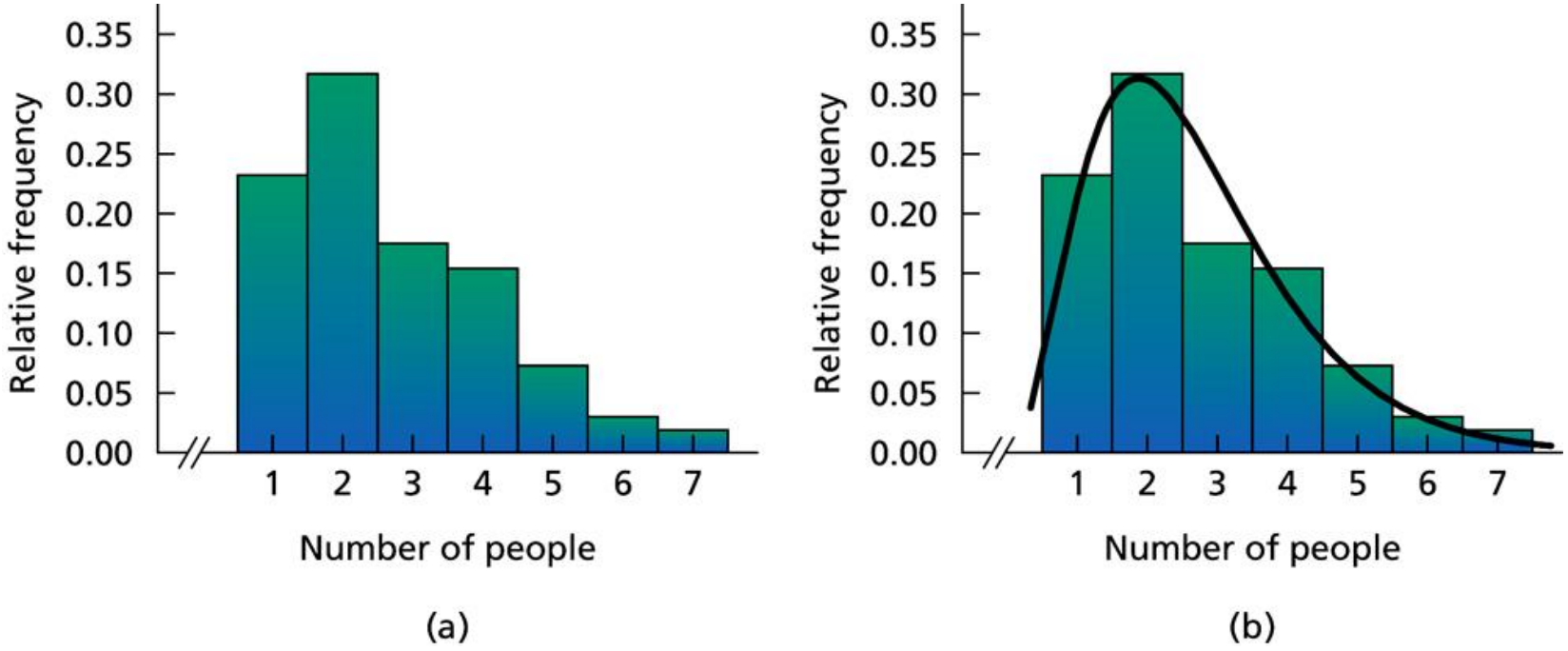


Figure 2.15

Relative-frequency histogram for household size



Definition 2.13

Population and Sample Data

Population data: The values of a variable for the entire population.

Sample data: The values of a variable for a sample of the population.

