2 THE ORGANIZATION AND GRAPHIC PRESENTATION OF DATA

CHAPTER LEARNING OBJECTIVES

2.1 Calculate frequency, proportion, and percentage distributions.

2.2 Compare and contrast frequency, percentage, and cumulative distributions for nominal, ordinal, and interval-ratio variables.

2.3 Calculate and interpret rates.

2.4 Construct and interpret pie charts, bar graphs, histograms, statistical maps, line graphs, and time-series charts.

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 two such methods used by social scientists: distribution tables and graphic presentations.

FREQUENCY AND PERCENTAGE 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. The U.S. Census Bureau uses the term *foreign born* to refer to those who are not U.S. citizens at birth and those who have become U.S. citizens through the naturalization process. The term *native* or *native born* refers to anyone born in the United States, Puerto Rico, a U.S. Island area, or abroad of a U.S. parent or parents. The U.S. Census estimates that 13.5% of the U.S. population, or approximately 44 million people, are 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 Region of Birth for Foreign-Born Population, 2020

Region of Birth	Frequency (f)
Africa	2,362,292
Asia	13,794,873
Europe	4,784,383
Oceania	293,505
Latin America	22,073,519
Northern America	826,548
Total	44,135,120

Source: U.S. Census Bureau, *Table DP02 American Community Survey 5-Year Estimates, Selected Social Characteristics in the United States, 2020.*

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, 2020." The table consists of two columns. The first column identifies the variable (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 ($N = 44{,}135{,}120$). Note also that the source of the table is clearly identified. It tells us that the data are from the 2020 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.

What can you learn from the information presented in Table 2.1? The table shows that as of 2020, approximately 44 million people were classified as foreign born. Out of this group, most—about 22 million people—were from Latin America, 13.8 million were from Asia, followed by about 4.8 million from Europe.

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

We've calculated the proportion for the three largest groups of foreign born. First, the proportion of foreign born originally from Latin America is:

$$\frac{22{,}073{,}519}{44{,}135{,}120} = 0.50$$

The proportion of foreign born who were originally from Asia is:

$$\frac{13{,}794{,}873}{44{,}135{,}120} = 0.31$$

The proportion of foreign born who were originally from Europe is:

$$\frac{4{,}784{,}383}{44{,}135{,}120} = 0.11$$

The proportion of foreign born who were originally from all other reported areas (combining the category "All other" with those from Africa, Oceania, and Northern America) is:

$$\frac{3{,}482{,}345}{44{,}135{,}120} = 0.08$$

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

$$0.50 + 0.31 + 0.11 + 0.08 = 1.0$$

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

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

For example, the frequency of foreign born from Asia can be calculated as:

$$0.31\,(44{,}135{,}120) = 13{,}681{,}887$$

The obtained frequency differs somewhat from the actual frequency of 13,794,873. This difference is due to rounding off the proportion. If we use the actual proportion instead of the rounded proportion, we obtain the correct frequency:

$$0.3125599976\,(44{,}135{,}120) = 13{,}794{,}873$$

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:

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

TABLE 2.2 ■ Frequency and Percentage Distribution for Categories of Region of Birth for Foreign-Born Population, 2020

Region of Birth	Frequency (f)	Percentage (%)
Africa	2,362,292	5
Asia	13,794,873	31
Europe	4,784,383	11
Oceania	293,505	1
Latin America	22,073,519	50
Northern America	826,548	2
Total	44,135,120	100

Source: U.S. Census Bureau, *Table DP02 American Community Survey 5-Year Estimates, Selected Social Characteristics in the United States, 2020.*

Or:

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

Thus, the percentage of respondents who were originally from Latin America is:

$$0.50\,(100) = 50\%$$

LEARNING CHECK 2.1

What proportion of students in your statistics class identify as women? What proportion identify as men? What proportion identify as non-binary?

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 $N = 44{,}135{,}120$.

THE CONSTRUCTION OF FREQUENCY, PERCENTAGE, AND CUMULATIVE DISTRIBUTIONS

In this section, you will learn how to construct distribution tables for nominal, ordinal, and interval data. In addition to frequency and percentage distributions, cumulative distribution tables are also used by social scientists to summarize and analyze information. Most often, we use statistical software to create these different tables, but it is important to go through the process to understand how they are actually constructed. Make sure that you are able to distinguish between nominal, ordinal, and interval-ratio variables. The distinction between these levels of measurement will be important throughout the book.

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: sex, a nominal-level variable; degree, an ordinal measurement of education; and number of children and age, 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), but as a reminder, sex is occasionally treated as synonymous with gender despite the complex distinctions between sex and gender.

The interviewer recorded the sex 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, or 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

TABLE 2.3 ■ A GSS Subsample of 40 Respondents

Sex 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

(Continued)

(*Continued*)

TABLE 2.3 ■ A GSS Subsample of 40 Respondents

Sex of Respondent	Degree	Number of Children	Age
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.

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.

Frequency and Percentage Distributions for Nominal Variables

Let's begin with the nominal variable, *sex*. 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 and percentage distribution presented in Table 2.4. The table has a title describing its content ("Frequency and Percentage Distribution of the Variable *sex*: GSS Subsample"). Its categories (male and female) and their associated frequencies are clearly listed; in addition, the total number of cases (represented by 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.

TABLE 2.4 ■ Frequency and Percentage Distribution of the Variable *Sex*, GSS Subsample

Sex	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

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 men, women, and non-binary persons in your statistics class.

Frequency and Percentage 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 and percentage 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 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).

TABLE 2.5 ■ Frequency and Percentage Distribution of the Variable *Degree*, GSS Subsample

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

Frequency and Percentage Distributions for Interval-Ratio Variables

We hope that you agree by now that constructing frequency and percentage 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. Percentages are calculated by using formulas 2.3 or 2.4. Building a frequency and percentage distribution for interval-ratio variables with relatively few values is also easy. For example, when constructing a frequency and percentage distribution for *number of children*, simply list the number of children and report the corresponding frequency and percentage, 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 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. 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.

