2

THE ORGANIZATION AND GRAPHIC PRESENTATION OF DATA

Demographers examine the size, composition, and distribution of human populations. Changes in the birth, death, and migration rates of a population affect its composition and social characteristics.[1] To examine a large population, researchers often have to deal with very large amounts of data. For example, imagine the amount of data it takes to describe the immigrant or elderly population in the United States. To make sense out of these data, a researcher must organize and summarize the data in some systematic fashion. In this chapter, we review three such methods used by social scientists: (1) the creation of frequency distributions, (2) the construction of bivariate tables and (3) the use of graphic presentation.

Chapter Learning Objectives

1. Construct and analyze frequency, percentage, and cumulative distributions.
2. Calculate proportions and percentages.
3. Compare and contrast frequency and percentage distributions for nominal, ordinal, and interval-ratio variables.
4. Create a bivariate table
5. Construct and interpret a pie chart, bar graph, histogram, the statistical map, line graph, and time-series chart.

FREQUENCY DISTRIBUTIONS

The most basic way to organize data is to classify the observations into a frequency distribution. A **frequency distribution** is a table that reports the number of observations that fall into each category of the variable we are analyzing. Constructing a frequency distribution is usually the first step in the statistical analysis of data.

Immigration has been described as "remaking America with political, economic, and cultural ramifications."[2] Globalization has fueled migration, particularly since the beginning of the 21st century. Workers migrate because of the promise of employment and higher standards of living than what is attainable in their home countries. Data reveal that many migrants seek specifically to move to the United States.[3] The U.S. Census Bureau uses the term foreign born to refer to those who are not U.S. citizens at birth. The U.S. Census estimates that 13.5% of the U.S. population, or approximately 44 million people, are foreign born.[4] Immigrants

Frequency distribution: A table reporting the number of observations falling into each category of the variable.

are not one homogeneous group but are many diverse groups. Table 2.1 shows the frequency distribution of the world region of birth for the foreign-born population.

The frequency distribution is organized in a table, which has a number (2.1) and a descriptive title. The title indicates the kind of data presented: "Frequency Distribution for Categories of Region of Birth for Foreign-Born Population." The table consists of two columns. The first column identifies the variable (*world region of birth*) and its categories. The second column, with the heading "Frequency (f)," tells the number of cases in each category as well as the total number of cases (N = 43,681,654). Note also that the source of the table is clearly identified. It tells us that the data are from a 2018 report by Jynnah Radford and Abby Budiman (although the information is based on 2016 American Community Survey data from the U.S. Census). The source of the data can be reported as a source note or in the title of the table. This table is also referred to as a **univariate frequency table**, as it presents frequency information for a single variable.

Univariate frequency table: A table that displays the distribution of one variable

What can you learn from the information presented in Table 2.1? The table shows that as of 2016, approximately 44 million people were classified as foreign born. Out of this group, most—about 11.7 million people—were from South and East Asia, just under 11.6 million were from Mexico, followed by about 5.8 million from Europe or Canada.

Table 2.1 Frequency Distribution for Categories of Region of Birth for Foreign-Born Population, 2016

Region of Birth	Frequency (f)
South and East Asia	11,731,584
Mexico	11,568,060
Europe/Canada	5,785,135
Caribbean	4,300,022
Central America	3,463,389
South America	2,927,145
Middle East	1,875,264
Sub-Saharan Africa	1,769,778
All other	261,277
Total	43,681,654

Source: "2016, Foreign-Born Population in the United States Statistical Portrait", Pew Research Center, Washington, D.C (September 14, 2018), https://www.pewhispanic.org/2018/09/14/2016-statistical-information-on-foreign-born-in-united-states/.

PROPORTIONS AND PERCENTAGES

Frequency distributions are helpful in presenting information in a compact form. However, when the number of cases is large, the frequencies may be difficult to grasp. To standardize these raw frequencies, we can translate them into relative frequencies—that is, proportions or percentages.

A **proportion** is a relative frequency obtained by dividing the frequency in each category by the total number of cases. To find a proportion (p), divide the frequency (f) in each category by the total number of cases (N):

Proportion: A relative frequency obtained by dividing the frequency in each category by the total number of cases.

$$p = \frac{f}{N} \tag{2.1}$$

where

f = frequency

N = total number of cases

We've calculated the proportion for the three largest groups of foreign born. First, the proportion of foreign born originally from South and East Asia is

$$\frac{11{,}731{,}584}{43{,}681{,}654} = .269$$

The proportion of foreign born who were originally from Mexico is

$$\frac{11{,}568{,}060}{43{,}681{,}654} = .265$$

The proportion of foreign born who were originally from Europe or Canada is

$$\frac{5{,}785{,}135}{43{,}681{,}654} = .132$$

The proportion of foreign born who were originally from all other reported areas (combining the category "All other" with those from the Caribbean, Central and South America, Middle East, and sub-Saharan Africa) is

$$\frac{14{,}596{,}875}{43{,}681{,}654} = .334$$

Proportions should always sum to 1.00 (allowing for some rounding errors). Thus, in our example, the sum of the six proportions is

$$0.27 + 0.27 + 0.13 + 0.33 = 1.0$$

To determine a frequency from a proportion, we simply multiply the proportion by the total N:

$$f = p(N) \tag{2.2}$$

Thus, the frequency of foreign born from South and East Asia can be calculated as

$$0.27\ (43{,}681{,}654) = 11{,}794{,}047$$

The obtained frequency differs somewhat from the actual frequency of 11,731,584. This difference is due to rounding off of the proportion. If we use the actual proportion instead of the rounded proportion, we obtain the correct frequency:

$$0.268570050026036\ (43{,}681{,}654) = 11{,}731{,}584$$

Percentage: A relative frequency obtained by dividing the frequency in each category by the total number of cases and multiplying by 100.

We can also express frequencies as percentages. A **percentage** is a relative frequency obtained by dividing the frequency in each category by the total number of cases and multiplying by 100. In most statistical reports, frequencies are presented as percentages rather than proportions. Percentages express the size of the frequencies as if there were a total of 100 cases.

To calculate a percentage, multiply the proportion by 100:

$$\text{Percentage}\ (\%) = \frac{f}{N}(100) \tag{2.3}$$

or

$$\text{Percentage}\ (\%) = p(100) \tag{2.4}$$

Thus, the percentage of respondents who were originally from Mexico is

$$0.27\ (100) = 27\%$$

LEARNING CHECK 2.1

Calculate the proportion and percentage of males and females in your statistics class. What proportion is female?

PERCENTAGE DISTRIBUTIONS

Percentage distribution: A table showing the percentage of observations falling into each category of the variable.

Percentages are usually displayed as percentage distributions. A **percentage distribution** is a table showing the percentage of observations falling into each

Table 2.2 Frequency Distribution for Categories of Region of Birth for Foreign-Born Population, 2016

Region of Birth	Frequency (*f*)	Percentage (%)
South and East Asia	11,731,584	27
Mexico	11,568,060	27
Europe/Canada	5,785,135	13
Caribbean	4,300,022	10
Central America	3,463,389	8
South America	2,927,145	7
Middle East	1,875,264	4
Sub-Saharan Africa	1,769,778	4
All other	261,277	1
Total	43,681,654	100* (rounded)

Source: "2016, Foreign-Born Population in the United States Statistical Portrait", Pew Research Center, Washington, DC (September 14, 2018), https://www.pewhispanic.org/2018/09/14/2016-statistical-information-on-foreign-born-in-united-states/.

category of the variable. For example, Table 2.2 presents the frequency distribution of categories of places of origin (Table 2.1) along with the corresponding percentage distribution. Percentage distributions (or proportions) should always show the base (*N*) on which they were computed. Thus, in Table 2.2, the base on which the percentages were computed is $N = 43{,}681{,}654$.

THE CONSTRUCTION OF FREQUENCY DISTRIBUTIONS

In this section, you will learn how to construct frequency distributions. Most often, we can use statistical software to accomplish this, but it is important to go through the process to understand how frequency distributions are actually constructed.

For nominal and ordinal variables, constructing a frequency distribution is quite simple. To do so, count and report the number of cases that fall into each category of the variable along with the total number of cases (*N*). For the purpose of illustration, let's take a small random sample of 40 cases from a General Social Survey (GSS) sample and record their scores on the following variables: gender, a nominal-level variable; degree, an ordinal measurement of education; and age and number of children, both interval-ratio variables. The use of "male" and "female" in parts of this book is in keeping with the GSS categories for the variable *sex* (respondent's sex).

The interviewer recorded the gender of each respondent at the beginning of the interview. To measure degree, researchers asked everyone to indicate the highest degree completed: less than high school, high school, some college, bachelor's degree, and graduate degree. The first category represented the lowest level of education. Researchers calculated respondents' age based on the respondent's birth year. The number of children was determined by the question, "How many children have you ever had?" The answers given by our subsample of 40 respondents are displayed in Table 2.3. Note that each row in the table represents a respondent, whereas each column represents a variable. This format is conventional in the social sciences.

