Chapter 2

The Organization and Graphic Presentation of Data

Chapter Learning Objectives

- Understanding how to construct and analyze frequency, percentage, and cumulative distributions
- Understanding how to calculate proportions and percentages
- Recognizing the differences in frequency distributions for nominal, ordinal, and interval-ratio variables
- Reading statistical tables in research literature
- Constructing and interpreting a pie chart, bar graph, histogram, line graph, and time-series chart
- Analyzing and interpreting charts and graphs in literature

As social researchers, we often have to deal with very large amounts of data. For example, in a typical survey, by the completion of your data collection phase you will have accumulated thousands of individual responses represented by a jumble of numbers. To make sense out of these data, you will have to organize and summarize them in some systematic fashion. The most basic method for organizing 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.

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

FREQUENCY DISTRIBUTIONS

Sweeping immigration proposals by the Bush administration and the U.S. Congress in 2006 and 2007 sparked nationwide debate about the status of immigrants in the United States. Globalization has fueled labor migration, particularly since the beginning of the 21st century. Workers migrate because of the promise of employment and higher standards of living than their home countries. Data reveal that the United States is the destination for many migrants.[1] 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 nearly 12% of the U.S. population or 37 million people are foreign born.[2] Today, one in eight U.S. residents is foreign born.[3] Immigrants 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.

Table 2.1 Frequency Distribution for Categories of World Region of Birth for Foreign-Born Population, 2008

World Region of Birth	*Frequency (f)*
Europe	4,644,000
Asia	9,978,000
Latin America	20,034,000
Other areas	2,602,000
Total (*N*)	37,258,000

Source: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 43.

Note that 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: "Categories of World 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, headed "Frequency (*f*)," tells the number of cases in each category as well as the total number of cases ($N = 37{,}258{,}000$). Note also that the source of the table is clearly identified in a source note. It tells us that the data are from a U.S. census report and the data come from a 2010 report. In general, the source of data for a table should appear as a source note unless it is clear from the general discussion of the data.

What can you learn from the information presented in Table 2.1? The table shows that as of 2008, approximately 37 million people were classified as foreign born. Out of this group, the majority, about 20 million people, were from Latin America, 10 million were from Asia, followed by 4.6 million from Europe.

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):

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

where

f = frequency

N = total number of cases

Thus, the proportion of foreign born originally from Latin America is

$$\frac{20{,}034{,}000}{37{,}258{,}000} = 0.54$$

The proportion of foreign born who were originally from Asia is

$$\frac{9{,}978{,}000}{37{,}258{,}000} = 0.27$$

The proportion of foreign born who were originally from Europe is

$$\frac{4{,}644{,}000}{37{,}258{,}000} = 0.12$$

And finally, the proportion of foreign born who were originally from other areas is

$$\frac{2{,}602{,}000}{37{,}258{,}000} = 0.07$$

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

$$0.54 + 0.27 + 0.12 + 0.07 = 1.00$$

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 Asia can be calculated as

$$0.27(37{,}258{,}000) = 10{,}059{,}660$$

Note that the obtained frequency differs somewhat from the actual frequency of 9,978,000. 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:

$$0.267808255945032(37{,}258{,}000) = 9{,}978{,}000$$

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

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, simply 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 Asia is

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

The percentage of respondents who were originally from Latin America is

$$0.54(100) = 54\%$$

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

✓ *Learning Check*

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

PERCENTAGE DISTRIBUTIONS

Percentages are usually displayed as percentage distributions. A **percentage distribution** is a table showing the percentage of observations falling into each 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 $N = 37{,}258{,}000$.

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

Table 2.2 Frequency and Percentage Distributions for Categories of World Region of Birth for Foreign Born, 2008

World Region of Birth	*Frequency (f)*	*Percentage*
Europe	4,644,000	12.5
Asia	9,978,000	26.8
Latin America	20,034,000	53.8
Other areas	2,602,000	7.0
Total (*N*)	37,258,000	100.1

Source: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 43.

COMPARISONS

In Table 2.2, we illustrated that there are four primary places of origin for foreign born in the United States. These distinctions help us understand the specific characteristics and backgrounds of each group. We can resist the temptation to group all foreign born in one category and ask, for instance, is one group more educated than another? Is one group younger than the other groups?

The decision to consider these groups separately or to pool them depends to a large extent on our research question. For instance, we know that in 2008, 16.5% of the foreign-born population were living in poverty.[4] Among the foreign born, poverty rates were highest among those from Latin America (21%) and lowest among those from Europe (9%). What do these figures tell us about the demographic characteristics of foreign borns? Of Latin American foreign borns? Of European foreign borns? To answer these questions and determine whether the two categories of region of birth have markedly different social characteristics, we need to *compare* them.

As students, as social scientists, and even as consumers, we are frequently faced with problems that call for some way to make a clear and valid comparison. For example, in 2008, 22.8% of children lived with only their mothers.[5] Is this figure high or low? In 2008, 26% of those 18 years of age or older had never been married.[6] Does this reflect a change in the American family? In each of these cases, comparative information is required to answer the question and reach a conclusion.

STATISTICS IN PRACTICE: LABOR FORCE PARTICIPATION AMONG LATINOS

Very often, we are interested in comparing two or more groups that differ in size. Percentages are especially useful for making such comparisons. For example, we know that differences in socioeconomic status mark divisions between populations, indicating differential access to economic opportunities. Labor participation (either employed or seeking employment) is an important indicator of access to economic opportunities and is strongly associated with socioeconomic status. Table 2.3 shows the raw frequency distributions for the variable *labor force participation* for three categories of Latino (or Hispanic) identity.

Table 2.3 Employment Status of the Civilian Population, Selected Hispanic Categories, 2008

Employment Status	*Mexican*	*Puerto Rican*	*Cuban*
Employed	12,931,000	1,634,000	841,000
Unemployed	1,078,000	188,000	57,000
Not in labor force	6,465,000	1,032,000	525,000
Total (*N*)	20,474,000	2,854,000	1,423,000

Source: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 39.

Which group has the highest relative number of persons who are not in the labor force? Because of the differences in the population sizes of the three groups, this is a difficult question to answer based on only the raw frequencies. To make a valid comparison, we have to compare the percentage distributions for all three groups. These are presented in Table 2.4. Note that the percentage distributions make it easier to identify differences between the groups. Compared with Puerto Ricans and Cubans, Mexicans have the highest percentage employed in the labor force (63% vs. 57% and 59%). Among the three groups, Cubans have the lowest percentage (4% vs. 5.3% and 6.6%) of persons who are unemployed.

Table 2.4 Employment Status of the Civilian Population, Selected Hispanic Categories, 2008

Employment Status	*Mexican (%)*	*Puerto Rican (%)*	*Cuban (%)*
Employed	63.2	57.2	59.1
Unemployed	5.3	6.6	4.0
Not in labor force	31.6	36.2	36.9
Total	100.1	100.0	100.0
(*N*)	20,474,000	2,854,000	1,423,000

Source: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 39.

✓ *Learning Check*

Examine Table 2.4 and answer the following questions: What is the percentage of Mexicans who are employed? What is the base (N) *for this percentage? What is the percentage of Cubans who are not in the labor force? What is the base* (N) *for this percentage?*

Whenever one group is compared with another, the most meaningful conclusions can usually be drawn based on comparison of the relative frequency distributions. In fact, we are seldom interested in a single distribution. Most interesting questions in the social sciences are about differences between two or more groups.[7] The finding that the labor force participation patterns vary among different

Latino groups raises doubt about whether Latinos can be legitimately regarded as a single, relatively homogeneous ethnic group. Further analyses could examine *why* differences in Latino identity are associated with differences in labor force participation patterns. Other variables that explain these differences could be identified. These kinds of questions can be answered using more complex multivariate statistical techniques that involve more than two variables. The comparison of percentage distributions is an important foundation for these more complex techniques.

THE CONSTRUCTION OF FREQUENCY DISTRIBUTIONS

In this section, you will learn how to construct frequency distributions. Most often, this can be done by your computer, but it is important to go through the process to understand how frequency distributions are actually put together.

For nominal and ordinal variables, constructing a frequency distribution is quite simple. 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).

You can see that it is going to be difficult to make sense of these data just by eyeballing Table 2.5. 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 the frequency distributions for all four variables.

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 2.6 for the purpose of illustration). The tally results are then used to construct the frequency distribution presented in Table 2.6. 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.

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

✓ *Learning Check*

Table 2.5 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
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 2.6 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 Ordinal Variables

To construct a frequency distribution for ordinal-level variables, follow the same procedures outlined for nominal-level variables. Table 2.7 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.

Table 2.7 Frequency Distribution of the Variable Degree: GSS Subsample

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

The major difference between frequency distributions for nominal and ordinal variables is the order in which the categories are listed. The categories 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 2.7 are presented in inclining order from "less than high school" (the lowest educational category) to "graduate" (the highest educational category).

✓ *Learning Check*

Figures 2.1, 2.2, and 2.3 illustrate the gender and degree data in stages as presented in Tables 2.5, 2.6, and 2.7. To convince yourself that classifying the respondents by gender (Figure 2.2) and by degree (Figure 2.3) makes the job of counting much easier, turn to Figure 2.1 and answer these questions: How many men are in the group? How many women? How many said that they completed a bachelor's degree? Now turn to Figure 2.2: How many men are in the group? How many women? Finally, examine Figure 2.3: How many said that they completed a bachelor's degree?