TABLE 2.6 ■ Frequency and Percentage 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	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 2.8 ■ Grouped Frequency and Percentage 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

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

Review the difference between nominal, ordinal, and interval-ratio variables first presented in Chapter 1. The distinction between these levels of measurement will be important throughout the book.

Cumulative Frequency and Percentage 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 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.

TABLE 2.9 Grouped Frequency and Cumulative Frequency Distribution 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	

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 2.9, 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).

A CLOSER LOOK 2.1

Real Limits, Stated Limits, and Midpoints of Class Intervals

The intervals presented in Table 2.8 constitute the categories of the variable *age* that we used to classify the survey's respondents. In Chapter 1, we noted that our variables need to be both exhaustive and mutually exclusive. These principles apply to the intervals here as well. This means that each of the 40 respondents can be classified into one and only one category. In addition, we should be able to classify all the possible scores.

In our example, these requirements are met: Each observation score fits into only one interval, and there is an appropriate category to classify each individual score as recorded in Table 2.8. However, if you looked closely at Table 2.8, you may have noticed that there is actually a gap of 1 year between adjacent intervals. A gap could create a problem with scores that have fractional values. Although age is conventionally rounded down, let's suppose for a moment that respondent's age had been reported with more precision. Where would you classify a woman who was 49.25 years old? Notice that her age would actually fall between the intervals 40–49 and 50–59! To avoid this potential problem, use the real limits shown in the following table rather than the stated limits listed in Table 2.8.

Real limits extend the upper and lower limits of the intervals by 0.5. For instance, the real limits for the interval 40–49 are 39.5–49.5, the real limits for the interval 50–59 are 49.5–59.5, and so on. (Scores that fall exactly at the upper real limit or the lower real limit of the interval [e.g., 59.5 or 49.5] are usually rounded to the closest even number. The number 59.5 would be rounded to 60 and would thus be included in the interval 59.5–69.5.) In the following table, we include both the stated limits and real limits for the grouped frequency distribution of respondent's age. So where would you classify a respondent who was 49.25 years old? (*Answer:* In the interval 39.5–49.5.) How about 19.9? (In the interval 19.5–29.5.)

The midpoint is a single number that represents the entire interval. A midpoint is calculated by adding the lower and upper real limits of the interval and dividing by 2. The midpoint of the interval 19.5–29.5, for instance, is $(19.5 + 29.5) \div 2 = 24.5$. The midpoints for all the intervals of the table are displayed in the third column.

Even though grouped frequency distributions are very helpful in summarizing information, remember that they are only a summary and therefore involve a considerable loss of detail. Since most researchers and students have access to computers, grouped frequencies are used only when the raw data are not available. Most of the statistical procedures described in later chapters are based on the raw scores.

Respondent's Age			
Stated Limits	**Real Limits**	**Midpoint**	**Frequency (*f*)**
20–29	19.5–29.5	24.5	7
30–39	29.5–39.5	34.5	7
40–49	39.5–49.5	44.5	12
50–59	49.5–59.5	54.5	3
60–69	59.5–69.5	64.5	3
70–79	69.5–79.5	74.5	6
80–89	79.5–89.5	84.5	2
Total (*N*)			40

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.

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. 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 used GSS 2021 data to contrast the opinions of white and Black Americans about government assistance to college students from low-income families. Respondents were asked to indicate their level of

TABLE 2.10 ■ Grouped Frequency and Cumulative Percentage Distribution 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	

TABLE 2.11 ■ Cumulative Percentage Distribution for "Government should give financial assistance to college students from low-income families, even if it might require a tax increase to pay for it" by Race, GSS 2021

Race	White		Black	
	Percentage (%)	Cumulative Percentage (*c*%)	Percentage (%)	Cumulative Percentage (*c*%)
Strongly Agree	14.7	14.7	25	25
Agree	39.9	54.6	42.9	67.9
Neither	24.2	78.8	23.2	91.1
Disagree	15.9	94.7	6	97.1
Strongly Disagree	5.4	100.1	3	100.1
Total	100.1		100.1	

support for the following statement, "The government should give financial assistance to college students from low-income families, even if it might require a tax increase to pay for it." The percentage distribution and the cumulative percentage distribution for white and Black respondents are shown in Table 2.11. (This table is referred to as a bivariate table, reporting the overlap between two variables—[1] *respondent race* and [2] level of agreement to the statement. We'll discuss bivariate tables in depth in Chapter 9.)

The cumulative percentage distributions suggest that a higher percentage of Black Americans are more likely than white Americans to agree with the statement about financial assistance to college students from low-income families. The two groups are separated by 13.3 percentage points—67.9% of Black respondents agree or strongly agree with the statement, while only 54.6% of white respondents said the same. These data prompt many other questions about the role that race or other variables may play in support for lower-income college students. What explains the difference between white and Black respondents? What would the differences be if we compared men with women? White respondents with Latinx respondents? College educated versus non-college educated individuals?

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.

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

For example, we can use data from the American Community Survey to determine the 2020 poverty rate by dividing the number of individuals below poverty (actual occurrences) by the total population in 2020 (possible occurrences). The 2020 rate can be expressed as:

$$\text{Poverty rate, 2020} = \frac{\text{Number of individuals below poverty}}{\text{Total population}}$$

Since 37,247,000 people were poor in 2020 and the number for the total population was 325,713,000, the poverty rate for 2020 was:

$$\text{Poverty rate, 2020} = \frac{37{,}247{,}000}{325{,}713{,}000} = 0.114$$

We can thus conclude that the poverty rate in 2020 was 11.4% (0.114 × 100). Rates are often multiplied by some power of 10 to eliminate decimal points and make the number easier to interpret. This means that for every 1,000 people, 114 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:

$$\text{Poverty rate for those under age 18 in 2020} = \frac{11{,}607{,}000}{72{,}295{,}000} = 0.161$$

We could even take a look at the poverty rate for older Americans and compare the difference in poverty rates between the two age groups:

$$\text{Poverty rate for those aged 65 and older in 2020} = \frac{5{,}000{,}000}{55{,}836{,}000} = 0.09$$

Many different variables can be expressed as rates. For example, 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 other variables that can be expressed as rates?