You can see that it is going to be difficult to make sense of these data just by eyeballing Table 2.3. How many of these 40 respondents are males? How many said that they had a graduate degree? How many were older than 50 years of age? To answer these questions, we construct a frequency distribution for each variable.

Table 2.3 A GSS Subsample of 40 Respondents

Gender of Respondent	Degree	Number of Children	Age
M	Bachelor	1	43
F	High school	2	71
F	High school	0	71
M	High school	0	37
M	High school	0	28
F	High school	6	34
F	High school	4	69
F	Graduate	0	51
F	Bachelor	0	76
M	Graduate	2	48
M	Graduate	0	49
M	Less than high school	3	62
F	Less than high school	8	71
F	High school	1	32

Gender of Respondent	Degree	Number of Children	Age
F	High school	1	59
F	High school	1	71
M	High school	0	34
M	Bachelor	0	39
F	Bachelor	2	50
M	High school	3	82
F	High school	1	45
M	High school	0	22
M	High school	2	40
F	High school	2	46
M	High school	0	29
F	High school	1	75
F	High school	0	23
M	Bachelor	2	35
M	Bachelor	3	44
F	High school	3	47
M	High school	1	84
F	Graduate	1	45
F	Less than high school	3	24
F	Graduate	0	47
F	Less than high school	5	67
F	High school	1	21
F	High school	0	24
F	High school	3	49
F	High school	3	45
F	Graduate	3	37

Note: M = male; F = female.

Table 2.4 Frequency Distribution of the Variable *Gender*: GSS Subsample

Gender	Tallies	Frequency (*f*)	Percentage (%)
Male	𝍸 𝍸 𝍸	15	37.5
Female	𝍸 𝍸 𝍸 𝍸 𝍸	25	62.5
Total (*N*)		40	100.0

Frequency Distributions for Nominal Variables

Let's begin with the nominal variable, *gender*. First, we tally the number of males, then the number of females (the column of tallies has been included in Table 2.4 for the purpose of illustration). The tally results are then used to construct the frequency distribution presented in Table 2.4. The table has a title describing its content ("Frequency Distribution of the Variable Gender: GSS Subsample"). Its categories (male and female) and their associated frequencies are clearly listed; in addition, the total number of cases (*N*) is also reported. The Percentage column is the percentage distribution for this variable. To convert the Frequency column to percentages, simply divide each frequency by the total number of cases and multiply by 100. Percentage distributions are routinely added to almost any frequency table and are especially important if comparisons with other groups are to be considered. Immediately, we can see that it is easier to read the information. There are 25 females and 15 males in this sample. Based on this frequency distribution, we can also conclude that the majority of sample respondents are female.

LEARNING CHECK 2.2

Construct a frequency and percentage distribution for males and females in your statistics class.

Frequency Distributions for Ordinal Variables

To construct a frequency distribution for ordinal-level variables, follow the same procedures outlined for nominal-level variables. Table 2.5 presents the frequency distribution for the variable *degree*. The table shows that 60.0%, a majority, indicated that their highest degree was a high school degree.

The major difference between frequency distributions for nominal and ordinal variables is the order in which the categories are listed. The categories

Table 2.5 Frequency Distribution of the Variable *Degree*: GSS Subsample

Degree	Tallies	Frequency (*f*)	Percentage (%)
Less than high school	||||	4	10.0
High school	卌 卌 卌 卌 ||||	24	60.0
Bachelor	卌 |	6	15.0
Graduate	卌 |	6	15.0
Total (*N*)		40	100.0

for nominal-level variables do not have to be listed in any particular order. For example, we could list females first and males second without changing the nature of the distribution. Because the categories or values of ordinal variables are rank-ordered, however, they must be listed in a way that reflects their rank—from the lowest to the highest or from the highest to the lowest. Thus, the data on degree in Table 2.5 are presented in declining order from "less than high school" (the lowest educational category) to "graduate" (the highest educational category).

Frequency Distributions for Interval-Ratio Variables

We hope that you agree by now that constructing frequency distributions for nominal- and ordinal-level variables is rather straightforward. Simply list the categories and count the number of observations that fall into each category. Building a frequency distribution for interval-ratio variables with relatively few values is also easy. For example, when constructing a frequency distribution for number of children, simply list the number of children and report the corresponding frequency, as shown in Table 2.6.

Very often interval-ratio variables have a wide range of values, which makes simple frequency distributions very difficult to read. For example, look at the frequency distribution for the variable *age* in Table 2.7. The distribution contains age values ranging from 21 to 84 years. For a more concise picture, the large number of different scores could be reduced into a smaller number of groups, each containing a range of scores. Table 2.8 displays such a grouped frequency distribution of the data in Table 2.7. Each group, known as a class interval, now contains 10 possible scores instead of 1. Thus, the ages of 21, 22, 23, 24, 28, and 29 all fall into a single class interval of 20–29. The second column of Table 2.8, Frequency, tells us the number of respondents

Table 2.6 Frequency Distribution of Variable *Number of Children*: GSS Subsample

Number of Children	Frequency (*f*)	Percentage (%)
0	13	32.5
1	9	22.5
2	6	15.0
3	8	20.0
4	1	2.5
5	1	2.5
6	1	2.5
7+	1	2.5
Total (*N*)	40	100.0

Table 2.7 Frequency Distribution of the Variable *Age*: GSS Subsample

Age of Respondent	Frequency (*f*)	Age of Respondent	Frequency (*f*)
21	1	40	1
22	1	43	1
23	1	44	1
24	2	45	3
28	1	46	1
29	1	47	2
32	1	48	1
34	2	49	2
35	1	50	1
37	2	51	1
39	1	59	1

Age of Respondent	Frequency (*f*)	Age of Respondent	Frequency (*f*)
62	1		
67	1		
69	1		
71	4		
75	1		
76	1		
82	1		
84	1		

Table 2.8 Grouped Frequency Distribution of the Variable *Age*: GSS Subsample

Age Category	Frequency (*f*)	Percentage (%)
20–29	7	17.5
30–39	7	17.5
40–49	12	30.0
50–59	3	7.5
60–69	3	7.5
70–79	6	15.0
80–89	2	5.0
Total (*N*)	40	100.0

who fall into each of the intervals—for example, that seven respondents fall into the class interval of 20–29. Having grouped the scores, we can clearly see that the biggest single age group is between 40 and 49 years (12 out of 40, or 30% of sample). The percentage distribution that we have added to Table 2.8 displays the relative frequency of each interval and emphasizes this pattern as well.

The decision as to how many groups to use and, therefore, how wide the intervals should be is usually up to the researcher and depends on what makes sense in terms of the purpose of the research. The rule of thumb is that an interval width should be large enough to avoid too many categories but not so large that significant differences between observations are concealed. Obviously, the number of intervals depends on the width of each. For instance, if you are working with scores ranging from 10 to 60 and you establish an interval width of 10, you will have five intervals.

LEARNING CHECK 2.3

Can you verify that Table 2.8 was constructed correctly? Use Table 2.7 to determine the frequency of cases that fall into the categories of Table 2.8.

LEARNING CHECK 2.4

If you are having trouble distinguishing between nominal, ordinal, and interval-ratio variables, review the section on levels of measurement in Chapter 1. The distinction between these levels of measurement will be important throughout the book.

CUMULATIVE DISTRIBUTIONS

Sometimes, we may be interested in locating the relative position of a given score in a distribution. For example, we may be interested in finding out how many or what percentage of our sample was younger than 40 or older than 60. Frequency distributions can be presented in a cumulative fashion to answer such questions. A **cumulative frequency distribution** shows the frequencies at or below each category of the variable.

Cumulative frequency distribution: A distribution showing the frequency at or below each category (class interval or score) of the variable.

Cumulative frequencies are appropriate only for variables that are measured at an ordinal level or higher. They are obtained by adding to the frequency in each category the frequencies of all the categories below it.

Let's look at Table 2.9. It shows the cumulative frequencies based on the frequency distribution from Table 2.8. The cumulative frequency column, denoted by *Cf*, shows the number of persons at or below each interval. For example, you can see that 14 of the 40 respondents were 39 years old or younger, and 29 respondents were 59 years old or younger.

To construct a cumulative frequency distribution, start with the frequency in the lowest class interval (or with the lowest score, if the data are

Table 2.9 Grouped Frequency Distribution and Cumulative Frequency for the Variable *Age*: GSS Subsample

Age Category	Frequency (*f*)	Cumulative Frequency (*Cf*)
20–29	7	7
30–39	7	14
40–49	12	26
50–59	3	29
60–69	3	32
70–79	6	38
80–89	2	40
Total (*N*)	40	

ungrouped), and add to it the frequencies in the next highest class interval. Continue adding the frequencies until you reach the last class interval. The cumulative frequency in the last class interval will be equal to the total number of cases (*N*). In Table 2.9, the frequency associated with the first class interval (20–29) is 7. The cumulative frequency associated with this interval is also 7, since there are no cases below this class interval. The frequency for the second class interval is 7. The cumulative frequency for this interval is 7 + 7 = 14. To obtain the cumulative frequency of 26 for the third interval, we add its frequency (12) to the cumulative frequency associated with the second class interval (14). Continue this process until you reach the last class interval. Therefore, the cumulative frequency for the last interval is equal to 40, the total number of cases (*N*).