Figure 2.1 Forty Respondents From the GSS Subsample, Their Gender and Their Level of Degree (see Table 2.5)

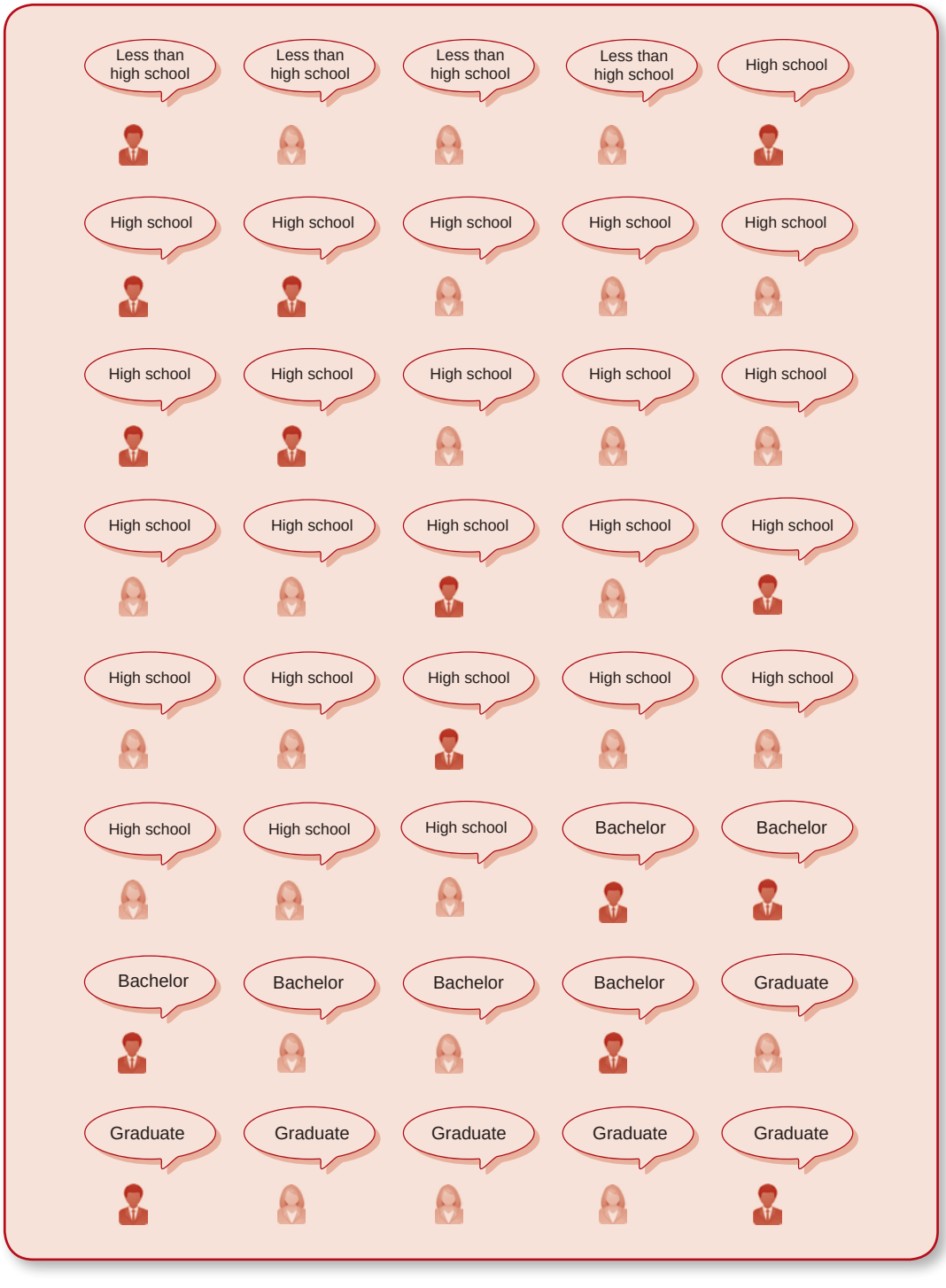


Figure 2.2 Forty Respondents From the GSS Subsample, Classified by Gender (see Table 2.6)

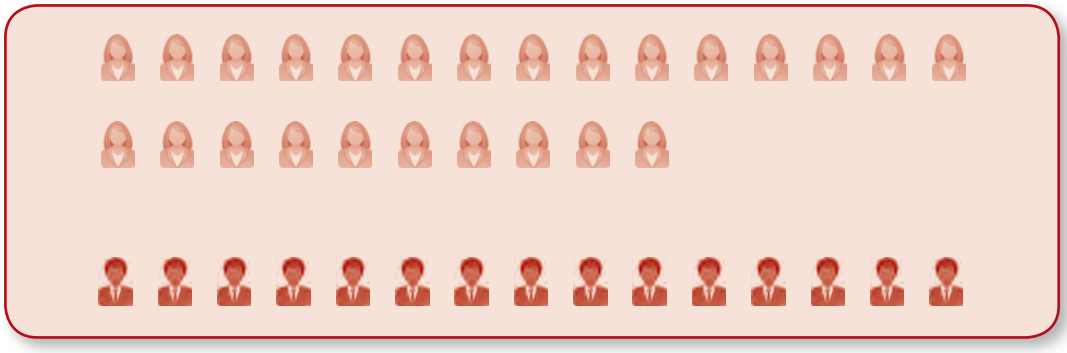
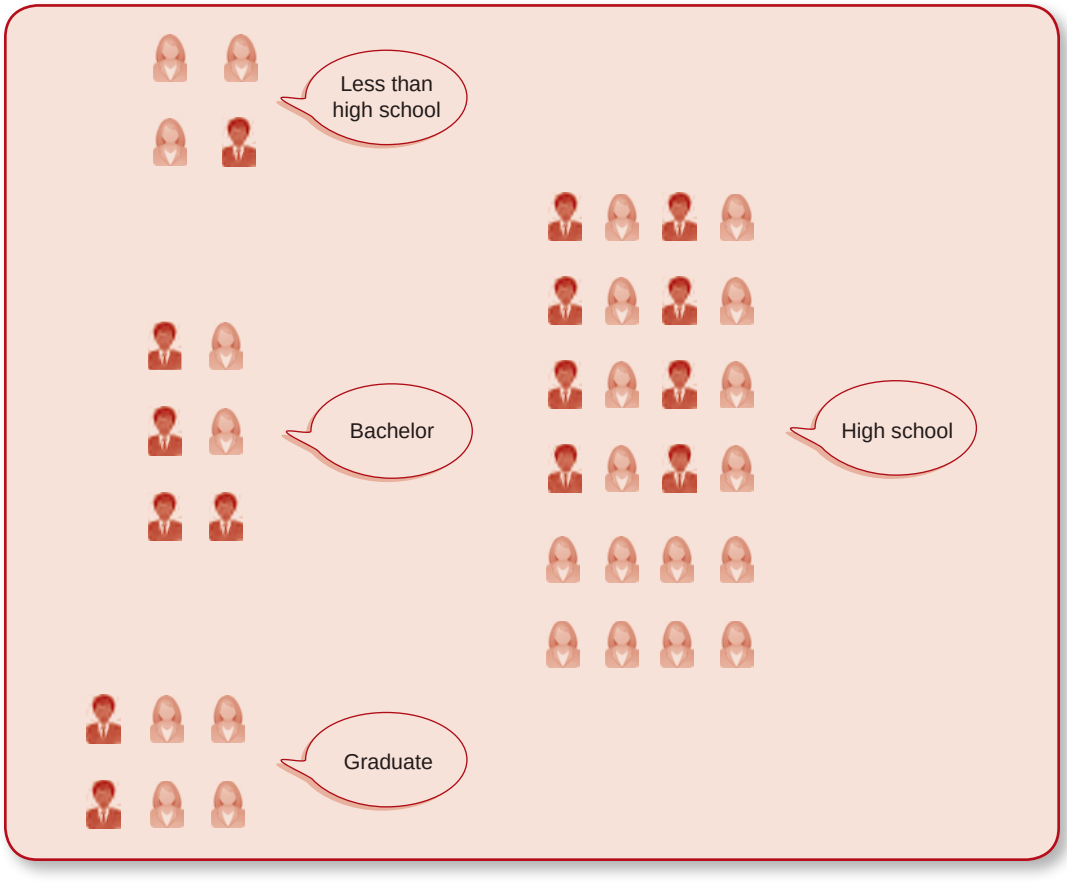


Figure 2.3 Forty Respondents From the GSS Subsample, Classified by Gender and Degree



Frequency Distributions for Interval-Ratio Variables

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 2.8.

Table 2.8 Frequency and Percentage Distribution for the 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

Very often interval-ratio variables have a wide range of values, which makes simple frequency distributions very difficult to read. For example, take a look at the frequency distribution for the variable *age* in Table 2.9. 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 2.10 displays such a grouped frequency distribution of the data in Table 2.9. 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 2.10, Frequency, tells us the number of respondents 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 2.10 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.[8] 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*

Can you verify that Table 2.10 was constructed correctly? Use Table 2.9 to determine the frequency of cases that fall into the categories of Table 2.10.

Table 2.9 Frequency Distribution of the Variable Age: GSS Subsample

Age of Respondent	*Frequency* (f)	*Age of Respondent*	*Frequency* (f)
21	1	59	1
22	1	62	1
23	1	67	1
24	2	69	1
28	1	71	4
29	1	75	1
32	1	76	1
34	2	82	1
35	1	84	1
37	2		
39	1		
40	1		
43	1		
44	1		
45	3		
46	1		
47	2		
48	1		
49	2		
50	1		
51	1		

Table 2.10 Grouped Frequency, Cumulative Frequency, and Percentage Distribution for the Variable Age: GSS Subsample

Age Category	*Frequency* (f)	Cf	*Percentage*
20–29	7	7	17.5
30–39	7	14	17.5
40–49	12	26	30.0
50–59	3	29	7.5
60–69	3	32	7.5
70–79	6	38	15.0
80–89	2	40	5.0
Total (*N*)	40		100.0

✓ *Learning Check*

If you are having trouble distinguishing between nominal, ordinal, and interval-ratio variables, go back to Chapter 1 and review the section on levels of measurement. The distinction between these three levels of measurement is 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 the cumulative frequencies in Table 2.10. The cumulative frequency column, denoted by *Cf*, shows the number of persons at or below each interval.

To construct a cumulative frequency distribution, start with the frequency in the lowest class interval (or with the lowest score, if the data are 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 2.10, 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 2.11, we have added the cumulative percentage distribution to the frequency and percentage distributions shown in Table 2.10. The cumulative percentage distribution shows, for example, that 35% of the sample was younger than 40 years of age—that is, 39 years or younger.

Like the percentage distributions described earlier, cumulative percentage distributions are especially useful when you want to compare differences between groups. For an example of how cumulative percentages are used in a comparison, we have used the 2008 GSS data to contrast the opinions of white women and white men about their standard of living compared with other American families. Respondents were asked the following question: "The way things are in America, people like me and my family have a good chance of improving our standard of living—do you agree or disagree?"

The percentage distribution and the cumulative percentage distribution for white women and white men are shown in Table 2.12. The cumulative percentage distributions suggest that slightly

Table 2.11 Grouped Frequency, Percentage, and Cumulative Percentage Distribution for the Variable Age: GSS Subsample

Age Category	*Frequency (f)*	*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	

Table 2.12 Will One's Standard of Living Improve: White Males Versus White Females

	White Males		*White Females*	
	%	C%	%	C%
Strongly agree	18.6	18.6	14.4	14.4
Agree	42.3	60.9	39.9	54.3
Neither	13.0	73.9	16.5	70.8
Disagree	24.2	98.1	24.8	95.6
Strongly disagree	2.0	100.1	4.4	100.0
Total	100.1		100.0	
(*N*)	355		411	

Source: General Social Survey, 2008.

more white men agree to the statement that their standard of living will improve. The two groups are separated by 6.6 percentage points—54.3% of the white women indicated that they either strongly agreed or agreed to the statement, while 60.9% of white males said the same. (Note that a higher percentage of women disagree with the statement than men.) What might explain these differences? Gender and wage discrimination might play a role in the gap between white women and white men.