Reading the Research Literature: The Impact of COVID on College Students' Diet Quality and Physical Activity

Statistical tables that display frequency distributions or other kinds of statistical information are found in virtually every book, article, or newspaper report that makes any use of statistics. However, the inclusion of statistical tables in a report or an article doesn't necessarily mean that the research is more scientific or convincing. You will always have to ask what the tables are saying and judge whether the information is relevant or accurately presented and analyzed. Most statistical tables presented in the social science literature are a good deal more complex than those we describe in this chapter.

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, as the title usually describes the central contents of the table. Check for any source notes to the table; such notes reveal 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 2.12 was included in an article written by Melissa Oflert and her colleagues (2022) about how the COVID pandemic impacted college students' diet quality and physical activity. The researchers hypothesized that students would have poorer diet quality and less physical activity as a result of the pandemic. For their analysis, the researchers relied on data from a survey of 2,018 students attending an Appalachian university in spring 2020. Students were asked a series of questions regarding their diet and eating patterns and exercise activities before and since COVID. For this discussion, we'll focus on students' physical activities. Table 2.12 summarizes students' physical activity before and since COVID-19. (Note: The total number of responses may vary by question.)

TABLE 2.12 ■ Physical Activity Before and Since COVID-19

Variable	Before COVID-19 Frequency N (%)	Since COVID-19 Frequency N (%)
Gym Use		
Never	710 (36.0%)	1553 (94.5%)
Rarely	257 (13.0%)	39 (2.4%)
Sometimes	436 (22.2%)	16 (0.9%)
Often	568 (28.8%)	36 (2.2%)
Student Recreation Centre Use		
Never	740 (37.7%)	1615 (98.5%)
Rarely	359 (18.3%)	16 (0.9%)
Sometimes	408 (20.8%)	4 (0.2%)
Often	457 (23.2%)	4 (0.2%)
Outdoor Activities		
Never	511 (26.0%)	738 (44.8%)
Rarely	543 (27.6%)	303 (18.4%)
Sometimes	655 (33.3%)	373 (22.6%)
Often	259 (13.2%)	233 (14.1%)
Running Outside		
Never	1020 (52.5%)	921 (56.1%)
Rarely	447 (22.8%)	255 (15.5%)
Sometimes	348 (17.8%)	267 (16.3%)
Often	137 (7.0%)	198 (12.1%)
Running Inside		
Never	1037 (52.7%)	1369 (83.6%)
Rarely	345 (17.5%)	108 (6.6%)
Sometimes	369 (18.8%)	105 (6.4%)
Often	217 (11.0%)	57 (3.5%)
Self-Led At-Home Workouts		
Never	827 (42.1%)	626 (37.6%)
Rarely	506 (25.8%)	295 (17.9%)
Sometimes	424 (21.6%)	334 (20.2%)
Often	206 (10.5%)	393 (24.0%)

(*Continued*)

(*Continued*)

TABLE 2.12 ■ Physical Activity Before and Since COVID-19

Variable	Before COVID-19 Frequency N (%)	Since COVID-19 Frequency N (%)
Guided at Home Workouts		
Never	1107 (56.5%)	829 (50.2%)
Rarely	365 (18.6%)	228 (13.8%)
Sometimes	320 (16.3%)	264 (16.0%)
Often	166 (8.5%)	329 (19.9%)
Workout Apps		
Never	1328 (67.7%)	1138 (69.1%)
Rarely	300 (15.3%)	174 (10.6%)
Sometimes	222 (11.3%)	157 (9.5%)
Often	112 (5.7%)	179 (10.9%)
Physical Activity Levels		
Low	561 (27.8%)	1048 (51.9%)
Moderate	257 (12.7%)	284 (14.1%)
High	1200 (59.5%)	686 (34.0%)

Source: Adapted from Melissa Olfert, Rachel Wattick, Emily Saurborn, and Rebecca Hegedorn, "Impact of COVID-19 on College Students Diet Quality and Physical Activity," *Nutrition and Health* 18, (2022): 8–9. Some cell percentage totals will not add up to 100% due to rounding.

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.

LEARNING CHECK 2.5

Inspect Table 2.12 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 last row of the table?

What do the authors tell us about the table?

> The physical activity habits of respondents significantly changed due to COVID-19. As could be expected, there was a significant decrease in the use of gyms and the student recreation center. There was a significant increase in the use of at-home workouts, both self-created and guided, as well as the use of workout apps. However, the overall amounts of physical activity decreased, with a significant number of individuals engaged in low levels of physical activity increasing, and a significant number of individuals with high levels of vigorous activity decreasing.[4]

The authors concluded that:

> Students' lifestyle habits have significantly changed since campus activities were disrupted and there is the potential that students heavily rely on campus resources, such as the student recreation centre, in order to achieve adequate physical activity levels. Further, the

> disruption of routine leading to increased boredom has negatively impacted student eating habits . . . Although home confinement and campus closures are important for public health, they undoubtedly lead to negative behaviors in a health compromising way. This is important for college students who are at a time period where their health is especially important to maintain success and long-term well-being.[5]

Further analyses could examine why these differences exist. Other variables that explain the differences between these groups could be identified (such as gender, year in school, and place of residence). For a more detailed analysis of the relationships between these variables, you need to consider some of the more complex techniques of bivariate (two-variable) analysis and statistical inference. We consider these advanced techniques beginning with Chapter 8.