We can also construct a cumulative percentage distribution (*C*%), which has wider applications than the cumulative frequency distribution (*Cf*). A **cumulative percentage distribution** shows the percentage at or below each category (class interval or score) of the variable. A cumulative percentage distribution is constructed using the same procedure as for a cumulative frequency distribution except that the percentages—rather than the raw frequencies—for each category are added to the total percentages for all the previous categories.

Cumulative percentage distribution: A distribution showing the percentage at or below each category (class interval or score) of the variable.

In Table 2.10, we have added the cumulative percentage distribution to the frequency and percentage distributions shown in Table 2.8. The cumulative percentage distribution shows, for example, that 35% of the sample was 39 years or younger.

Table 2.10 Grouped Frequency Distribution and Cumulative Percentages for the Variable *Age*: GSS Subsample

Age Category	Frequency (*f*)	Percentage (%)	Cumulative Percentage (*C*%)
20–29	7	17.5	17.5
30–39	7	17.5	35.0
40–49	12	30.0	65.0
50–59	3	7.5	72.5
60–69	3	7.5	80.0
70–79	6	15.0	95.0
80–89	2	5.0	100.0
Total (*N*)	40	100.0	

RATES

Terms such as birthrate, unemployment rate, and marriage rate are often used by social scientists and demographers and then quoted in the popular media to describe population trends. But what exactly are rates, and how are they constructed? A **rate** is obtained by dividing the number of actual occurrences in a given time period by the number of possible occurrences.

Rate: A number obtained by dividing the number of actual occurrences in a given time period by the number of possible occurrences.

$$\text{Rate} = \frac{f}{\text{Population}} \tag{2.5}$$

For example, we can use data from the American Community Survey to determine the 2017 poverty rate by dividing (actual occurrences) by the total population in 2017 (possible occurrences). The 2017 rate can be expressed as

$$\text{Poverty rate, 2017} = \frac{\text{Number of people in poverty in 2017}}{\text{Total population in 2017}}$$

Since 42,583,651 people were poor in 2017 and the number for the total population was 317,741,588, the poverty rate for 2017 is

$$\text{Poverty rate, 2017} = \frac{42{,}583{,}651}{317{,}741{,}588} = 0.13$$

We can thus conclude that the poverty rate in 2017 was 13% (.13 × 100). This means that for every 1,000 people, 130 were poor according to the American

Community Survey definition. Rates are often expressed as rates per thousand or hundred thousand to eliminate decimal points and make the number easier to interpret.

The preceding poverty rate can be referred to as a crude rate because it is based on the total population. Rates can be calculated on the general population or on a more narrowly defined select group. For instance, poverty rates are often given for the number of people who are 18 years or younger—highlighting how our young are vulnerable to poverty. The poverty rate for those 18 years or younger is as follows:

$$\text{Poverty rate for those 18 years or younger, 2017} = \frac{13{,}353{,}202}{72{,}452{,}925} = 0.18$$

We could even take a look at the poverty rate for older Americans:

$$\text{Poverty rate for those 65 years of age or older, 2017} = \frac{4{,}581{,}772}{49{,}500{,}479} = 0.09$$

LEARNING CHECK 2.5

Law enforcement agencies routinely record crime rates (the number of crimes committed relative to the size of a population), arrest rates (the number of arrests made relative to the number of crimes reported), and conviction rates (the number of convictions relative to the number of cases tried). What other variables can be expressed as rates?

BIVARIATE TABLES

Cross-tabulation is a technique for analyzing the relationship between two variables (an independent and a dependent variable) that have been organized in a table. A cross-tabulation is a type of **bivariate analysis**, a method designed to detect and describe the relationship between two nominal or ordinal variables. We demonstrate not only how to detect whether two variables are associated but also how to determine the strength of the association and, when appropriate, its direction in Chapter 8 The Chi-Square Test and Measures of Association.[5]

Cross-tabulation: A technique for analyzing the relationship between two variables that have been organized in a table.

Bivariate analysis: A statistical method designed to detect and describe the relationship between two variables.

Bivariate table: A table that displays the distribution of one variable across the categories of another variable.

HOW TO CONSTRUCT A BIVARIATE TABLE

A **bivariate table** displays the distribution of one variable across the categories of another variable. It is obtained by classifying cases based on their joint scores on two nominal or ordinary variables. It can be thought of as a series of frequency distributions joined to make one table. The data in Table 2.11 represent a sample of General Social Survey (GSS) respondents by race and whether they own or rent their home (in this case, both variables are nominal-level measurements).

Table 2.11 Race and Home Ownership for 20 GSS Respondents

Respondent	Race	Home Ownership
1	Black	Own
2	Black	Own
3	White	Rent
4	White	Rent
5	White	Own
6	White	Own
7	White	Own
8	Black	Rent
9	Black	Rent
10	Black	Rent
11	White	Own
12	White	Own
13	White	Rent
14	White	Own
15	Black	Rent
16	White	Own
17	Black	Rent
18	White	Rent
19	Black	Own
20	Black	Rent

Table 2.12 Home Ownership by Race (Absolute Frequencies), GSS

	Race: Black	Race: White	
Own	3	7	10
Rent	6	4	10
	9	11	20

(Row total): 10, 10 · Total cases (*N*): 20 · (Column total): 9, 11

To make sense out of these data, we must first construct the table in which these individual scores will be classified. In Table 2.12, the 20 respondents have been classified according to joint scores on race and home ownership.

The table has the following features typical of most bivariate tables:

1. The table's title is descriptive, identifying its content in terms of the two variables.

2. It has two dimensions, one for race and one for home ownership. The variable *home ownership* is represented in the rows of the table, with one row for owners and another for renters. The variable *race* makes up the columns of the table, with one column for each racial group. A table may have more columns and more rows, depending on how many categories the variables represent. For example, had we included a group of Latinx people, there would have been three columns (not including the row total column). Usually, the independent variable is the **column variable** and the dependent variable is the **row variable**.

3. The intersection of a row and a column is called a **cell**. For example, the two individuals represented in the upper left cell are blacks who are also home owners.

4. The column and row totals are the frequency distribution for each variable, respectively. The column total is the frequency distribution for *race*, and the row total is for *home ownership*. Row and column totals are sometimes called **marginals**. The total number of cases (*N*) is the number reported at the intersection of the row and column totals. (These elements are all labeled in the table.)

Column variable: A variable whose categories are the columns of a bivariate table.

Row variable: A variable whose categories are the rows of a bivariate table.

Cell: The intersection of a row and a column in a bivariate table.

Marginals: The row and column totals in a bivariate table.

5. The table is a 2 × 2 table because it has two rows and two columns (not counting the marginals). We usually refer to this as an $r \times c$ table, in which r represents the number of rows and c the number of columns. Thus, a table in which the row variable has three categories and the column variable has two categories would be designated as a 3 × 2 table.

6. The source of the data should also be clearly noted in a source note to the table.

LEARNING CHECK 2.6

Examine Table 2.12. Make sure you can identify all the parts just described and that you understand how the numbers were obtained. Can you identify the independent and dependent variables in the table? You will need to know this to convert the frequencies to percentages.

HOW TO COMPUTE PERCENTAGES IN A BIVARIATE TABLE

To compare home ownership status for blacks and whites, we need to convert the raw frequencies to percentages because the column totals are not equal. Percentages are especially useful for comparing two or more groups that differ in size. There are two basic rules for computing and analyzing percentages in a bivariate table:

1. Calculate percentages within each category of the independent variable.
2. Interpret the table by comparing the percentage point difference for different categories of the independent variable.

Calculating Percentages Within Each Category of the Independent Variable

The first rule means that we have to calculate percentages within each category of the variable that the investigator defines as the independent variable. When the independent variable is arrayed in the columns, we compute percentages within each column separately. The frequencies within each cell and the row marginals are divided by the total of the column in which they are located, and the column totals should sum to 100%. When the

Table 2.13 Home Ownership by Race (in Percentages)			
Home Ownership	Race		Total
	Black	White	
Own	33%	64%	50%
Rent	67%	36%	50%
Total	100%	100%	100%
(*N*)	(9)	(11)	(20)

independent variable is arrayed in the rows, we compute percentages within each row separately. The frequencies within each cell and the column marginals are divided by the total of the row in which they are located, and the row totals should sum to 100%.

In our example, we are interested in *race* as the independent variable and in its relationship with *home ownership*. Therefore, we are going to calculate percentages by using the column total of each racial group as the base of the percentage. For example, the percentage of black respondents who own their homes is obtained by dividing the number of black home owners by the total number of blacks in the sample.

Table 2.13 presents percentages based on the data in Table 2.12. Notice that the percentages in each column add up to 100%, including the total column percentages. Always show the *N*s that are used to compute the percentages—in this case, the column totals.