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 a number obtained by dividing the number of actual occurrences in a given time period by the number of possible occurrences. For example, to determine the poverty rate for 2008, the U.S. Census Bureau took the number of men and women in poverty in 2008 (actual occurrences) and divided it by the total population in 2008 (possible occurrences). The rate for 2008 can be expressed as

$$\text{Poverty rate, } 2008 = \frac{\text{Number of people in poverty in 2008}}{\text{Total population in 2008}}$$

Since 40,136,000 people were poor in 2008 and the number for the total population was 304,060,000, the poverty rate for 2008 can be expressed as

$$\text{Poverty rate, } 2008 = \frac{40{,}136{,}000}{304{,}060{,}000} = 0.132$$

The poverty rate in 2008 as reported by the U.S. Census Bureau was 13.2% (0.132×100). Rates are often expressed as rates per thousand or hundred thousand to eliminate decimal points and make the number easier to interpret. For example, to express the poverty rate per thousand we multiply it by 1,000: $0.132 \times 1{,}000 = 132$. This means that for every 1,000 people, 132 were poor according to the U.S. Census Bureau definition.

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 under 18 years—highlighting how our young are vulnerable to poverty. The poverty rate for those under 18 years is as follows:

$$\begin{aligned}\text{Poverty rate (under 18 years, } 2008) &= \frac{13{,}454{,}000}{73{,}922{,}000} \\ &= 0.182 \times 1{,}000 \\ &= 182\end{aligned}$$

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

$$\begin{aligned}\text{Poverty rate (65 years or older, 2008)} &= \frac{3{,}842{,}000}{38{,}812{,}000} \\ &= 0.099 \times 1{,}000 \\ &= 99\end{aligned}$$

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

✓ *Learning Check*

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). Can you think of some other variables that could be expressed as rates?

READING THE RESEARCH LITERATURE: STATISTICAL TABLES[9]

In this section, we present some guidelines for how to read and interpret statistical tables displaying frequency distributions. The purpose is to help you see that some of the techniques described in this chapter are actually used in a meaningful way. Remember that it takes time and practice to develop the skill of reading tables. Even experienced researchers sometimes make mistakes when interpreting tables. So take the time to study the tables presented here, do the chapter exercises, and you will find that reading, interpreting, and understanding tables will become easier in time.

Basic Principles

The first step in reading any statistical table is to understand what the researcher is trying to tell you. Begin your inspection of the table by reading its title. It usually describes the central contents of the table. Check for any source notes to the table. These tell the source of the data or the table and any additional information that the author considers important. Next, examine the column and row headings and subheadings. These identify the variables, their categories, and the kind of statistics presented, such as raw frequencies or percentages. The main body of the table includes the appropriate statistics (frequencies, percentages, rates, etc.) for each variable or group as defined by each heading and subheading.

Table 2.13 was taken from an article written by Professors Eric Fong and Kumiko Shibuya about the residential patterns of different racial and ethnic groups. In their study, the researchers attempted to compare different home ownership patterns and suburbanization patterns among whites, blacks, five major Asian groups, and three major Hispanic groups. Data from the 1990 1% Public Use Microdata

Sample (PUMS) based on 15 Primary Metropolitan Statistical Areas (PMSAs) and Metropolitan Statistical Areas (MSAs) are used for their analysis.

In Table 2.13, the researchers display the percentages for housing status by location for each racial/ethnic group for 1990. Note that the frequency (*f*) for each category is not reported. Although the table is quite simple, it is important to examine it carefully, including its title and headings, to make sure that you understand what the information means.

Table 2.13 Percentage Distribution of Housing Tenure Status by Residential Location of Major Racial and Ethnic Groups, 1990

	Suburban Owner	*Central City Owner*	*Suburban Renter*	*Central City Renter*	*Total*
Whites	50.4	16.8	17.1	15.7	100.0
Blacks	14.6	27.2	13.9	44.3	100.0
Asians	34.2	20.6	14.7	30.5	100.0
Chinese	33.7	26.1	9.9	30.3	100.0
Japanese	39.2	20.8	17.8	22.2	100.0
Koreans	29.8	13.8	17.1	39.3	100.0
Vietnamese	31.0	11.8	24.3	32.9	100.0
Filipino	38.1	21.5	15.8	24.6	100.0
Hispanics	22.9	15.3	20.8	40.9	99.9
Cubans	36.3	16.6	16.2	30.9	100.0
Mexicans	26.0	15.5	26.8	31.8	100.1
Puerto Ricans	7.1	14.2	8.4	70.3	100.0
Total	41.7	18.5	16.8	23.0	100.0

Source: 1990 U.S. 1% PUMS Data. Adapted from Eric Fong and Kumiko Shibuya, "Suburbanization and Home Ownership: The Spatial Assimilation Process in U.S. Metropolitan Cities," *Sociological Perspectives* 43, no. 1 (2000): 143.

✓ *Learning Check*

Inspect Table 2.13 and answer the following questions:

- *What is the source of this table?*
- *How many variables are presented? What are their names?*
- *What is represented by the numbers presented in the first column? In the third column?*

What do the authors tell us about the table?

Table 2.13 [Table 1], which shows tenure status cross-classified by housing location, reveals four distinctive patterns. Column 1, which presents the proportions of suburban home owners, shows remarkable variations among groups. Although 42 percent of the total population of these 15 metropolitan areas have

been able to own a house in the suburbs, the figures suggest that it is atypical for some groups. Whites have the highest share: about 50 percent are suburban home owners. Blacks and Puerto Ricans, on the other hand, show an extremely low level of suburban home ownership. Asians in general have higher suburban home ownership rates (34%) than Hispanics (23%). There are, however, substantial variations within these two groups, ranging from 39 percent for Japanese to only 7 percent for Puerto Ricans.

Home ownership rates in the central city (Column 2) show a different picture. Blacks have the highest rate: about 27 percent of blacks own homes in central city areas. In fact, two-thirds of black home owners are found in those areas. Their higher home ownership rates in the central city may be explained by the fact that the cost of owning a house in the suburbs is higher. It is also possible that blacks are barred from the suburban areas because of discrimination in the housing market. Asian ethnic groups vary substantially in home ownership rates in the central city. The rate ranges from 12 percent for Vietnamese to 26 percent for Chinese. A higher rate of Chinese home owners in central cities may reflect the fact that Chinatowns have traditionally been located there. However, the rates among Hispanic ethnic groups are similar to one another. They range from 14 percent for Cubans and 17 percent for Mexicans.

The results in column 3 suggest a pattern for suburban renters that is distinct from home ownership patterns in either the suburbs or the central city. The results indicate that whites and Hispanics have the highest proportions of suburban renters. Although the percentage of Asian suburban renters is low on average, some specific Asian ethnic groups, such as Vietnamese, have higher rates than whites. The relatively high number of Vietnamese renters may be related to the results of the dispersion by refugee settlement programs. The higher rate of suburban renters among these Asian groups may be related to the absence of their ethnic communities in central cities. . . .

The results in column 4 reveal another unique pattern for central city renters. Among all groups, blacks and Puerto Ricans have the highest rates in this category. About 70 percent of Puerto Ricans and 44 percent of blacks are central city renters. Asians have a moderate percentage of central city renters, ranging from a low of 22 percent for Japanese to a high of 39 percent for Koreans. Whites have the lowest percentage of central city renters: about 16 percent.

Overall, these data show the complicated nature of the residential distribution patterns of racial/ethnic groups in contemporary cities. The results suggest four distinctive pictures of tenure status in suburbs and central cities, which would be undetected if suburbanization or home ownership were studied separately. These distinctive patterns, however, may simply reflect the differences in the socioeconomic resources and acculturation levels of each group in both locations.[10]

For a more detailed analysis of the relationships between these variables, you need to consider some of the more complex techniques of bivariate analysis and statistical inference. We consider these more advanced techniques beginning with Chapter 7.

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. Many people are intimidated by statistical information presented in frequency distributions or in other tabular forms, but they find the same information to be readable and understandable when presented graphically. Graphs tell a story in "pictures" rather than in words or numbers. They are supposed to make us think about the substance rather than the technical details of the presentation.

As we introduce the various graphical techniques, we also show you how to use graphs to tell a "story." The particular story we tell in this section is that of the elderly in the United States. The different types of graphs introduced in this section demonstrate the many facets of the aging of society over the next four decades. People have tended to talk about seniors as if they were a homogeneous group, but the different graphical techniques we illustrate here dramatize the wide variations in economic characteristics, living arrangements, and family status among people aged 65 and older. Most of the statistical information presented in this chapter is based on reports prepared by statisticians from the U.S. Census Bureau and other government agencies that gather information about the elderly in the United States and internationally.

Numerous graphing techniques are available to you, but here we focus on just a few of the most widely used ones in the social sciences. The first two, the pie chart and bar graph, are appropriate for nominal and ordinal variables. The next two, histograms and 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: RACE AND ETHNICITY OF THE ELDERLY

The elderly population of the United States is racially heterogeneous. As the data in Table 2.14 show, of the total elderly population (defined as persons 65 years and older) in 2008, about 33.7 million were white,[11] about 3.3 million black, 225,000 American Indian, 1.3 million Asian American, 36,000 Native Hawaiian or Pacific Islander, and 261,000 were two or more races combined.

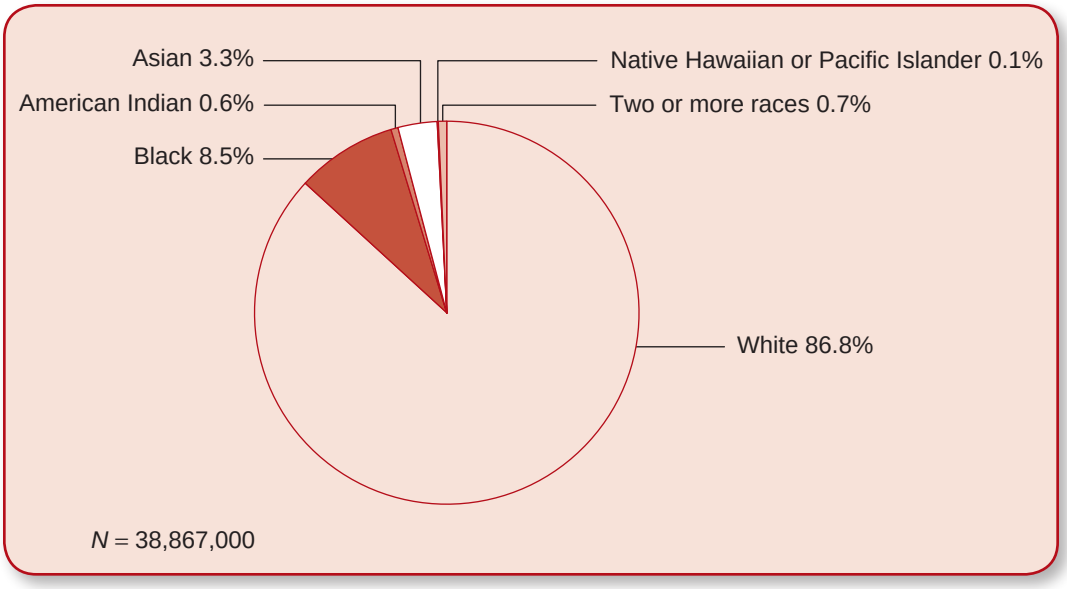
Table 2.14 Annual Estimates of the U.S. Population 65 Years and Over by Race, 2008

Race	*Frequency* (f)	*Percentage* (%)
White alone	33,737,000	86.8
Black alone	3,314,000	8.5
American Indian alone	225,000	0.6
Asian alone	1,294,000	3.3
Native Hawaiian or Pacific Islander	36,000	0.1
Two or more races combined	261,000	0.7
Total	38,867,000	100.0

Source: U.S. Census Bureau, *The 2010 Statistical Abstract*, Table 10.