GRAPHIC PRESENTATION OF DATA

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.

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.

As we introduce various graphical techniques, we also show you how to use graphs to tell a story. The particular story we tell here is that of the older population 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-age and older Americans, what demographers have referred to as the graying of America. The U.S. Census predicts that in 2034 older adults will for the first time outnumber children under the age of 18. In that year, people age 65 or older are expected to number 77 million, while the number of children under age 18 will be 76.5 million.[6]

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 statistical maps and time-series charts. The statistical map is most often used with interval-ratio data. Finally, time-series charts are used to show how some variables change over time.

The Pie Chart

The older population of the United States is racially heterogeneous. As the data in Table 2.13 show, of the total 50,783,796 older people (defined as persons 65 years and older) in 2013–2017, the two largest racial groups were whites = 82.9% and Blacks = 9.2%.

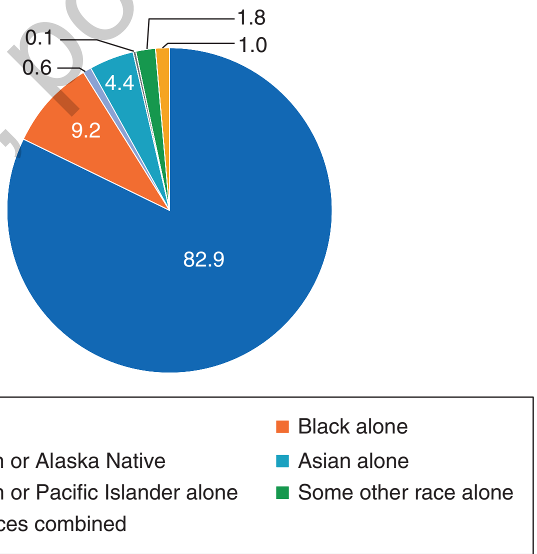
A **pie chart** shows the differences in frequencies or percentages among the categories of a nominal or an ordinal variable. The categories are displayed as segments of a circle whose pieces add up to 100% of the total frequencies. The pie chart shown in Figure 2.1 displays the same information that Table 2.13 presents. Although you can inspect these data in Table 2.13, you can interpret the information more easily by seeing it presented in the pie chart in Figure 2.1.

TABLE 2.13 ■ Five-Year Estimates of the U.S. Population 65 Years and Over by Race, 2015–2019, Percentages

Race	Percentage (%)
White alone	82.9
Black alone	9.2
American Indian or Alaska Native	0.6
Asian alone	4.4
Native Hawaiian or Pacific Islander alone	0.1
Some other race alone	1.8
Two or more races combined	1.0
Total	100

Source: U.S. Census Bureau, *American Community Survey*, Table S0103, 2019.

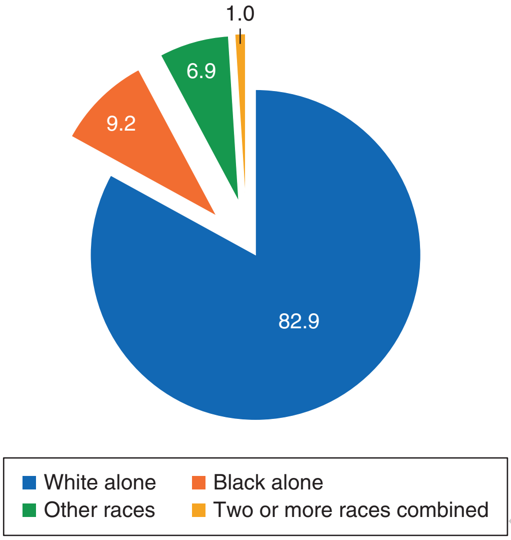
FIGURE 2.1 ■ Five-Year Estimates of the U.S. Population 65 Years and Over by Race, 2015–2019, Percentages



Source: U.S. Census Bureau, *American Community Survey*, Table S0103, 2019.

Did you notice that the percentages for several of the racial groups are 4.4% 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.

FIGURE 2.2 ■ Five-Year Estimates of U.S. Population 65 Years and Over by Race, 2015–2019, Percentages



Source: U.S. Census Bureau, *American Community Survey*, Table S0103, 2019.

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 older persons comprised 82.9% of the U.S. older population from 2015 to 2019.

The Bar Graph

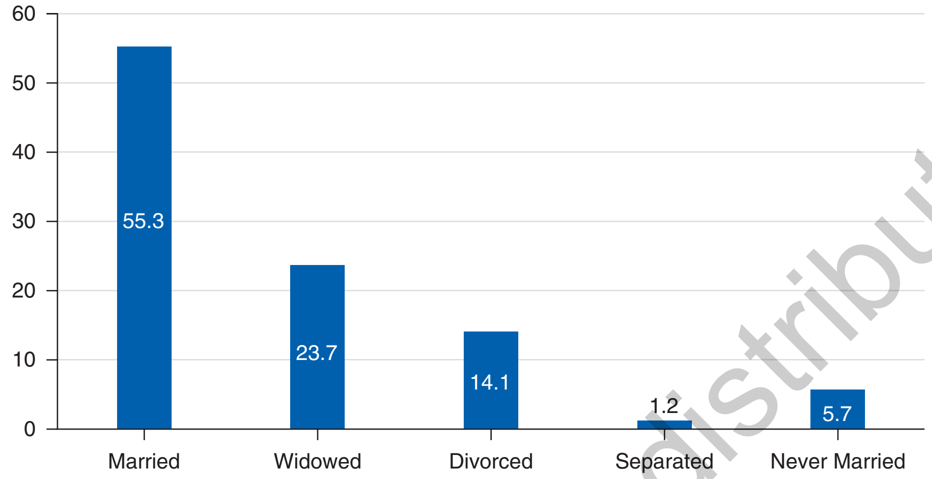
The **bar graph** provides an alternative way to graphically present nominal or ordinal data. It shows the differences in frequencies or percentages among categories of a nominal or an ordinal variable. The categories are displayed as rectangles of equal width with their height proportional to the frequency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

Let's illustrate the bar graph with an overview of the marital status of the older population. Figure 2.3 is a bar graph displaying the percentage distribution of persons 65 years old and over by marital status from 2015 to 2019. 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 from 2015 to 2019, the majority of the older population were married. Specifically, 55.3% were married, 23.7% were widowed, 14.1% 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.