Comparing the Percentages Across Different Categories of the Independent Variable

The second rule tells us to compare how home ownership varies between blacks and whites. Comparisons are made by examining differences between percentage points across different categories of the independent variable. Some researchers limit their comparisons to categories with at least a 10–percentage point difference. In our comparison, we can see that there is a 31–percentage point difference between the percentage of white home owners (64%) and black home owners (33%). In other words, in this group, whites are more likely to be home owners than blacks.[5] Therefore, we can conclude that one's race appears to be associated with the likelihood of being a home owner.

Note that the same conclusion would be drawn had we compared the percentage of black and white renters. However, since the percentages of home owners and renters within each racial group sum to 100%, we need to make only one comparison. In fact, for any 2 × 2 table, only one comparison needs to be made to interpret the table. For a larger table, more than one comparison can be made and used in interpretation.

LEARNING CHECK 2.7

Practice constructing a bivariate table. Use Table 2.11 to create a percentage bivariate table. Compare your table with Table 2.12. Did you remember all the parts? Are your calculations correct? If not, go back and review this section.

GRAPHIC PRESENTATION OF DATA

You have probably heard that "a picture is worth a thousand words." The same can be said about statistical graphs because they summarize hundreds or thousands of numbers. Graphs communicate information visually, rather than in words or numbers, and are often used in news stories, research reports, and government documents. Information that is presented graphically may seem more accessible than the same information when presented in frequency distributions or in other tabular forms.

In this section, you will learn about some of the most commonly used graphical techniques. We concentrate less on the technical details of how to create graphs and more on how to choose the appropriate graphs to make statistical information coherent. We also focus on how to interpret graphically presented information.

The particular story we tell here is that of the elderly in the United States and throughout the world. Demographers predict that over the next several decades, the U.S. overall population growth will be among middle-aged and older Americans, what demographers have referred to as the graying of America. "Population aging is a long-range trend that will characterize our society as we continue into the 21st century. It is a force we all will cope with for the rest of our lives," warns gerontologist Harry Moody.[7]

The different types of graphs demonstrate the many facets and challenges of our aging society. People have tended to talk about seniors as if they were a homogeneous group, but the different graphical techniques illustrate the wide variation in economic characteristics, living arrangements, and family status among people aged 65 years and older.

Here we focus on those graphical techniques most widely used in the social sciences. The first two, (1) the pie chart and (2) bar graph, are appropriate for nominal and ordinal variables. The next two, (3) histograms and (4) line graphs, are used with interval-ratio variables. We also discuss time-series charts. Time-series charts are used to show how some variables change over time.

THE PIE CHART

The elderly population of the United States is racially heterogeneous. As the data in Table 2.14 show, of the total 47,732,389 elderly (defined as persons 65 years and older) in 2013–2017, the two largest racial groups were whites (83.5%) and blacks (8.9%).

A **pie chart** shows the differences in frequencies or percentages among the categories of a nominal or an ordinal variable. The categories are displayed as segments of a circle whose pieces add up to 100% of the total frequencies. The pie chart shown in Figure 2.1 displays the same information that Table 2.14 presents (notice that due to rounding, the percentages in Table 2.14 do not add up to 100%). Although you can inspect these data in Table 2.14, you can interpret the information more easily by seeing it presented in the pie chart in Figure 2.1.

Pie chart: A graph showing the differences in frequencies or percentages among categories of a nominal or an ordinal variable. The categories are displayed as segments of a circle whose pieces add up to 100% of the total frequencies.

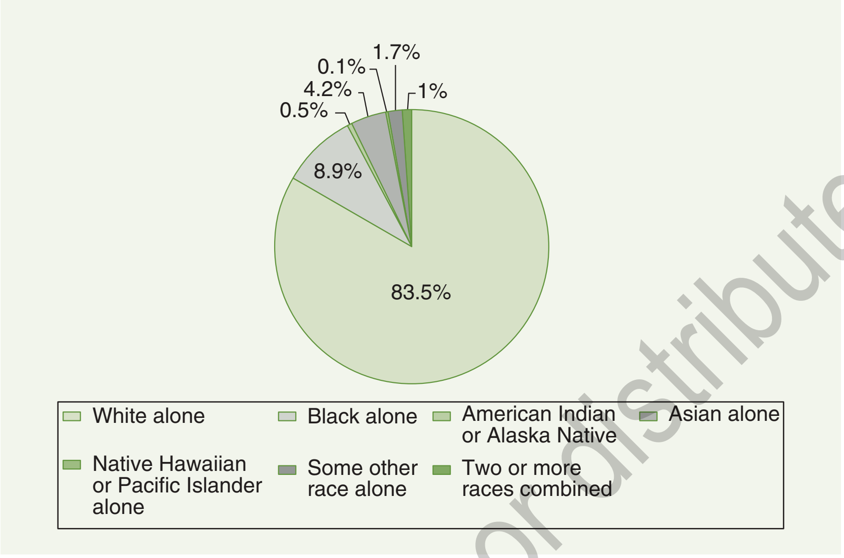
Did you notice that the percentages for several of the racial groups are 4.2% or less? It might be better to combine categories—American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, and some other race—into an “other races” category. This will leave us with four distinct categories: (1) white, (2) black, (3) two or more races, and (4) other. The revised pie chart is presented in Figure 2.2. We can highlight the diversity of the elderly population by “exploding” the pie chart, moving the nonwhite segments representing these groups slightly outward to draw them to the viewer’s attention. This also highlights the largest slice of the pie chart—white elderly comprised 83.5% of the U.S. elderly population in 2013–2017.

Table 2.14 Five-Year Estimates of the U.S. Population 65 Years and Over by Race, 2013–2017

Race	Percentage (%)
White alone	83.5
Black alone	8.9
American Indian or Alaska Native	0.5
Asian alone	4.2
Native Hawaiian or Pacific Islander alone	0.1
Some other race alone	1.7
Two or more races combined	1.0
Total	99.9

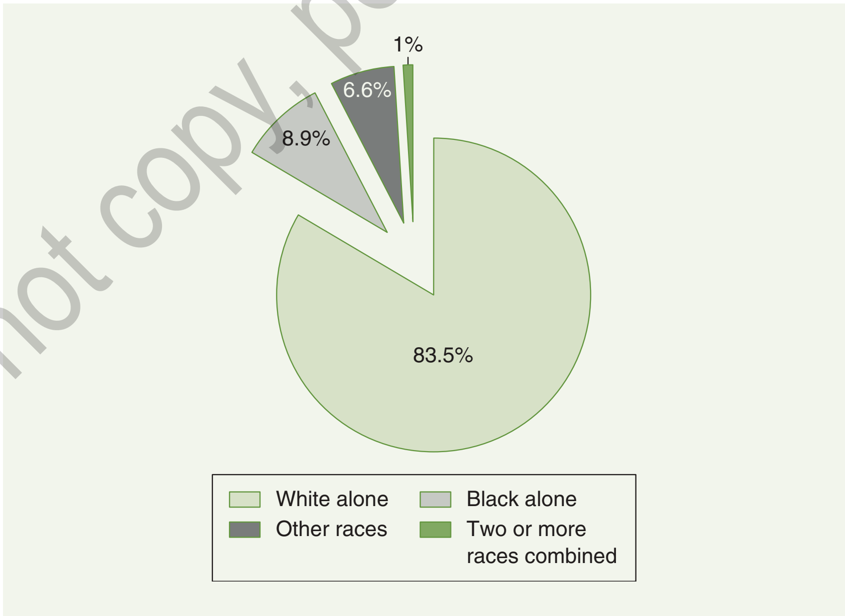
Source: U.S. Census Bureau, *American Fact Finder*, Table S0103, 2017.

Figure 2.1 Five-Year Estimates of the U.S. Population 65 Years and Over by Race, 2013–2017



Source: U.S. Census Bureau, *American Fact Finder*, Table S0103, 2017.

Figure 2.2 Five-Year Estimates of the U.S. Population 65 Years and Over by Race, 2013–2017



Source: U.S. Census Bureau, *American Fact Finder*, Table S0103, 2017.

THE BAR GRAPH

The **bar graph** provides an alternative way to graphically present nominal or ordinal data. It shows the differences in frequencies or percentages among categories of a nominal or an ordinal variable. The categories are displayed as rectangles of equal width with their height proportional to the frequency or percentage of the category.

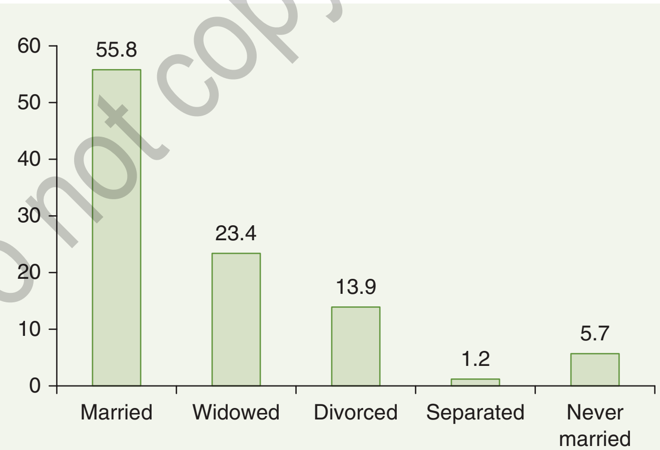
Bar graph: A graph showing the differences in frequencies or percentages among categories of a nominal or an ordinal variable. The categories are displayed as rectangles of equal width with their height proportional to the frequency or percentage of the category.