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 2.4 displays the same information that Table 2.14 presents. 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 2.4. It shows that the elderly population is predominantly white (86.8%), followed by black (8.5%).

Figure 2.4 Annual Estimates of the U.S. Population 65 Years and Over by Race, 2008



Source: U.S. Census Bureau, *The 2010 Statistical Abstract*, Table 10.

Notice that the pie chart contains all the information presented in the frequency distribution. Like the frequency distribution, charts have an identifying number, a title that describes the content of the figure, and a reference to a source. The frequency or percentage is represented both visually and in numbers.

✓ *Learning Check*

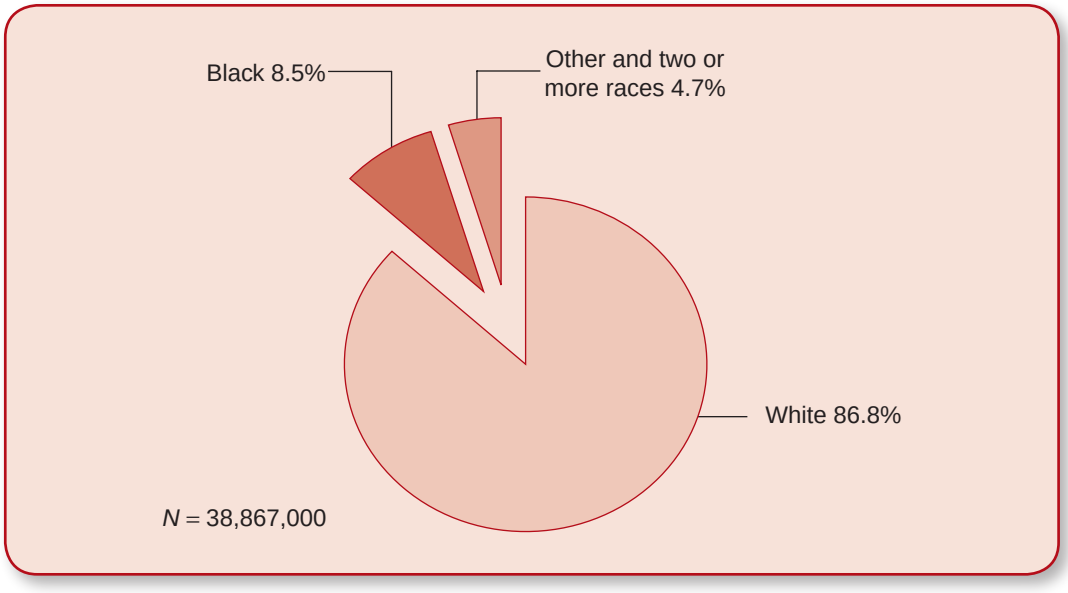
Pie chart A graph showing the differences in frequencies or percentages among the categories of a nominal or an ordinal variable. The categories are displayed as segments of a circle whose pieces add up to 100% of the total frequencies.

Note that the percentages for several of the racial groups are about 3% or less. It might be better to combine categories—American Indian, Asian, Native Hawaiian—into an "other races" category. This will leave us with three distinct categories: White, Black, and Other and Two or More Races. The revised pie chart is presented in Figure 2.5. We can highlight the diversity of the elderly population by "exploding" the pie chart, moving the segments representing these groups slightly outward to draw them to the viewer's attention.

THE BAR GRAPH: LIVING ARRANGEMENTS AND LABOR FORCE PARTICIPATION OF THE ELDERLY

The **bar graph** provides an alternative way to present nominal or ordinal data graphically. It shows the differences in frequencies or percentages among categories of a nominal or an ordinal variable. The

Figure 2.5 Annual Estimates of U.S. Population 65 Years and Over by Race, 2008



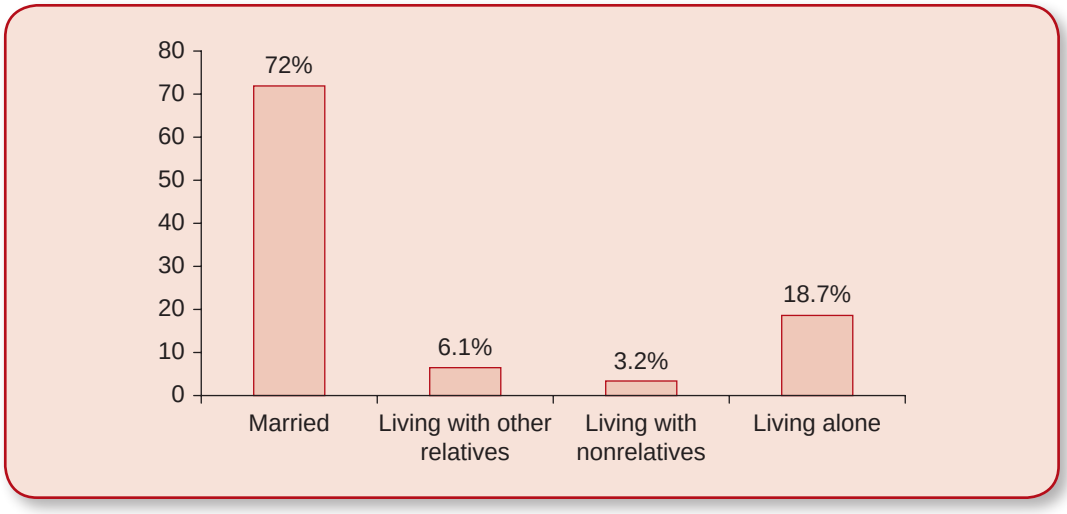
Source: U.S. Census Bureau, *The 2010 Statistical Abstract*, Table 10.

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 the 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 living arrangements of the elderly. Living arrangements change considerably with advancing age—an increasing number of the elderly live alone or with other relatives. Figure 2.6 is a bar graph displaying the percentage distribution of elderly males by living arrangements in 2009. 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). It shows that in 2009, 18.7% of elderly males lived alone, 72% were married and living with their spouses, 6.1% were living with other relatives, and the remaining 3.2% were living with nonrelatives.

Figure 2.6 Living Arrangement of Males (65 and older) in the United States, 2009



Source: U.S. Census Bureau, Current Population Survey, 2009, Annual Social and Economic Supplement, Tables A1 and A2.

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. For example, there is an increasing likelihood that women will live alone as they age. The longevity of women is the major factor in the gender differences in living arrangements.[12] In addition, elderly widowed men are more likely to remarry than elderly widowed women. Also, it has been noted that the current generation of elderly women has developed more protective social networks and interests.[13]

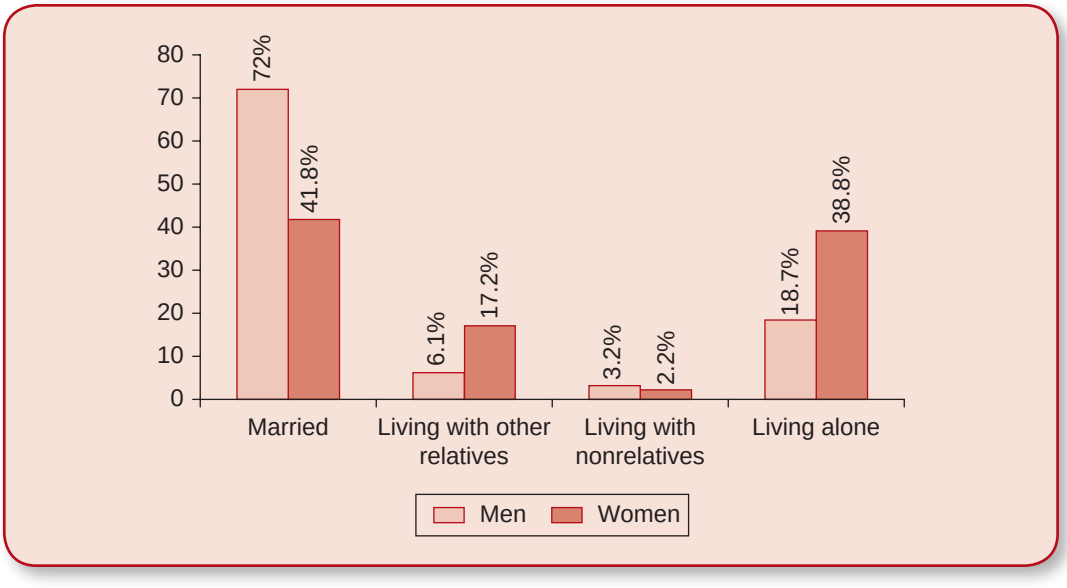
Suppose we want to show how the patterns in living arrangements differ between men and women. Figure 2.7 compares the percentage of women and men 65 years and older who lived with others or alone in 2009. It clearly shows that elderly women are more likely than elderly men to live alone.

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. This format is illustrated in Figure 2.8, which compares the percentage of men and women 55 years and over in the civilian labor force for 2008. We see that for all age categories, men were more likely to be employed than women.

THE HISTOGRAM

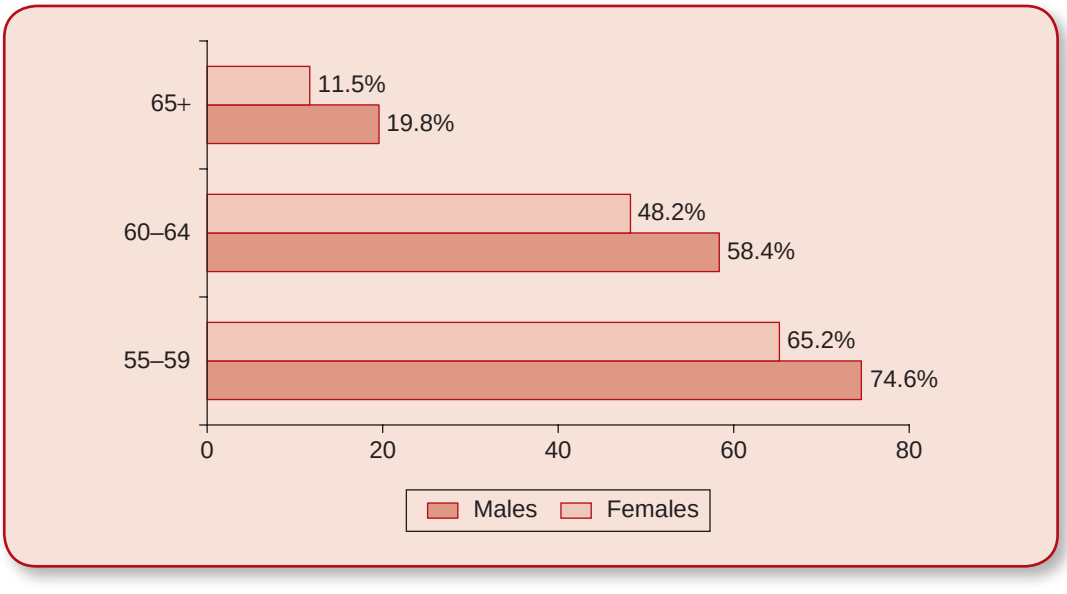
The **histogram** is used to show the differences in frequencies or percentages among categories of an interval-ratio variable. The categories are displayed as contiguous bars, with width proportional to the width of the category and height proportional to the frequency or percentage of that category. A histogram looks very similar to a bar chart except that the bars are contiguous to each other (touching) and may not be of equal width. In a bar chart, the spaces between the bars visually indicate that the categories are

Figure 2.7 Living Arrangement of U.S. Elderly (65 and older) by Gender, 2009



Source: U.S. Census Bureau, Current Population Survey, 2009, Annual Social and Economic Supplement, Tables A1 and A2.