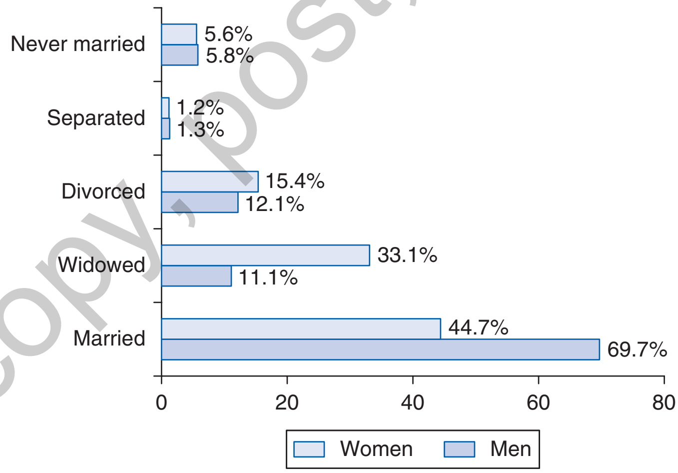
Bar graphs are often used to compare one or more categories of a variable among different groups. Suppose we want to show how the patterns in marital status differ between men and

FIGURE 2.3 ■ Marital Status of U.S. Elderly (65 Years and Older), Percentages, 2015–2019



Source: U.S. Census Bureau, *American Community Survey*, Table S0103, 2019.

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

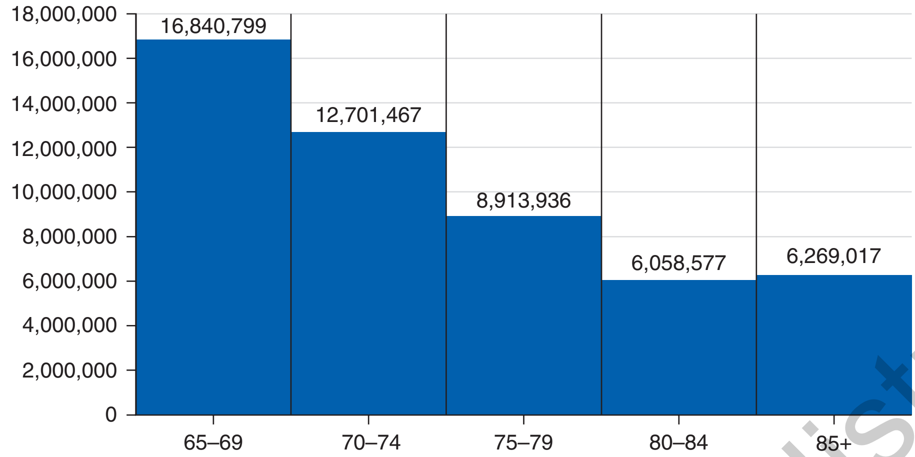


Source: U.S. Census Bureau, *American Fact Finder*, Table S1201, 2017.

women. The longevity of women is a major factor in the gender differences in marital and living arrangements. Additionally, older widowed men are more likely to remarry than older 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 women over 65 are more likely than men over 65 to be widowed (33.1% vs. 11.1%), and older men are more likely to be married than older women (69.7% vs. 44.7%).

FIGURE 2.5 ■ Age Distribution of U.S. Elderly (65 Years and Older), 2015–2019



Source: U.S. Census Bureau, *American Community Survey*, Table S0101, 2019.

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 Histogram

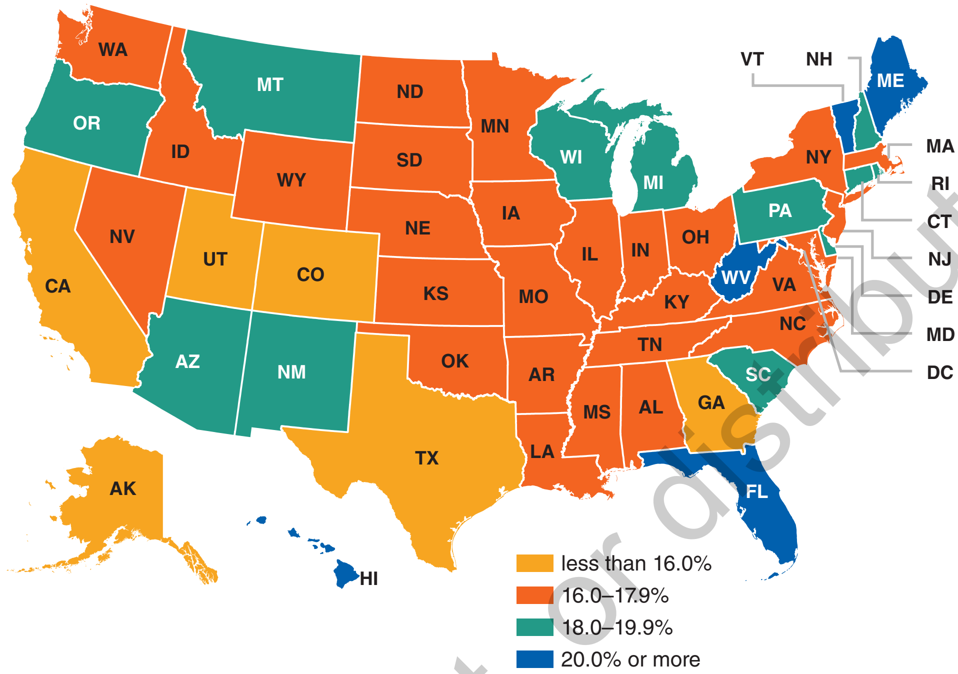
The **histogram** is used to show the differences in frequencies or percentages among categories of an interval-ratio or ordinal variable. The categories are displayed as contiguous bars, with width proportional to the width of the category and height proportional to the frequency or percentage of that 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 (e.g. married, single), gender (men, women, non-binary people), 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. The total number of elderly 65 years and over can be found by summing all the reported frequencies.