Let's illustrate the bar graph with an overview of the marital status of the elderly. Figure 2.3 is a bar graph displaying the percentage distribution of persons 65 years old and over by marital status in 2017. This chart is interpreted similar to a pie chart except that the categories of the variable are arrayed along the horizontal axis (sometimes referred to as the *X*-axis) and the percentages along the vertical axis (sometimes referred to as the *Y*-axis). This bar graph is easily interpreted: It shows that in 2017, the majority of the elderly population were married. Specifically, 55.8% were married, 23.4% were widowed, 13.9% divorced, 1.2% separated, and 5.7% never married.

Construct a bar graph by first labeling the categories of the variables along the horizontal axis. For these categories, construct rectangles of equal width, with the height of each proportional to the frequency or percentage of the category. Note that a space separates each of the categories to make clear that they are nominal categories.

Bar graphs are often used to compare one or more categories of a variable among different groups. Suppose we want to show how the patterns in marital status differ between men and women. The longevity of women is a major factor

Figure 2.3 Marital Status of U.S. Elderly (65 Years and Older), Percentages, 2017



Source: U.S. Census Bureau, *American Fact Finder*, Table S0103, 2017.

Figure 2.4 Marital Status of U.S. Elderly (65 Years and Older) by Gender (Percentages), 2017

Source: U.S. Census Bureau, *American Fact Finder*, Table S1201, 2017.

in the gender differences in marital and living arrangements.[8] Additionally, elderly widowed men are more likely to remarry than elderly widowed women.

Figure 2.4 compares the marital status for women and men 65 years and older in 2017. We can also construct bar graphs horizontally, with the categories of the variable arrayed along the vertical axis and the percentages or frequencies displayed on the horizontal axis, as displayed in Figure 2.4. This presentation allows for a side-by-side visual comparison. It shows that elderly women are more likely than elderly men to be widowed (33% vs. 11%), and elderly men are more likely to be married than elderly women (70% vs. 45%).

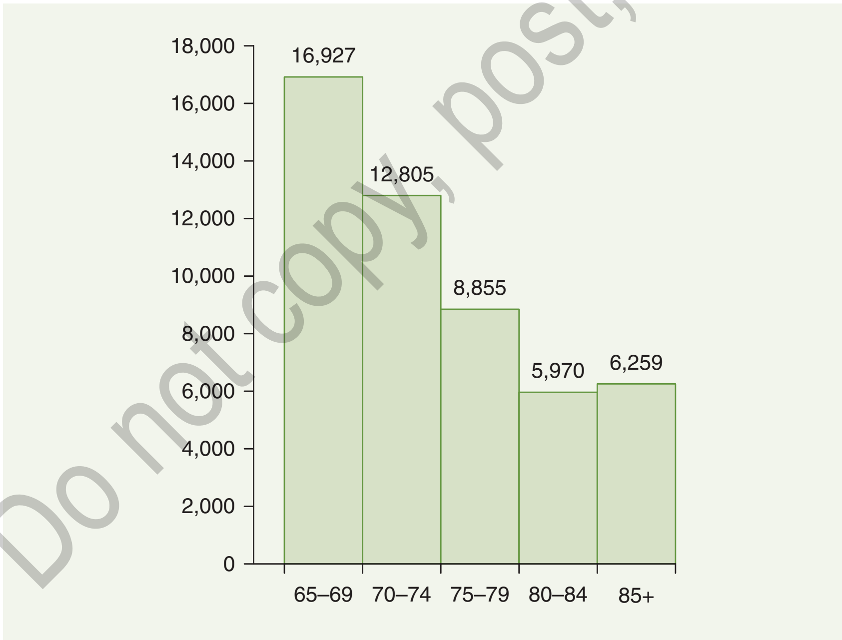
Histogram: A graph showing the differences in frequencies or percentages among categories of an interval-ratio variable. The categories are displayed as contiguous bars, with width proportional to the width of the category and height proportional to the frequency or percentage of that category.

THE HISTOGRAM

The **histogram** is used to show the differences in frequencies or percentages among categories of an interval-ratio or ordinal variable. The categories are displayed as contiguous bars, with width proportional to the width of the category and height proportional to the frequency or percentage of that category. A histogram looks very similar to a bar graph except that the bars are contiguous to each other (touching) and may not be of equal width. In a bar graph, the spaces between the bars visually indicate that the categories are separate. Examples of variables with separate categories are marital status (married, single), gender (male, female), and employment status (employed, unemployed). In a histogram, the touching bars indicate that the categories or intervals are ordered from low to high in a meaningful way. For example, the categories of the variables hours spent studying, age, and years of school completed are contiguous, ordered intervals.

Figure 2.5 is a histogram displaying the frequency distribution of the population 65 years and over by age. To construct the histogram, arrange the age intervals along the horizontal axis and the frequencies (or percentages) along the vertical axis. For each age category, construct a bar with the height corresponding to the frequency of the elderly in the population in that age category. The width of each bar corresponds to the number of years that the age interval represents. And, in histograms, the bar for each category is touching the bar associated with the category above and below. The area that each bar occupies tells us the number of individuals that falls into a given age interval. Note that the figure title includes the notation "numbers in thousands." You should multiply each reported frequency by 1,000. For example, the largest age category is 65–69 years with 16,927,000 (16,927 × 1,000). The smallest age group is 80–84 years with 5,970,000. The total number of elderly 65 years and over can be found by summing all the reported frequencies.

Figure 2.5 Age Distribution of U.S. Elderly (65 Years and Older), 2017 (Numbers in Thousands)



Source: U.S. Census Bureau, *American Fact Finder*, Table S0101, 2017.

Note: Ages were collapsed into categories in this example for visual purposes only. In general, histograms should be displayed with interval-ratio data that haven't been collapsed.

THE LINE GRAPH

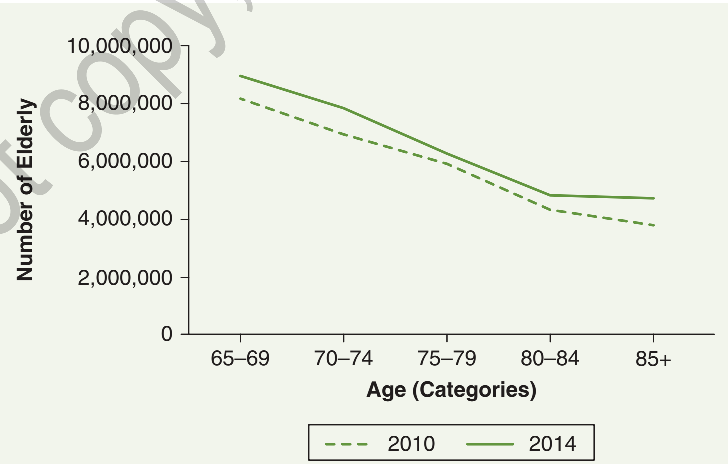
The elderly population is growing worldwide in both developed and developing countries. In 1994, 30 nations had elderly populations of at least 2 million; demographic projections indicate that there will be 55 such nations by 2020. Japan is one of the nations that is experiencing dramatic growth of its elderly population. Figure 2.6 is a line graph displaying the elderly population of Japan for 2010 and 2014.

Line graph: A graph showing the differences in frequencies or percentages among categories of an interval-ratio variable. Points representing the frequencies of each category are placed above the midpoint of the category and are joined by a straight line.

The **line graph** is another way to display interval-ratio distributions; it shows the differences in frequencies or percentages among categories of an interval-ratio variable. Compared with histograms, line graphs are better suited for comparing how a variable is distributed across two or more groups or across two or more time periods, as we've done in Figure 2.6. Points representing the frequencies of each category are placed above the midpoint of the category and are joined by a straight line. Notice that in Figure 2.6, the age intervals are arranged on the horizontal axis and the frequencies along the vertical axis. Instead of using bars to represent the frequencies, however, points representing the frequencies of each interval are placed above the midpoint of the intervals. Adjacent points are then joined by straight lines.