Figure 2.8 Percentage of Men and Women 55 Years and Over Who Are Employed, 2008



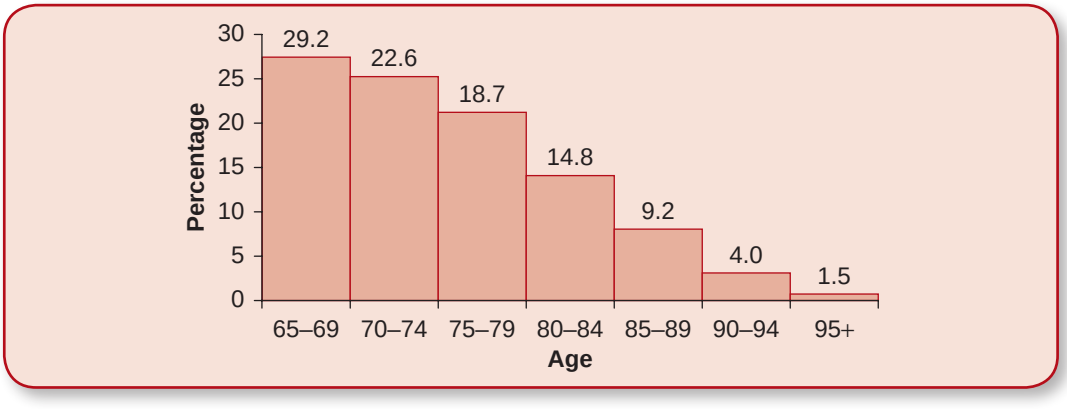
Source: U.S. Census Bureau, American Community Survey 2006–2008, Table B23001.

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.

Histogram A graph showing the differences in frequencies or percentages among the 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.

Figure 2.9 is a histogram displaying the percentage distribution of the population 65 years and over by age. The data on which the histogram is based are presented in Table 2.15. To construct the histogram of Figure 2.9, arrange the age intervals along the horizontal axis and the percentages (or frequencies) along the vertical axis. For each age category, construct a bar with the height corresponding to the percentage 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. The area that each bar occupies tells us the proportion of the population that falls into a given age interval. The histogram is drawn with the bars touching each other to indicate that the categories are contiguous.

Figure 2.9 Age Distribution of U.S. Population 65 Years and Over, 2008



Source: U.S. Census Bureau, *The 2010 Statistical Abstract*, Table 10.

When bar charts or histograms are used to display the frequencies of the categories of a single variable, the categories are shown on the X-axis and the frequencies on the Y-axis. In a horizontal bar chart or histogram, this is reversed.

✓ *Learning Check*

Table 2.15 Percentage Distribution of U.S. Population 65 Years and Over by Age, 2008

Age (years)	*Percentage (%)*
65–69	29.2
70–74	22.6
75–79	18.7
80–84	14.8
85–89	9.2
90–94	4.0
95+	1.5
Total (*N*)	100.0 (38,870,000)

THE LINE GRAPH

Numerical growth of the elderly population is taking place worldwide, occurring 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 experiencing dramatic growth of its elderly population. Figure 2.10 is a line graph displaying the elderly population of Japan by age.

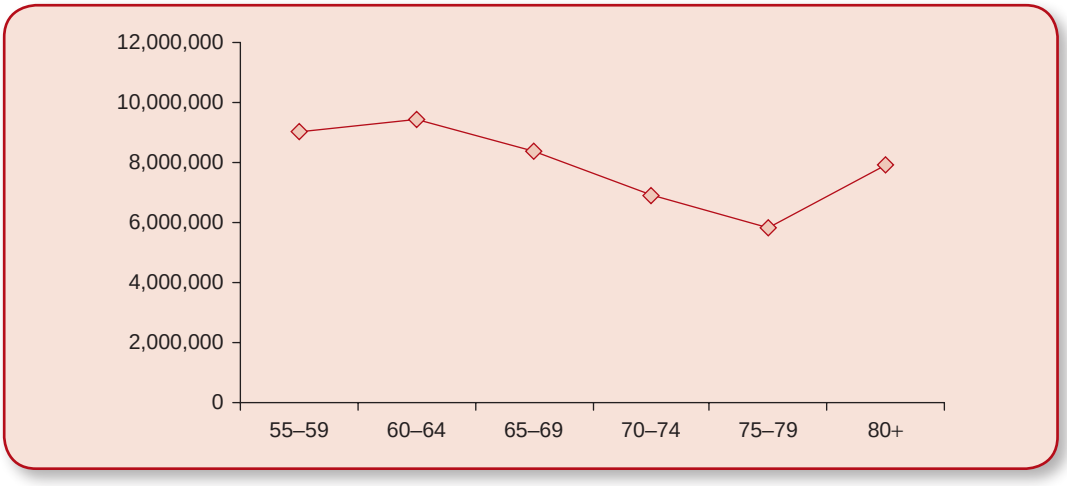
The **line graph** is another way to display interval-ratio distributions; it shows the differences in frequencies or percentages among categories of an interval-ratio variable. Points representing the frequencies of each category are placed above the midpoint of the category and are joined by a straight line. Notice that in Figure 2.10 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.

> ***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.

Both the histogram and the line graph can be used to depict distributions and trends of interval-ratio variables. How do you choose which one to use? To some extent, the choice is a matter of individual preference, but in general, line graphs are better suited for comparing how a variable is distributed across two or more groups or across two or more time periods. For example, Figure 2.11 compares the elderly population in Japan for 2000 with the projected elderly population for the years 2010 and 2020.

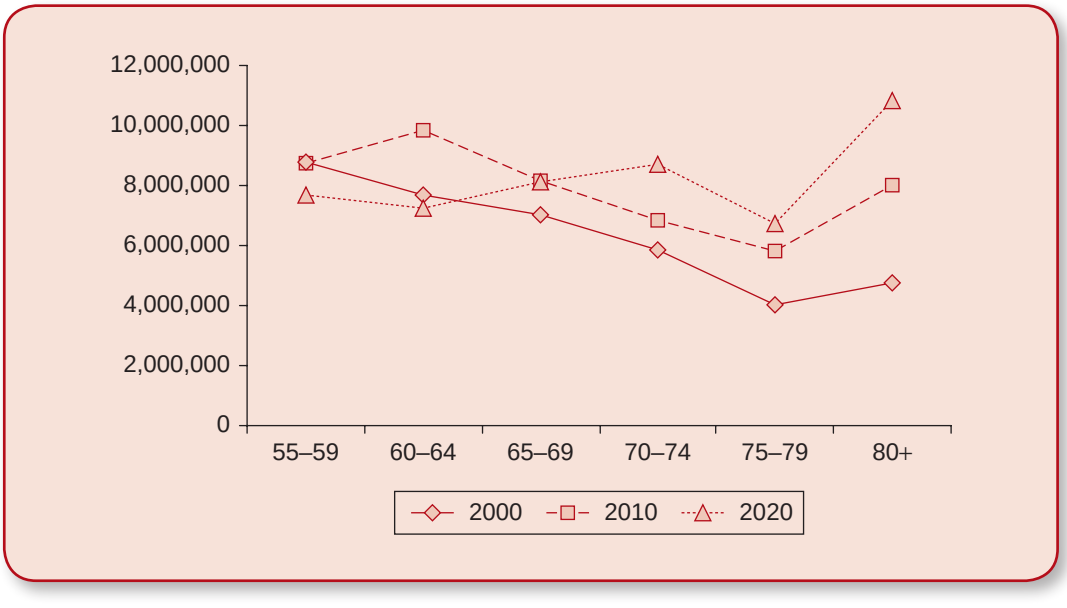
Let's examine this line graph. It shows that Japan's population of age 65 and over is expected to grow dramatically in the coming decades. According to projections, Japan's oldest-old population, those 80 years or older, is also projected to grow rapidly, from about 4.8 million (less than 4% of the total population) to 10.8 million (8.9%) by 2020. This projected rise has already led to a reduction in retirement benefits and other adjustments to prepare for the economic and social impact of a rapidly aging society.[14]

Figure 2.10 Population of Japan, Age 55 and Above, 2009



Source: Adapted from Ministry of Internal Affairs and Communications of Japan, Statistics Bureau, *Monthly Report April 2010*, Population Estimates.

Figure 2.11 Population of Japan, Age 55 and Over, 2000, 2010, and 2020



Source: Adapted from U.S. Census Bureau, Center for International Research, International Database, 2007.

Look closely at the line graph shown in Figure 2.11, comparing 2010 and 2020 data. How would you characterize the population increase among the Japanese elderly?

✓ *Learning Check*

TIME-SERIES CHARTS

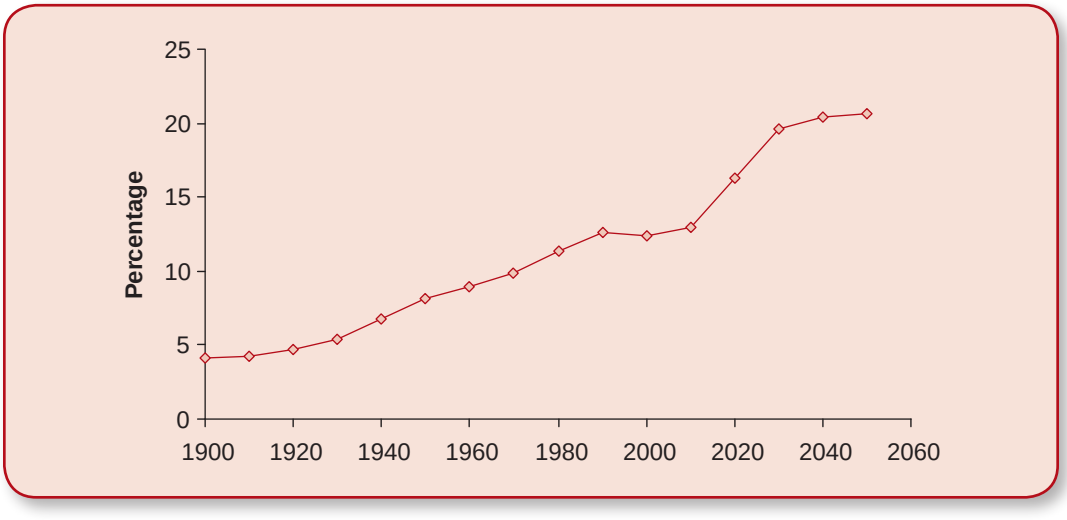
We are often interested in examining how some variables change over time. A **time-series chart** displays changes in a variable at different points in time. It involves two variables: (1) *time*, which is labeled across the horizontal axis, and (2) another variable of interest whose values (frequencies, 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 2.12 shows a time series from 1900 to 2050 of the percentage of the total U.S. population that is 65 years or older (the figures for the years 2000 through 2050 are projections made by the Social Security Administration, as reported by the U.S. Census Bureau). The number of elderly increased from a little less than 5% in 1900 to about 12.4% in 2000. The rate is expected to increase to 20% of the total population. This dramatic increase in the elderly population, especially beginning in the year 2010, is associated with the "graying" of the baby boom generation.