The Statistical Map

Since the 1960s, the elderly have been relocating to the South and the West of the United States. The concentration of elderly in these areas has increased by as much as 80% (although recent census data reveal that the Great Recession halted this dominant migration trend). We can display these dramatic geographical changes in American society by using a **statistical map**.

FIGURE 2.6 ■ Percentage of the Population 65 Years and Over by State, 2020



Source: U.S. Census Bureau, *Vintage 2020 Population Estimates, 2020.*

A statistical map presents geographic data patterns or variations, such as population distribution, voting patterns, crime rates, or labor force composition.

Let's look at Figure 2.6. It presents a statistical map, by state, of the percentage of the population 65 years and over for 2020. The variable percentage of the population has four categories: (1) Less than 16%, (2) 16%–17.9%, (3) 18%–19.9%, and (4) 20% or more. Each category is represented by a different shading (or color code), and the states are shaded depending on their classification into the different categories. To make it easier to read a map that you construct and to identify its patterns, keep the number of categories relatively small—say, not more than five.

Maps may also display geographical patterns on the level of cities, counties, city blocks, census tracts, and other units. Your choice of whether to display variations on the state level or for smaller units will depend on the research question you wish to explore.

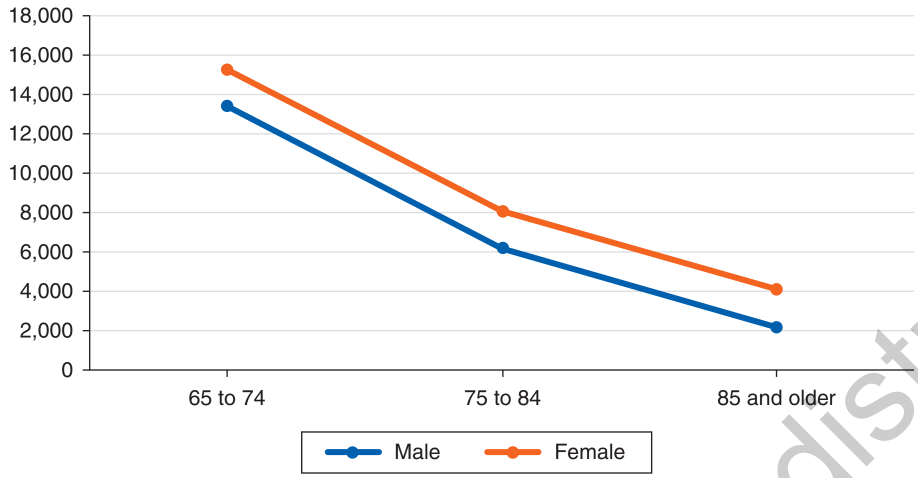
LEARNING CHECK 2.6

Can you think of a few other examples of data that could be described using a statistical map? What type of data are organized or reported at the state level?

The Line Graph

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.

FIGURE 2.7 ■ Population 65 and Older by Age and Sex (numbers in thousands), 2016



Source: Andrew Roberts, Stella Ogunwole, Laura Blakeslee, and Megan Rabe. *The Population 65 Years and Older in the United States: 2016.* American Community Survey Reports, ACS-38, October 2018.

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 2.7. Figure 2.7 is a line graph displaying the number of elderly males and females for 2016.

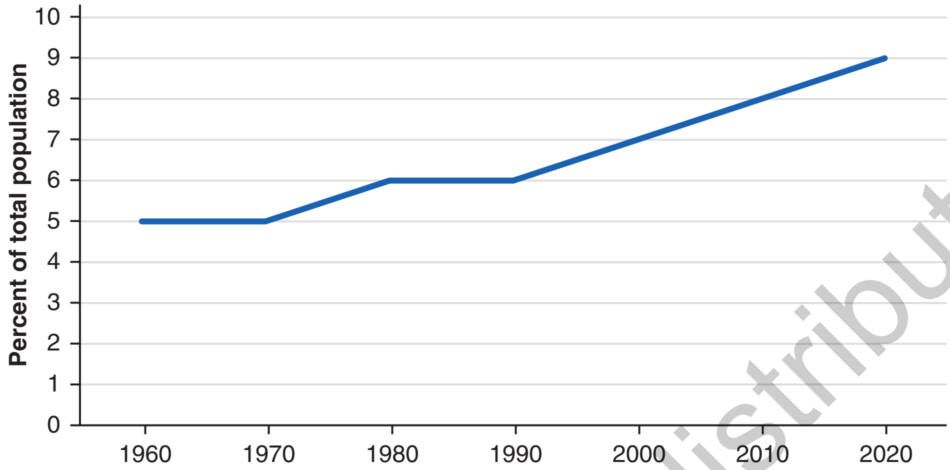
Points representing the frequencies of each category are placed above the midpoint of the category and are joined by a straight line. Notice that in Figure 2.7, the age intervals are arranged on the horizontal axis and the frequencies along the vertical axis. Instead of using bars to represent the frequencies, however, points representing the frequencies of each interval are placed above the midpoint of the intervals. Adjacent points are then joined by straight lines.