Figure 2.6 shows how Japan's population of age 65 and over increased from 2010 to 2014. According to projections, Japan's oldest-old population—those 80 years or older—is projected to grow rapidly, from about 4.8 million (less than 4% of the total population) in 2014 to 10.8 million (8.9%) by 2020 (not depicted in the figure). This projected rise has already led to a reduction in

Figure 2.6 Population of Japan, Age 65 and Above, 2010 and 2014



Source: United Nations, Statistics Division, *Population by Age, Sex, and Urban/Rural Residence*, 2015. Retrieved from http://data.un.org/Data.aspx?d=POP&f=tableCode%3A22

retirement benefits and other adjustments to prepare for the economic and social impact of a rapidly aging society.[9]

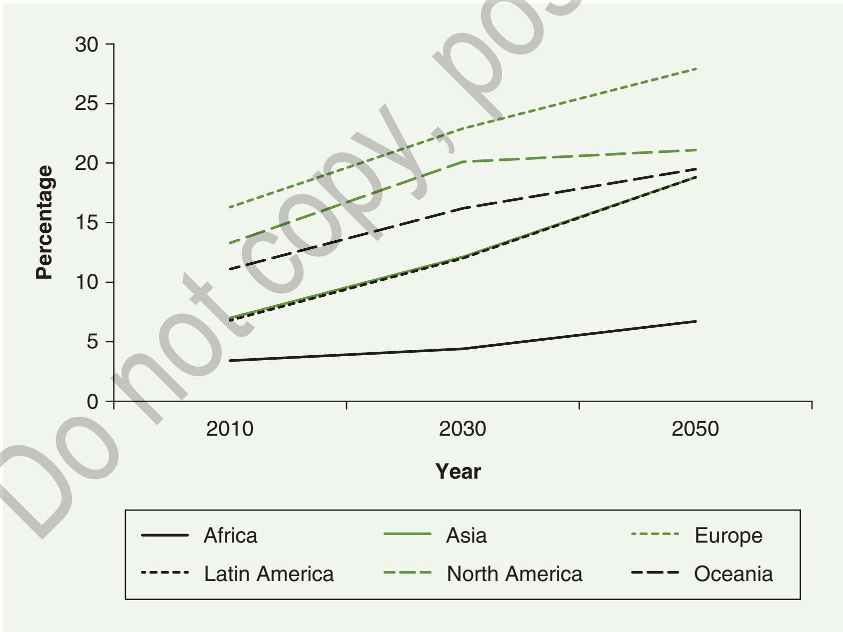
THE TIME-SERIES CHART

We are often interested in examining how some variables change over time. For example, we may be interested in showing changes in the labor force participation of Latinas over the past decade, changes in the public's attitude toward same-sex marriage, or changes in divorce and marriage rates. A **time-series chart** displays changes in a variable at different points in time. It involves two variables: (1) *time*, which is labeled across the horizontal axis, and (2) another variable of interest whose values (frequencies, percentages, or rates) are labeled along the vertical axis. To construct a time-series chart, use a series of dots to mark the value of the variable at each time interval and then join the dots by a series of straight lines.

Time-series chart: A graph displaying changes in a variable at different points in time. It shows time (measured in units such as years or months) on the horizontal axis and the frequencies (percentages or rates) of another variable on the vertical axis.

Figure 2.7 shows a time series from 2010 to 2050 of the percentage of the total population that is 65 years or older (the percentages for 2030 and 2050 are

Figure 2.7 Percentage of Total Population 65 Years and Above for Selected World Regions, 2010, 2030, and 2050



Source: Loraine West, Samantha Cole, Daniel Goodkind, and Wan He, *65+ in the United States: 2010*, Current Population Report, P23–212, 2014.

projections, as reported by the U.S. Census Bureau) for selected world regions. This time series enables us to see clearly the increase in the elderly population worldwide. As we have already mentioned, these demographic changes will have significant social, political, and economic implications, capturing the attention of policy makers and social scientists.

LEARNING CHECK 2.8

How does the time-series chart differ from a line graph? The difference is that line graphs display frequency distributions of a single variable, whereas time-series charts display two variables. In addition, time is always one of the variables displayed in a time-series chart.

STATISTICS IN PRACTICE: FOREIGN-BORN POPULATION 65 YEARS AND OVER

In their 2014 report *65+ in America*, U.S. Census Bureau researchers Loraine West, Samantha Cole, Daniel Goodkind, and Wan He describe the foreign-born population aged 65 and over in a series of tables and graphs. We present several for your review.

Frequencies and percentages presented in Table 2.15 summarize three characteristics of the 5,000,000 foreign-born elderly. The majority of these older men and women entered the U.S. prior to 1990. Almost 73% were naturalized citizens in 2010. In the same year, the largest percentage lived in the West (36%), followed by the South (29%).

Figure 2.8 is a pie chart, presenting one variable—*world region of birth*. We learn that the majority of the foreign-born elderly originally came from Latin America (37%), Asia (29%), and Europe (28%). The bar graph (Figure 2.9) presents the percentage of foreign-born elderly from each world region by their period of entry. Prior to 1990 and during 2000–2010, the largest percentage of foreign-born elderly came from Latin America and the Caribbean. However, from 1990 to 1999, the largest percentage of foreign-born elderly emigrated from Asia.

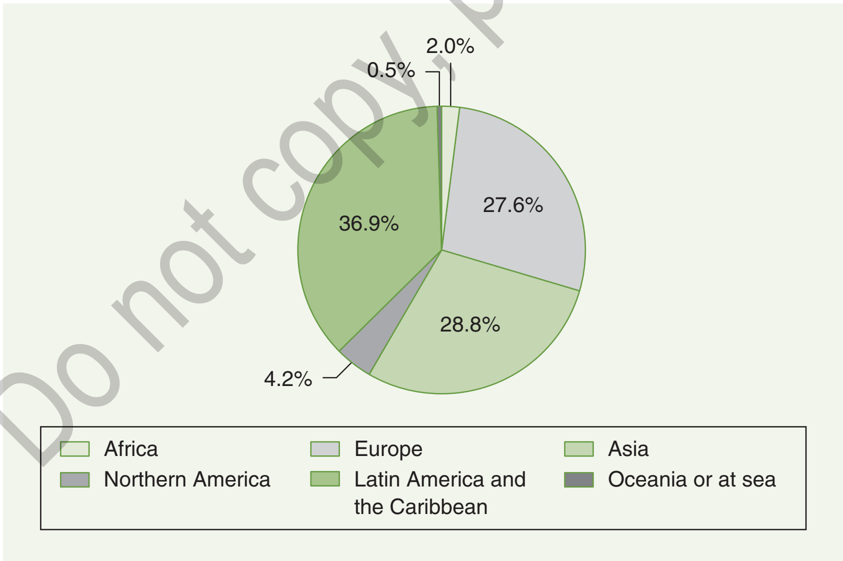
Table 2.15 Foreign-Born Population Aged 65 and Over by Period of Entry, Citizenship Status, and Region, 2010

Characteristic	Population (in Thousands)	Percentage (%)
Total	4,963	100
Period of entry		
Prior to 1990	3,769	76

Characteristic	Population (in Thousands)	Percentage (%)
1990 to 1999	644	13.0
2000 to 2010	550	11.1
Citizenship status		
Naturalized citizen	3,582	72.2
Not a U.S. citizen	1,381	27.8
Region		
Northwest	1,232	24.8
Midwest	504	10.1
South	1,442	29.1
West	1,784	36.0

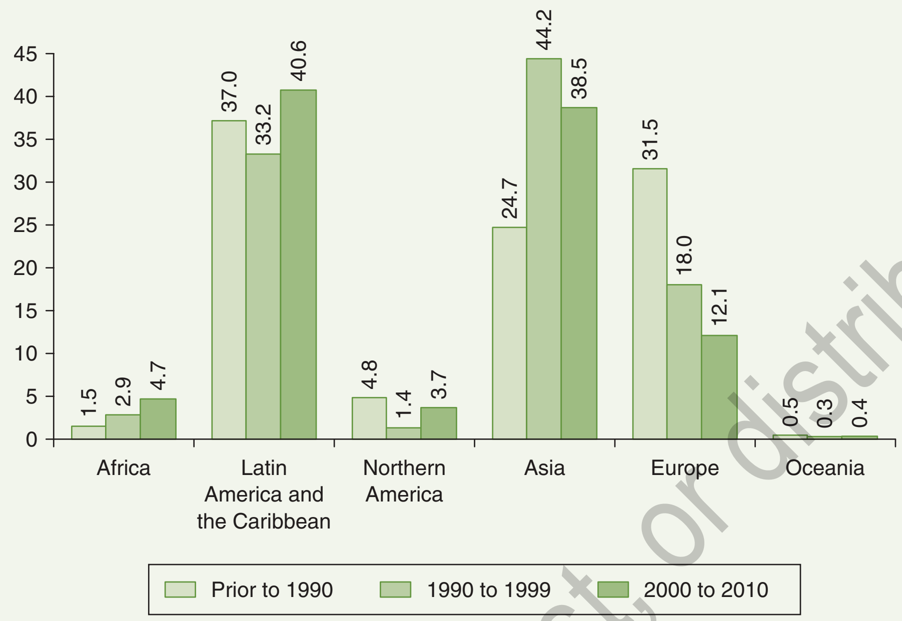
Source: Loraine West, Samantha Cole, Daniel Goodkind, and Wan He, *65+ in the United States: 2010*, Current Population Report, P23–212, 2014.

Figure 2.8 Foreign-Born Population Aged 65 Years and Over by World Region of Birth, 2010



Source: Loraine West, Samantha Cole, Daniel Goodkind, and Wan He, *65+ in the United States: 2010*, Current Population Report, P23–212, 2014.