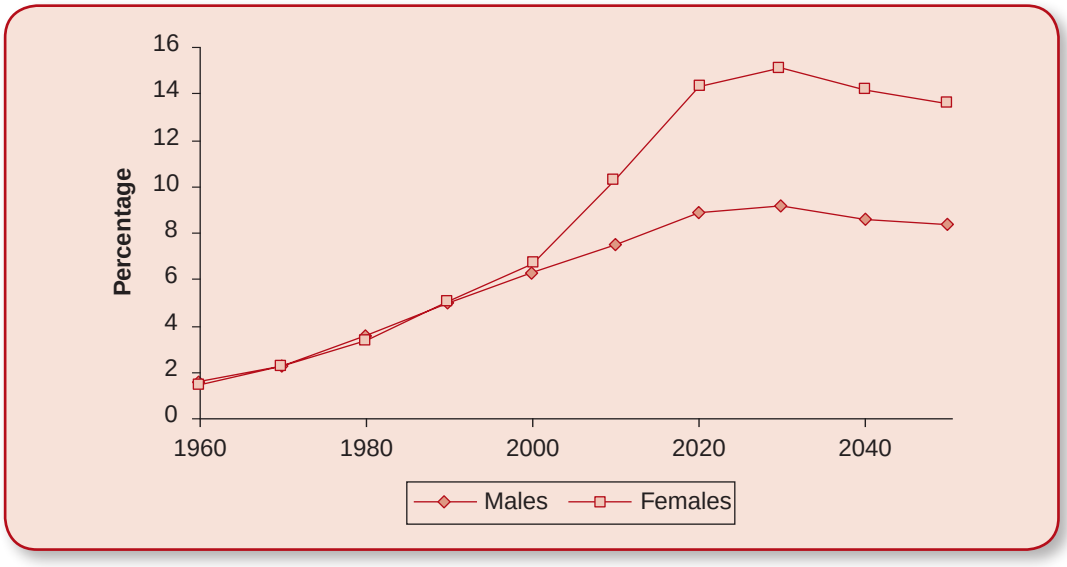
Often, we are interested in comparing changes over time for two or more groups. Let's examine Figure 2.13, which charts the trends in the percentage of divorced elderly from 1960 to 2050 for men

Figure 2.12 Percentage of Total U.S. Population 65 Years and Above, 1900–2050



Source: Federal Interagency Forum on Aging Related Statistics, *Older Americans 2004: Key Indicators of Well Being*, 2004.

Figure 2.13 Percentage Currently Divorced Among U.S. Population 65 Years and Over, by Gender, 1960–2050



Source: U.S. Bureau of the Census, *65+ in America, Current Population Reports, Special Studies*, P23–190, 1996, Table 6-1.

and women. This time-series graph shows that the percentage of divorced elderly men and elderly women was about the same until 2000. For both groups, the percentage increased from less than 2% in 1960 to about 5% in 1990.[15] According to projections, however, there will be significant increases in the percentage of men and especially women who are divorced: from 5% of all the elderly in 1990 to 8.4% of all elderly men and 13.6% of all elderly women by the year 2050. This sharp upturn and the gender divergence are clearly emphasized in Figure 2.13.

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.

✓ *Learning Check*

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 can not only quickly inform us, they 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.

Figure 2.14 Female Representation in National Parliaments, 2009: (a) Using 0 as the Baseline and (b) Using 30 as the Baseline

Source: Inter-Parliamentary Union. 2009. Women in National Parliaments. Retrieved December 22, 2009, from www.ipu.org/wmn-e/classif.htm.

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 Figures 2.14a and 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 Figures 2.14a and b, 2009 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 2.14b).

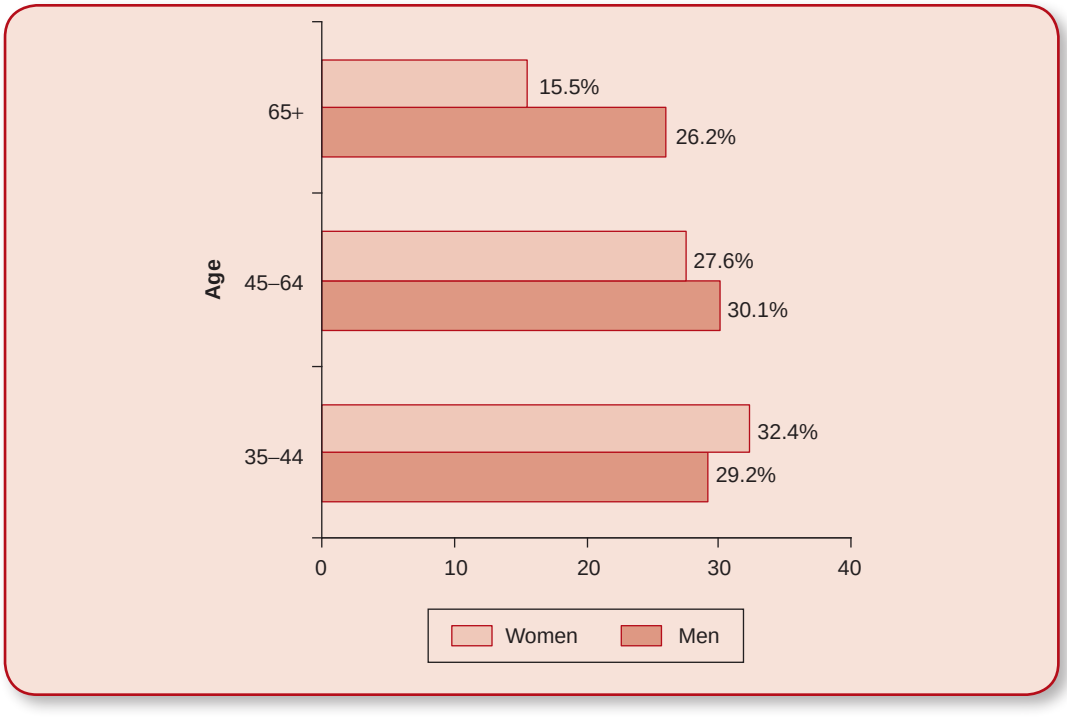
Remember: Always interpret the graph in the context of the numerical information the graph represents.

STATISTICS IN PRACTICE: DIVERSITY AT A GLANCE

We now illustrate some additional ways in which graphics can be used to highlight diversity visually. In particular, we show how graphs can help us (a) explore the differences and similarities among the many social groups coexisting within American society and (b) emphasize the rapidly changing composition of the U.S. population. Indeed, because of the heterogeneity of American society, the most basic question to ask when you look at data is "compared with what?" This question is not only at the heart of quantitative thinking[16] but underlies inclusive thinking as well.

Three types of graphs—the bar chart, the line graph, and the time-series chart—are particularly suitable for making comparisons among groups. Let's begin with the bar chart displayed in Figure 2.15. It compares the college degree attainment of those 35 years of age and over by gender.

Figure 2.15 Percentage of College Graduates Among People 35 Years and Over by Age and Gender, 2008



Source: U.S. Census Bureau, American Community Survey, 2008, Table B15001.

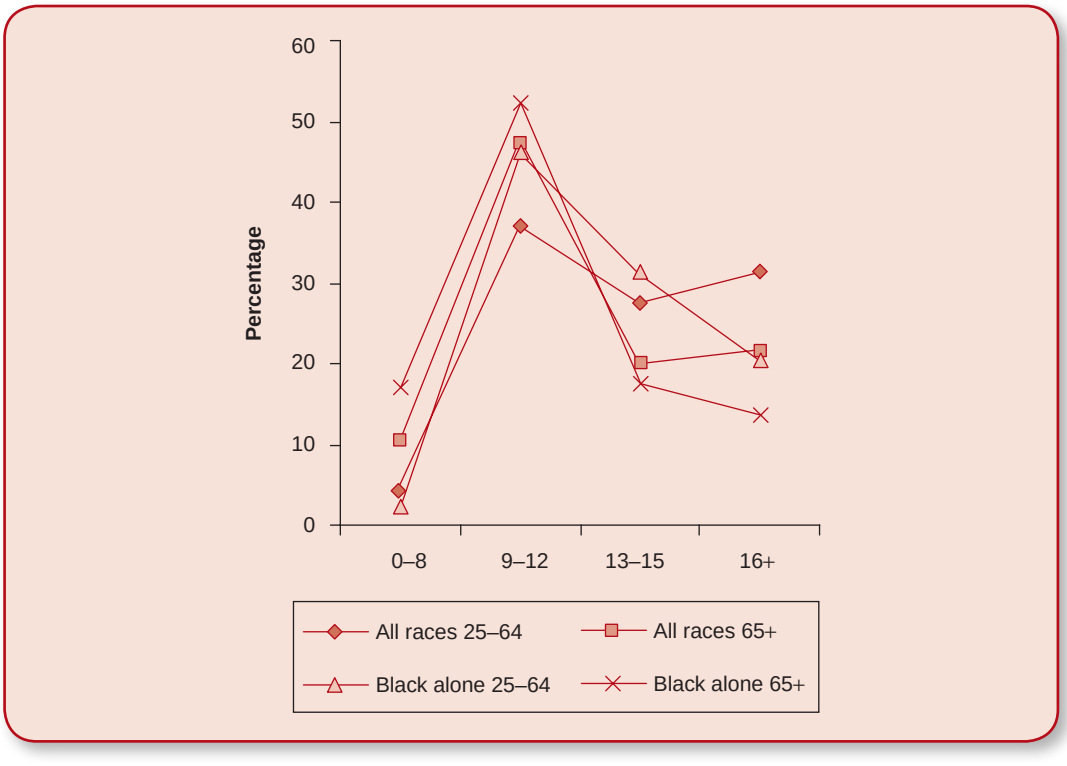
It shows that the percentage of men with a bachelor's degree or higher is greater than women in the older age groups (45 to 64 and 65+), whereas the percentage of women having at least a bachelor's degree exceeds that of men in the younger (35 to 44) age group. The smallest gap is among those who are 35 to 44 years, of whom 29.2% of men had a bachelor's degree or higher compared with 32.4% of women.