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.

Figure 2.8 shows a time series from 1960 to 2020 of the percentage of the total population that is 65 years or older as reported by the World Bank. 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.

FIGURE 2.8 ■ Percentage of World Population 65 Years and Above, 1960–2020



Source: The World Bank, DataBank, Retrieved July 13, 2022 (https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS).

Statistics in Practice: Foreign-Born Population 65 Years and Over

In their 2019 report, *The Older Foreign-Born Population in the United States: 2012–2016*, U.S. Census Bureau researchers Nobuko Mizoguchi, Laquitta Walker, Edward Trevelyan, and Bashiruddin Ahmed describe the foreign-born population ages 65 and over in a series of tables and graphs. We present several for your review.

Frequencies and percentages presented in Table 2.14 summarize three characteristics of the 7.3 million foreign-born older population using data from 2012-2016. About 73% were naturalized citizens. The largest percentage lived in the West (36.1%), followed by the South (29.9%).

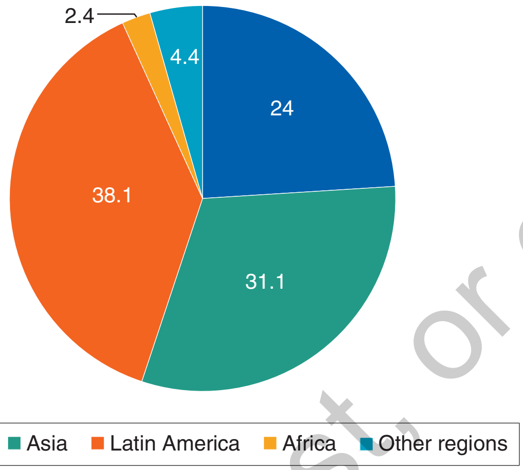
TABLE 2.14 ■ Foreign-Born Population Ages 65 and Over by Citizenship Status, and Region, 2012–2016

Characteristic	Population	Percentage (%)
Citizenship status		
Naturalized citizen	4,453,934	73.4
Not a U.S. citizen	1,611,318	26.6
Region		
Northwest	1,456,012	24.0
Midwest	602,799	9.9
South	1,814,453	29.9
West	2,191,988	36.1

Source: Nobuko Mizoguchi, Laquitta Walker, Edward Trevelyan, and Bashiruddin Ahmed, *The Older Foreign Born Population in the United States: 2012–2016*, American Community Survey Reports, ACS-42, 2019.

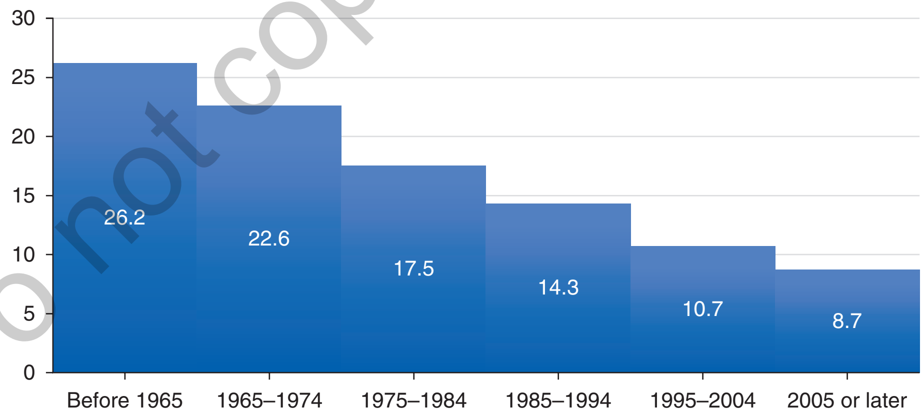
Figure 2.9 is a pie chart, presenting one variable—*world region of birth*. We learn that the majority of the foreign-born older population originally came from Latin America (38.1%), Asia (31.1%), and Europe (24%). The histogram (Figure 2.10) presents the percentage of the older foreign-born by their period of entry. Most of those 65 or older entered the United States before 1984 (26.2% + 22.6% + 17.5%).

FIGURE 2.9 ■ Percentages of Foreign-Born Population Ages 65 Years and Over by World Region of Birth, 2012–2016



Source: Nobuko Mizoguchi, Laquitta Walker, Edward Trevelyan, and Bashiruddin Ahmed, *The Older Foreign Born Population in the United States: 2012–2016*, American Community Survey Reports, ACS-42, 2019.

FIGURE 2.10 ■ Percentages of Foreign-Born Population Ages 65 and Over by Period of Entry, 2012–2016



Source: Nobuko Mizoguchi, Laquitta Walker, Edward Trevelyan, and Bashiruddin Ahmed, *The Older Foreign Born Population in the United States: 2012–2016*, American Community Survey Reports, ACS-42, 2019.

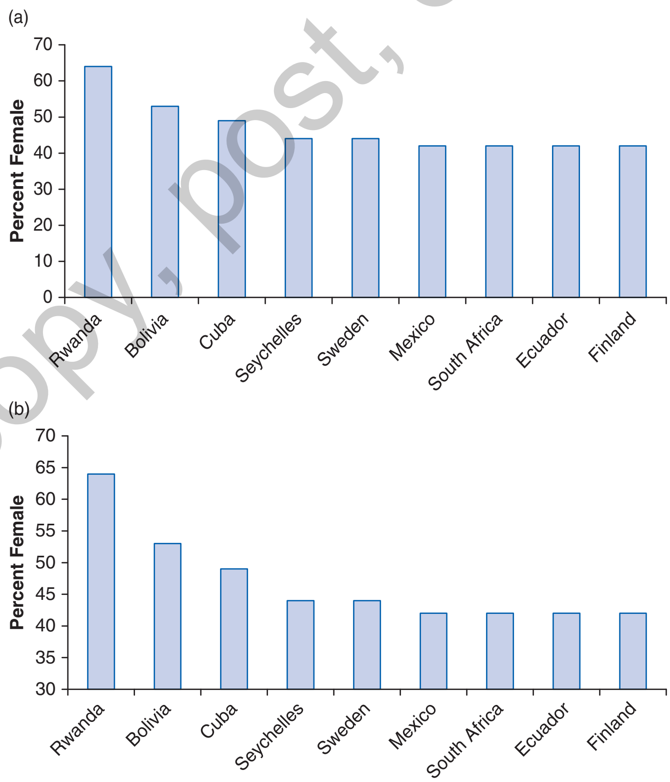
A CLOSER LOOK 2.2

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 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.[7]