Figure 2.9 Foreign-Born Population Aged 65 and Over by World Region of Birth and Period of Entry, 2010



Source: Loraine West, Samantha Cole, Daniel Goodkind, and Wan He, *65+ in the United States: 2010*, Current Population Report, P23–212, 2014.

A CLOSER LOOK 2.1

A Cautionary Note: Distortions in Graphs

In this chapter, we have seen that statistical graphs can give us a quick sense of the main patterns in the data. However, graphs not only can quickly inform us but also can also quickly deceive us. Because we are often more interested in general impressions than in detailed analyses of the numbers, we are more vulnerable to being swayed by distorted graphs. Edward Tufte in his 1983 book *The Visual Display of Quantitative Information* not only demonstrates the advantages of working with graphs but also offers a detailed discussion of some of the pitfalls in the application and interpretation of graphics.[10]

Probably the most common distortions in graphical representations occur when the distance along the vertical or horizontal axis is altered either by not using 0 as the baseline (as demonstrated in Figure 2.10a,b) or in relation to the other axis. Axes may be stretched or shrunk to create any desired result to exaggerate or disguise a pattern in the data. In Figure 2.10a,b, 2015 international data on female representation in national parliaments are presented. Without altering the data in any way, notice how the difference between the countries is exaggerated by using 30 as a baseline (as in Figure 2.10b).

Remember to interpret the graph in the context of the numerical information the graph represents.

Figure 2.10 Female Representation in National Parliaments, 2015: (a) Using 0 as the Baseline and (b) Using 30 as the Baseline

Source: Inter-Parliamentary Union, *Women in National Parliaments*, 2016. Retrieved from www.ipu.org/wmn-e/classif.htm

DATA AT WORK

Spencer Westby: Senior Editorial Analyst

Photo courtesy of Kurt Taylor Gaubatz

As a senior editorial analyst for an academic publishing company, Spencer uses data to examine the driving factors behind a successful journal publication. He tracks journal usage article submissions and citations to determine ways to improve reader outreach and revenue.

"I use bivariate tables and pie charts more than anything else as they are the quickest way to display relationships between variables. It is a simple way to display complex data for the publishing editors I am working with. Having worked at this position for a while now, I find it is just as important to

(Continued)

(Continued)

make sure the data and analysis are understandable by someone who does not know statistics well. The work needs to be beneficial to the whole business."

Spencer was introduced to this work through an internship. He hoped to gain career experience for college, but during his internship, he discovered an interest in publishing. "It was especially interesting to see how this industry worked from the inside compared to how simple it seemed on the outside as a student."

"If you are interested in a career using quantitative research, my biggest advice would be to make sure that it is something that you at least find interesting," says Spencer. "Working with numbers all day can be fairly demanding mental work if you are not ready for it. However, an inquisitive mind can turn any analytical project into a puzzle waiting to be solved. A career utilizing statistics is very satisfying for those who pursue it with creative minds."

MAIN POINTS

- The most basic way to organizing data is to classify the observations into a frequency distribution—a table that reports the number of observations that fall into each category of the variable being analyzed. A frequency distribution for a single variable is referred to as univariate table.

- Constructing a frequency distribution is usually the first step in the statistical analysis of data. To obtain a frequency distribution for nominal and ordinal variables, count and report the number of cases that fall into each category of the variable along with the total number of cases (*N*). To construct a frequency distribution for interval-ratio variables that have a wide range of values, first combine the scores into a smaller number of groups—known as class intervals—each containing a number of scores.

- Proportions and percentages are relative frequencies. To construct a proportion, divide the frequency (*f*) in each category by the total number of cases (*N*). To obtain a percentage, divide the frequency (*f*) in each category by the total number of cases (*N*) and multiply by 100.

- Percentage distributions are tables that show the percentage of observations that fall into each category of the variable. Percentage distributions are routinely added to almost any frequency table and are especially important if comparisons between groups are to be considered.

- Cumulative frequency distributions allow us to locate the relative position of a given score in a distribution. They are obtained by adding to the frequency in each category the frequencies of all the categories below it.

- Cumulative percentage distributions have wider applications than cumulative frequency distributions. A cumulative percentage distribution is constructed by adding to the percentages in each category the percentages of all the categories below it.
- A rate is a number that expresses raw frequencies in relative terms. A rate can be calculated as the number of actual occurrences in a given time period divided by the number of possible occurrences for that period. Rates are often multiplied by some power of 10 to eliminate decimal points and make the number easier to interpret.
- The bivariate table displays the distribution of one variable across the categories of another variable. It is obtained by classifying cases based on their joint scores for two variables.
- Percentaging bivariate tables are used to examine the relationship between two variables that have been organized in a bivariate table. The percentages are always calculated within each category of the independent variable.
- A pie chart shows the differences in frequencies or percentages among categories of a nominal or an ordinal variable. The categories of the variable are segments of a circle whose pieces add up to 100% of the total frequencies.
- A bar graph shows the differences in frequencies or percentages among categories of a nominal or an ordinal variable. The categories are displayed as rectangles of equal width with their height proportional to the frequency or percentage of the category.
- Histograms display the differences in frequencies or percentages among categories of interval-ratio variables. The categories are displayed as contiguous bars with their width proportional to the width of the category and height proportional to the frequency or percentage of that category.
- A line graph shows the differences in frequencies or percentages among categories of an interval-ratio variable. Points representing the frequencies of each category are placed above the midpoint of the category (interval). Adjacent points are then joined by a straight line.
- A time-series chart displays changes in a variable at different points in time. It displays two variables: (1) *time*, which is labeled across the horizontal axis, and (2) another variable of interest whose values (e.g., frequencies, percentages, or rates) are labeled along the vertical axis.

KEY TERMS

bar graph 49
bivariate analysis 41
bivariate table 41
cell 43
column variable 43
cross-tabulation 41
cumulative frequency distribution 38
cumulative percentage distribution 39
frequency distribution 27
histogram 50
line graph 52
marginals 43
percentage 30
percentage distribution 30
pie chart 47
proportion 29
rate 40
row variable 43
time-series chart 53
univariate frequency table 28

DIGITAL RESOURCES

Access key study tools at https://edge.sagepub.com/ssdsess4e

- eFlashcards of the glossary terms
- Datasets and codebooks
- SPSS and Excel walk-through videos
- SPSS and Excel demonstrations and problems to accompany each chapter
- Appendix F: Basic Math Review

CHAPTER EXERCISES

1. Suppose you surveyed 30 people and asked them whether they are white (W) or nonwhite (N) and how many traumas (serious accidents, rapes, or crimes) they have experienced in the past year. You also asked them to tell you whether they perceive themselves as being in the upper, middle, working, or lower class. Your survey resulted in the raw data presented in the following table:
   a. Identify the level of measurement for each variable.
   b. Construct raw frequency tables for race.
   c. What proportion of the 30 individuals is nonwhite? What percentage is white?

Race	Class	Trauma	Race	Class	Trauma
W	L	1	W	W	0
W	M	0	W	M	2

Race	Class	Trauma	Race	Class	Trauma
W	M	1	W	W	1
N	M	1	W	W	1
N	L	2	N	W	0
W	W	0	N	M	2
N	W	0	W	M	1
W	M	0	W	M	0
W	M	1	N	W	1
N	W	1	W	W	0
N	W	2	W	W	0
N	M	0	N	M	0
N	L	0	N	W	0
W	U	0	N	W	1
W	W	1	W	W	0

Note: Race: W = white; N = nonwhite; Class: L = lower class; M = middle class; U = upper class; W = working class.

2. Using the data from Exercise 1, construct a frequency and percentage distribution for class.
   a. Which is the smallest perceived class group?
   b. Which two classes include the largest percentages of people?
3. Using the data from Exercise 1, construct a frequency distribution for trauma.
   a. What level of measurement is used for the trauma variable?
   b. Are people more likely to have experienced no traumas or only one trauma in the past year?
   c. What proportion has experienced one or more traumas in the past year?
4. Using the data from Exercise 1, construct appropriate graphs showing percentage distributions for race, class, and trauma.
5. GSS 2018 respondents were asked to describe how much confidence they had in the press. Results are provided in the following table for the percentage in each category by whom the respondent voted for in the 2016 U.S. presidential election. Do these data support the statement

that those who voted for Donald Trump have lower levels of confidence in the press than those who didn't vote for Trump? Why or why not?

	Trump (%)	Clinton (%)	Other (%)
A great deal	2.9	21.8	11.8
Only some	17.9	58.9	29.4
Hardly any	79.2	19.3	58.8
Total	100.0	100.0	100.0

6. How many hours per week do you spend on e-mail? Data are presented here for a GSS sample of 99 men and women, who each reported the number of hours they spent per week on e-mail.