The line graph provides another way of looking at differences based on gender, race/ethnicity, or other attributes such as class, age, or sexual orientation. For example, Figure 2.16 compares years of school completed by black Americans aged 25 to 64 and 65 years and older with that of all Americans in the same age groups.

The data illustrate that in the United States the percentage of Americans who have completed only 8 years of education has declined dramatically from about 10.7% among Americans 65 years and older to 4.2% for those 25 to 64 years old. The decline for black Americans is even more dramatic, from 17% of the black elderly to about 2.2% for those 25 to 64 years old. The corresponding trend illustrated in Figure 2.16 is the increase in the percentage of Americans (all races as well as black Americans) who have completed 13 to 15 years of schooling, or 16 years or more. For example, about 17.3% of black Americans 65 years or older completed 13 to 15 years of schooling compared with 31.3% of those 25 to 64 years.

The trends shown in Figure 2.16 reflect the development of mass education in the United States during the past 50 years. The percentage of Americans who have completed 4 years of high school or

Figure 2.16 Educational Attainment in the United States by Race and Age, 2008



Source: The U.S. Census Bureau, *Educational Attainment in the United States: 2009*, Detailed Tables, Table 1.

more has risen from about 40% in 1940 to 86.6% in 2008 (record high according to the U.S. Census Bureau). Similarly, in 1940 only about 5% of Americans completed 4 years or more of college, compared with 29.4% in 2008.[17]

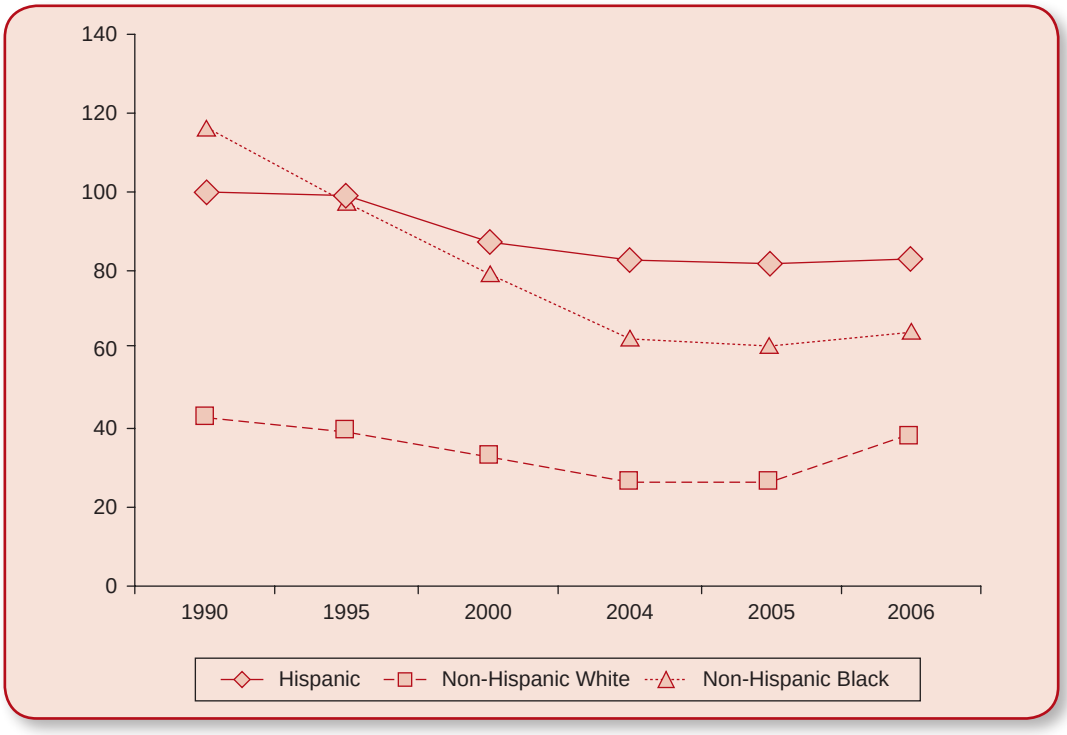
Figure 2.16 illustrates that overall, younger Americans (25–64 years old) are better educated than elderly Americans. However, despite these overall trends, there are differences between the number of years of schooling completed by "blacks" and "all races." Examine Figure 2.16, and find these differences. What do they tell you about schooling in America?

✓ *Learning Check*

Finally, Figure 2.17 is a time-series chart showing changes over time in the teenage birthrate (number of births per 1,000 women) among white, black, and Hispanic women. It shows that between 1990 and 2006, birthrates for teenagers declined; however, rates remain highest for Hispanic and non-Hispanic black women and lowest for non-Hispanic white women.

To conclude, the three examples of graphs in this section as well as other examples throughout this chapter have illustrated how graphical techniques can portray the complexities of the social world by emphasizing the distinct characteristics of age, gender, and ethnic groups. By depicting similarities and differences, graphs help us better grasp the richness and complexities of the social world.

Figure 2.17 Birthrates for U.S. Teenagers by Racial/Ethnic Group, 1990, 1995, 2000, 2004, 2005, 2006



Source: U.S. Census Bureau, *The 2010 Statistical Abstract*, Table 84.

Note: Rates per 1,000 women in specified group.

MAIN POINTS

- The most basic method for 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.

- 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.

- One other method of expressing raw frequencies in relative terms is known as a rate. Rates are defined as the number of actual occurrences in a given time period divided by the number of possible occurrences. Rates are often multiplied by some power of 10 to eliminate decimal points and make the number easier to interpret.

- A pie chart shows the differences in frequencies or percentages among categories of nominal or 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
cumulative frequency distribution
cumulative percentage distribution
frequency distribution
histogram
line graph
percentage
percentage distribution
pie chart
proportion
rate
time-series chart

ON YOUR OWN

Log on to the web-based student study site at **www.sagepub.com/ssdsessentials** for additional study questions, web quizzes, web resources, flashcards, codebooks and datasets, web exercises, appendices, and links to social science journal articles reflecting the statistics used in this chapter.

CHAPTER EXERCISES

1. Suppose you have 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 table below:
   a. What level of measurement is the variable race? Class?
   b. Construct raw frequency tables for race and for class.
   c. What proportion of the 30 individuals is nonwhite? What percentage is white?
   d. What proportion of the 30 individuals identified themselves as middle class?

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

Source: Data based on General Social Survey files for 1987 to 1991.

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

2. We selected a sample of people from the International Social Survey Program (ISSP) 2000. Raw data are presented for their sex (SEX), social class (CLASS), and number of household members (HOMPOP). CLASS is a subjective measure, with respondents indicating L = *lower*, W = *working*, M = *middle*, and U = *upper.*

Sex	*Hompop*	*Class*	*Sex*	*Hompop*	*Class*
F	1	L	F	2	U
F	3	W	M	2	M
F	1	M	F	4	W
M	2	M	M	2	U
F	3	M	M	4	M
M	2	U	M	4	W
F	7	L	M	2	M
M	3	M	F	1	W
M	4	M	M	2	U
F	1	M	M	4	W
F	2	M	F	7	M
M	5	L	M	3	M
M	4	M	M	4	M
M	4	L	F	4	M
M	3	W	F	5	L

   a. Construct a pie chart depicting the percentage distribution of sex. (*Hint:* Remember to include a title, percentages, and appropriate labels.)
   b. Construct a pie chart showing the percentage distribution of social class.
   c. Construct two pie charts comparing the percentage distribution of social class membership by sex.

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 2, construct bar graphs showing percentage distributions for sex and class. Remember to include appropriate titles, percentages, and labels.

5. The Gallup Organization conducted a survey in June 2007, asking Americans whether they approve or disapprove of the U.S. government's recent efforts to deal with illegal immigration. Results are provided in the table below, noting the percentage by ethnic background (not all responses are reported here so totals

will not add up to 100%). Do these data support the statement that people's views on illegal immigration are related to their ethnic background? Why or why not?

	Non-Hispanic Whites	*Blacks*	*Hispanics*
Approve	32	29	37
Disapprove	60	60	58

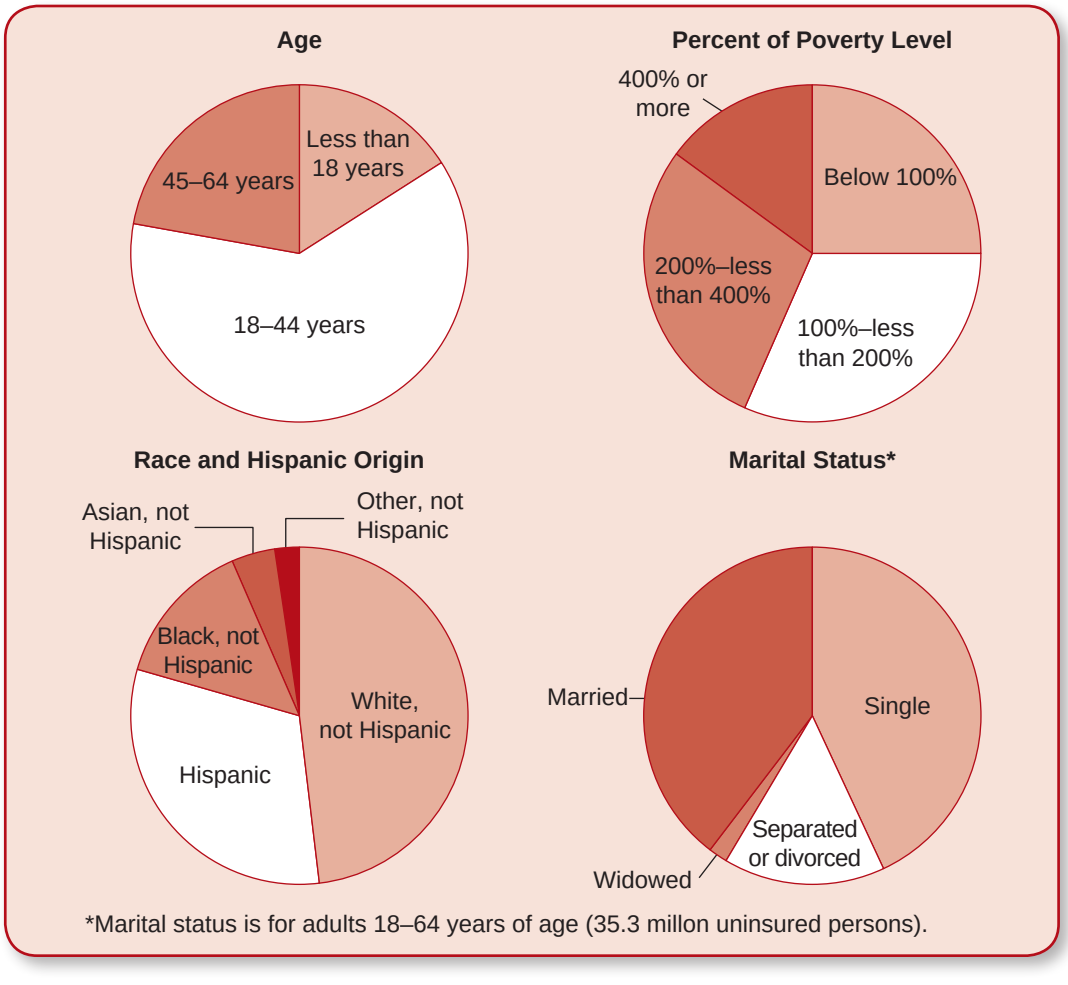
6. During the 2008 presidential election, health care and health insurance were identified by voters as important issues. Policy analysts have noted that the number of uninsured is increasing in the United States. Data from the National Center for Health Statistics are presented in Figure 2.18 and in the percentage table. What can be said about who did not have health insurance during 2005? How does the percentage of those without health insurance vary by ethnicity/race, age, marital, and poverty status?