Definition 2.14

Population and Sample Distributions; Distribution of a Variable

The distribution of population data is called the **population distribution,** or the **distribution of the variable.**

The distribution of sample data is called a **sample distribution.**

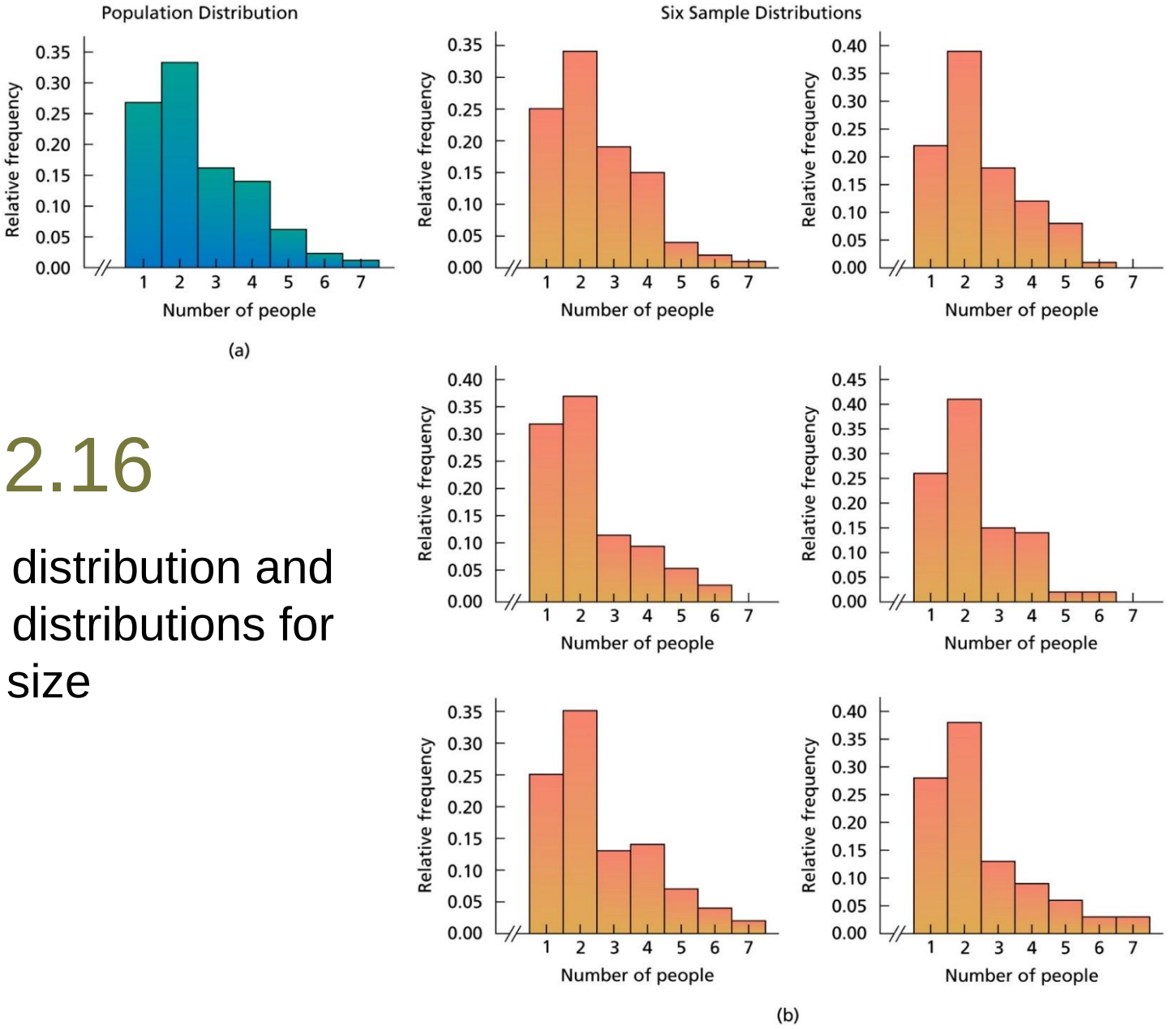


Figure 2.16

Population distribution and six sample distributions for household size

Key Fact 2.1

Population and Sample Distributions

For a simple random sample, the sample distribution approximates the population distribution (i.e., the distribution of the variable under consideration). The larger the sample size, the better the approximation tends to be.