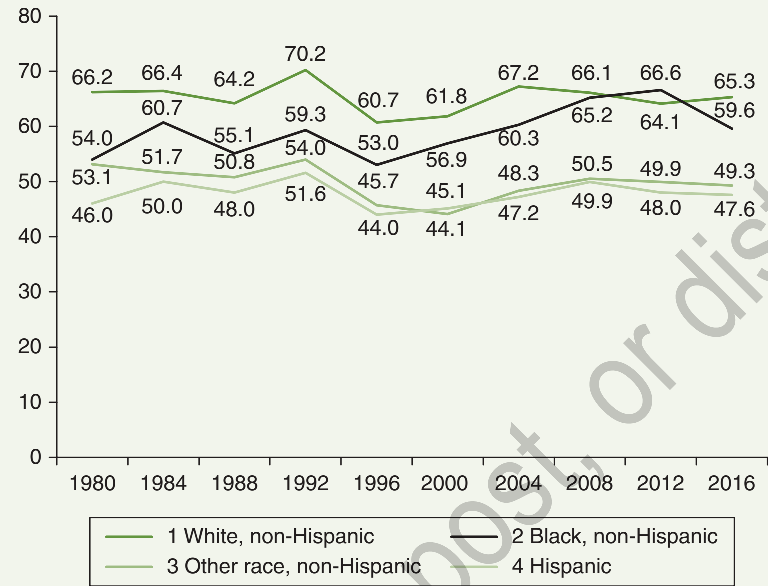
   a. Compute the cumulative frequency and cumulative percentage distribution for the data.
   b. What proportion of the sample spent 3 hours or less per week on e-mail?
   c. What proportion of the sample spent 6 or more hours per week on e-mail?
   d. Construct a graph that best displays these data. Explain why the graph you selected is appropriate for these data.

E-mail Hours per Week	Frequency
0	19
1	20
2	13
3	5
4	2
5	6
6	5
7	2
8	3
9	1
10 or more	23

7. The time-series chart shown below displays trends for presidential election voting rates by race and Hispanic origin for 1980–2016. U.S. Census Senior Sociologist Thom File noted how for the first time in the 2012 presidential

election, black voting rates exceeded the rates for non-Hispanic whites. However, in the 2016 presidential election, there was a remarkable shift in voting rates by race and Hispanic origin. Describe the variation in voting rates in 2016 for the four racial and Hispanic origin groups and then compare that variation with that presented for the four groups from 1980 onward.

Reported Voting Rates by Race and Hispanic Origin: 1980–2016



Source: Thom File, *Voting in America: A Look at the 2016 Presidential Election*, 2017. Retrieved from https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html

8. According to the Pew Research Center (2015), recent immigrants are better educated than earlier immigrants to the United States. The change was attributed to the availability of better education in each region or country of origin. The percentage of immigrants 25 years of age and older who completed at least high school is reported in this table for 1970 to 2013. Write a statement describing the change over time in the percentage who completed at least a high school degree.

	1970	1980	1990	2000	2013
Mexico	14	17	26	30	48
Other Central/ South America	52	57	53	60	66

(*Continued*)

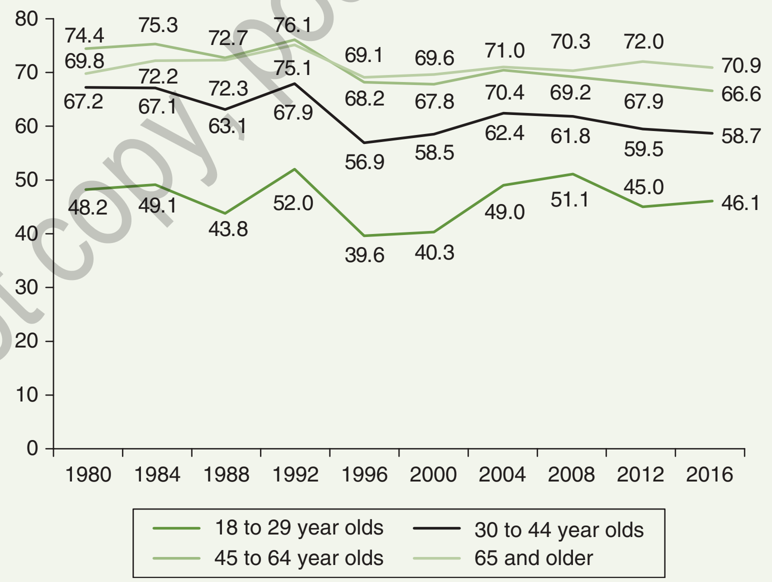
(Continued)

	1970	1980	1990	2000	2013
Asia	75	72	75	82	84
Europe	48	68	81	87	95
Caribbean	36	48	52	58	72
Africa	81	91	88	85	85

Source: Pew Research Center, *Modern Immigration Wave Brings 59 Million to U.S., Driving Population Growth and Change Through 2065*, 2015. Retrieved from https://www.pewresearch.org/hispanic/2015/09/28/modern-immigration-wave-brings-59-million-to-u-s-driving-population-growth-and-change-through-2065/

9. Older Americans are often described as more politically engaged than younger Americans. One measure of political engagement is election voting. In a 2017 report, U.S. Census senior sociologist Thom File presented the following time-series chart of voting rates by age in presidential elections from 1980 to 2016. Based on this time-series chart, are older Americans more likely to vote in the presidential election than younger Americans? (Before you answer, define which age groups are younger vs. older.)

Reported Voting Rates by Age: 1980–2016



Source: Thom File, *Voting in America: A Look at the 2016 Presidential Election*, 2017. Retrieved from https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html

10. The following cross-tabulation, based on GSS18SSDS-B data, examines the relationship between age (the independent variable) and presidential candidate one voted for in the 2016 U.S. presidential election (dependent variable). Notice we have recoded age into four categories: 18–29, 30–39, 40–49, and 50–59. Respondents of all other ages have been excluded from the analysis.

Vote Clinton or Trump * RE_AGE Crosstabulation

			RE_AGE				
			18–29	30–39	40–49	50–59	Total
Vote Clinton or Trump	Clinton	Count	65	96	77	78	316
		% within RE_AGE	65.7%	61.1%	53.8%	45.6%	55.4%
	Trump	Count	19	42	55	86	202
		% within RE_AGE	19.2%	26.8%	38.5%	50.3%	35.4%
	Other candidate (specify)	Count	12	17	10	5	44
		% within RE_AGE	12.1%	10.8%	7.0%	2.9%	7.7%
	Didn't vote for president	Count	3	2	1	2	8
		% within RE_AGE	3.0%	1.3%	0.7%	1.2%	1.4%
Total		Count	99	157	143	171	570
		% within RE_AGE	100.0%	100.0%	100.0%	100.0%	100.0%

a. Which is the independent variable?

b. What is the level of measurement for the independent variable and the dependent variable?

c. How would you describe the relationship between the two variables? Use percentages in your answer.

11. One of your classmates hypothesizes that people of color are far less likely to own their own home than white people. Use the following data, which draws from GSS18SSDS-B, to test your classmate's hypothesis.

Home Ownership	Race		Total
	White	Black	
Own or is buying	475	64	539
	67.7%	41.3%	63%

(*Continued*)

(Continued)

Home Ownership	Race		Total
	White	Black	
Pays rent	227	91	318
	32.3%	58.7%	37%
Total	702	155	857
	100%	100%	100%

a. Based on your classmate's argument, what is the dependent variable? The independent variable?

b. What percentage of those surveyed own their own home?

c. Using the percentages in the table, describe the relationship between race and home ownership.

12. Youth were asked in the Monitoring the Future (MTF) 2017 survey to report how often they were drunk in the past 12 months. Responses for 3,176 twelfth graders are reported by race.

Drunk in the past 12 Months	Race			Total
	Black	White	Hispanic	
None	359	1,081	480	1,920
1–2 times	63	342	125	530
3–5 times	24	168	50	242
6 or more times	31	389	64	484
Total	477	1,980	719	3,176

Calculate the percentages using *race* as the independent variable. Is there a relationship between race and frequency of drunkenness?

13. Drawing on International Social Survey Programme data, Tsui-o Tai and Janeen Baxter (2018) examined variation in household division of labor. The following table has been adapted from their analysis to show the relationship between perceived unfairness and frequency of housework disagreement for women in their sample.

a. What is the dependent variable?

b. How many people are included in this table?

c. What percentage of the sample felt housework divisions were unfair to women?

d. Is there a relationship between perceived unfairness and frequency of housework disagreement for women?

Housework Disagreement	Women's Perceptions of Housework Divisions, % (*N*)		
	Fair to Women	Unfair to Women	Total
Never	43.65% (1,393)	25.62% (1,013)	33.68% (2,406)
Rarely	32.09% (1,024)	30.16% (1,192)	31.02% (2,216)
Several times a year	13.19% (421)	17.56% (694)	15.61% (1,115)
Several times a month	7.80% (249)	17.28% (683)	13.05% (932)
Several times a week	3.27% (104)	9.39% (371)	6.65% (475)
Total	100% (3,191)	100% (3,953)	100% (7,144)

Source: Adapted from Tsui-o Tai and Janeen Baxter, "Perceptions of Fairness and Housework Disagreement: A Comparative Analysis," *Journal of Family Issues* 39, no. 8 (2018): 2461–2485 at p. 2471.