Figure 2.18 The Uninsured Population Below 65 Years of Age by Selected Characteristics, 2005

Characteristic	Percentage
Age	
Below 18 years	16.2
18–44 Years	61.6
45–64 Years	22.2
Poverty	
Below 100%	25.1
100 to less than 200%	31.6
200 to less than 400%	28.7
400% or more	14.6
Race and Hispanic origin	
Hispanic	31.3
White only	48.0
Black	13.9
Asian	4.3
Other	2.4
Marital status	
Single	40.2
Married	41.8
Separated/divorced	16.3
Widowed	1.7

(Continued)

(Figure 2.18 Continued)



Exercises

7. The following tables present the frequency distributions for education by gender and race based on the GSS 2008. Use them to answer the following questions.

	Gender	
Education	***Male* (f)**	***Female* (f)**
Less than high school	100	116
High school graduate	327	407
Some college	68	76
College graduate	190	216

Source: General Social Survey, 2008.

	Race/Ethnicity		
Education	*White* (f)	*Black* (f)	*Hispanic* (f)
Less than high school	131	50	21
High school graduate	567	109	25
Some college	113	19	5
College graduate	335	28	5

Source: General Social Survey, 2008.

a. Construct tables based on percentages and cumulative percentages of educational attainment for gender and race.
b. What percentage of males has continued their education beyond high school? What is the comparable percentage for females?
c. What percentage of whites has completed high school or less? What is the comparable percentage for blacks? For Hispanics?
d. Are the cumulative percentages more similar for men and women or for the racial and ethnic groups? (In other words, where is there more inequality?) Explain.

8. You are writing a research paper on grandparents who had one or more of their grandchildren living with them. In 2000, 2.4 million grandparents were defined as caregivers by the U.S. Census, meaning that they had primary responsibility for raising their grandchildren below the age of 18. You discover the following information from the U.S. Census Report, "Grandparents Living With Grandchildren: 2000" (C2KBR-31, October 2003): Among grandparent caregivers, 12% cared for a grandchild for less than 6 months, 11% for 6 to 11 months, 23% for 1 to 2 years, 15% for 3 to 4 years, and 39% for 5 or more years.
   a. Construct a graph or chart that best displays this information on how long grandparents care for their grandchildren.
   b. Explain why the graph you selected is appropriate.
9. From the MTF 2008, we present data on the number of moving violation tickets in the past 12 months for males and females.

	Males (%)	*Females (%)*
None	71.3	80.4
1	17.1	13.0
2	7.4	5.0
3	2.1	0.7
4	2.1	0.9
Total	100.0	100.0

Source: Monitoring the Future, 2008.

a. What is the level of measurement for number of moving violation tickets?
b. What percentage of males and females had 3 tickets or more? Calculate the percentages for males and females separately.
c. Convert the percentages to frequencies. The total number of male respondents is 624 and the total number of female respondents is 699.
d. What can be said about the differences in the number of tickets between males and females?

10. The 2008 General Social Survey (GSS) data on educational level can be further broken down by race as follows:
    a. Construct two histograms for education, one for whites ($N = 636$) and one for blacks ($N = 136$).
    b. Now use the two graphs to describe the differences in educational attainment by race.

Years of Education	*Whites*	*Blacks*
0	2	0
1	0	0
2	1	1
3	2	0
4	0	0
5	3	1
6	5	0
7	2	0
8	13	1
9	8	6
10	17	8
11	27	6
12	180	36
13	56	18
14	101	33
15	27	5
16	103	13
17	32	1
18	32	2
19	12	1
20	13	4

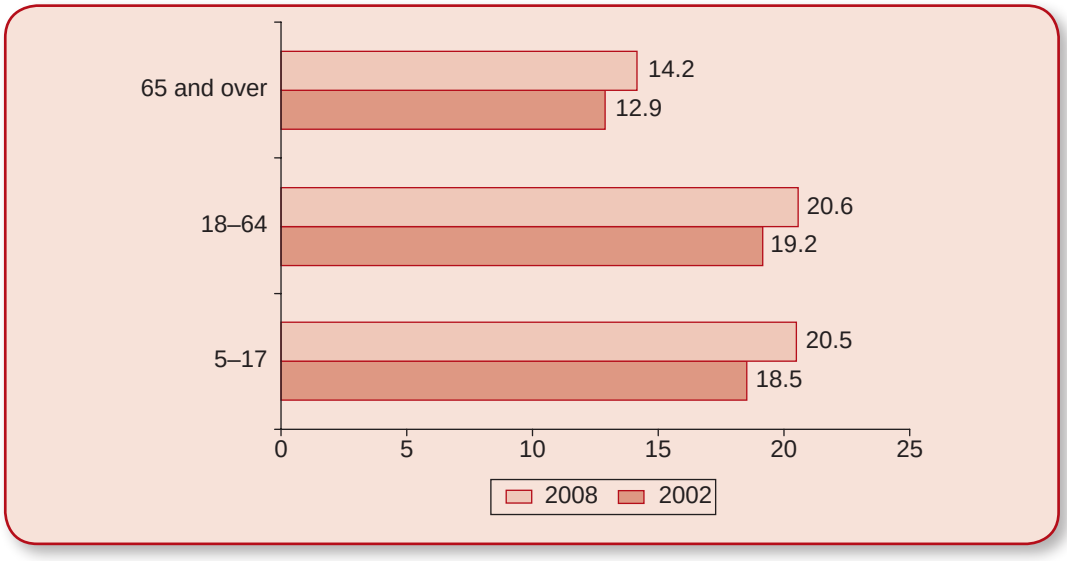
11. A Gallup Poll (June 9–30, 2004) compared 2,250 Americans' attitudes on minority rights and how much of a role the government should play in helping minorities improve their social and economic positions. Results from the survey are presented below (not all responses are reported). How would you characterize the differences between the attitudes of whites and minorities?

What Should Be the Government's Role in Improving Economic and Social Position of Minorities?	*Non-Hispanic Whites (%)*	*Blacks (%)*	*Hispanics (%)*
Major role	32	69	67
Minor role	51	22	21
No role	16	9	8

Source: Jeffrey Jones, "Blacks More Pessimistic Than Whites About Economic Opportunities," *Gallup Poll News Service*, The Gallup Organization. July 9, 2004. Used with permission.

12. Examine the bar chart representing the percentage of people speaking a language other than English at home as shown in Figure 2.19.
    a. Overall, which age group had the lowest percentage speaking a language other than English at home? Which age group had the highest?
    b. Describe the differences between the 2002 and 2008 percentages.

Figure 2.19 Percentage of People Speaking a Language Other Than English at Home Among the Population Aged 5 and Over, by Age, 2002 and 2008



Source: U.S. Census Bureau, American Community Survey, 2002, Table P035, and American Community Survey, 2008, Table C16007.

13. The Gallup Organization reported that public tolerance for gay rights reached a "high watermark" in 2007 (in contrast with attitudes measured over the past three decades). In its report, the organization examined the difference between different demographic groups to the question of whether homosexuality is an acceptable alternative lifestyle. Which groups support homosexuality as an acceptable alternative lifestyle? Review each demographic variable and summarize the pattern or level of support.

	Yes (%)	*No (%)*
Sex		
Men	53	44
Women	61	35
Age		
18–34 years	75	23
35–54 years	58	39
55+ years	45	51

(Continued)

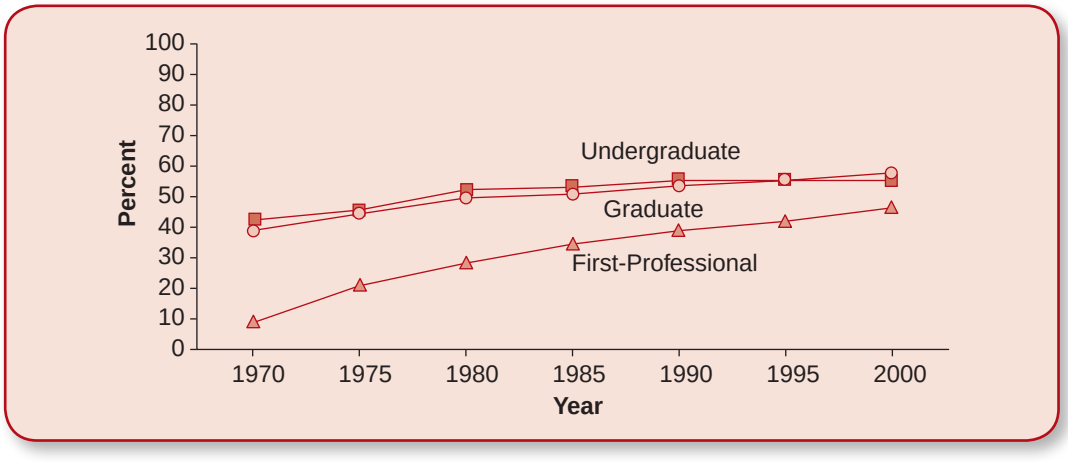
(Continued)

	Yes (%)	*No (%)*
Political affiliation		
Republican	36	58
Independent	60	36
Democrat	72	27
Religious service attendance		
Attend weekly	33	64
Attend nearly weekly/monthly	57	40
Attend less often/never	74	22

Source: Lydia Saad, "Tolerance for Gay Rights at High-Water Mark," 2007. Retrieved January 19, 2008, from www.gallup.com/poll/27694/Tolerance-Gay-Rights-HighWater-Mark.aspx. Used with permission.

14. As reported by Catherine Freeman (2004),[18] females have more success in postsecondary education than male students. They are more likely to enroll in college immediately after high school and have higher college graduate rates than males. In her report, Freeman provides the following time-series chart (Figure 2.20), documenting the percentage of women enrolled in undergraduate, graduate, and professional programs. Prepare a brief statement on the enrollment trends from 1970 to 2000.

Figure 2.20 Females as a Percentage of Total Enrollment in Undergraduate, Graduate, and First-Professional Education: Various Years, Fall 1970 to Fall 2000



Source: Catherine Freeman, *Trends in Educational Equity for Girls and Women: 2004* (NCES 2005–016). U.S. Department of Education, National Center for Education Statistics (Washington, DC: Government Printing Office, 2004).

Exercises