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 2.11a, 2.11b) 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 2.11a and 2.11b, 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 2.11b).

FIGURE 2.11 ■ 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. www.ipu.org/wmn-e/classif.htm

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

DATA AT WORK

Kurt Taylor Gaubatz: Graduate Program in International Studies


Photo courtesy of Kurt Taylor Gaubatz

Although Kurt began his college career as a music major, after taking an economics class, he was fascinated by the challenge of understanding and modeling human behavior. He came to see the importance of public policy. "All of the biggest problems we face as a society, indeed, as human beings, comes down to questions in the social sciences. . . . This is a very good area to work on the hardest and most important issues."

According to Kurt, "A research career is a life of posing and answering questions, of trying to think about things in new and more interesting ways. In that sense, then, in addition to getting to work on important social problems, research is a natural outlet for human curiosity."

He advises students to develop two essential research skills. "Information, a huge amount of information, is now increasingly available to everyone who carries a phone. The critical skill[s] [are] knowing how to build new ideas from the organization and analysis of that information, and being able to communicate those ideas effectively. Students need to focus on filling their toolbox with those analytic and communication skills."

MAIN POINTS

- The most basic way to organize 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.
- 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.
- 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 statistical map displays geographic data patterns or variations among categories of an interval-ratio variable.
- 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
statistical map
time-series chart

SPSS DEMONSTRATIONS [GSS21SSDS-A]

To access the data sets for this chapter and view tutorial videos for using SPSS, visit edge.sagepub.com/frankfort10e.

Demonstration 1: Producing Frequency Distributions

In SPSS, you can review the frequency distribution for a single variable or for several variables at once. The frequency procedure is found in the *Descriptive Statistics* menu under *Analyze.* For this chapter, we will use the GSS21SSDS-A data set.

In the Frequencies dialog box, click on the variable name(s) in the left column and transfer the name(s) to the Variable(s) box. More than one variable can be selected at one time.

For our demonstration, let's select the variable POLVIEWS (respondent's political views). Click on *OK* to process the frequency. Respondents were asked to answer the question by indicating 1 = extremely liberal, 2 = liberal, 3 = slightly liberal, 4 = moderate, 5 = slightly conservative, 6 = conservative, and 7 = extremely conservative.

SPSS will produce two tables in a separate Output window, a small statistics table (not presented here), and a frequency table. Use the Window scroll keys to move up and down the window to find the statistics and frequency tables for POLVIEWS. What level of measurement is this variable? (Refer to Chapter 1 to review definitions.)

In the first table, Statistics, SPSS identifies all the valid and missing responses to this question. Responses are coded missing if no answer was given.

In the frequency table (see Figure 2.12), the variable labels for POLVIEWS are shown. The first column lists the value and value label for each category of POLVIEWS. The next four columns contain important frequency information about the variable. The Frequency column shows the number of respondents who gave a particular response. Thus, we can see that 1,500 respondents are included in the data set, but only 1,478 provided a valid response, with 22 responses missing.

The Percent column calculates what percentage of the whole sample (1,500 cases) each of the responses represents. Thus, 11.3% of the total sample indicated slightly liberal political views. In most instances, percentages reported in the third column, Valid Percent, are more useful.

FIGURE 2.12 Frequency Table of POLVIEWS

polviews think of self as liberal or conservative

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 extremely liberal	80	5.3	5.4	5.4
	2 liberal	233	15.5	15.8	21.2
	3 slightly liberal	169	11.3	11.4	32.6
	4 moderate, middle of the road	547	36.5	37.0	69.6
	5 slightly conservative	176	11.7	11.9	81.5
	6 conservative	216	14.4	14.6	96.1
	7 extremely conservative	57	3.8	3.9	100.0
	Total	1478	98.5	100.0	
Missing	System	22	1.5		
Total		1500	100.0		

This column removes all the cases defined as missing and recalculates percentages based only on the valid responses. Recalculated based only on valid cases (1,478), the percentage of those who answered slightly liberal is 11.4. The last column, Cumulative Percent, calculates cumulative percentages beginning with the first response. We know that 32.6% of the valid sample reported that they were liberal (extremely liberal, liberal, or slightly liberal).

Demonstration 2: Recoding Variables

Some variables may need to be recoded or reduced into a smaller number of categories or intervals to better present and understand the data. We could, for example, collapse POLVIEWS into a variable with three categories: (1) liberal, (2) moderate, and (3) conservative. To accomplish this, we could use the SPSS commands *Transform–Recode Into Different Variables*.

For more detailed instruction on recoding variables, please refer to the section on Recoding Variables in the SPSS Appendix on the text's study site, which explains how to recode the variable EDUC (respondent's years of education).

After reviewing the SPSS Appendix, recode POLVIEWS into a new variable called RPOLVIEWS. Frequencies for Recoded POLVIEWS, 3 Categories (RPOLVIEWS) should look like Figure 2.13.

Demonstration 3: Producing a Bar Graph

SPSS greatly simplifies and improves the production of graphics. The program offers a separate choice from the main menu bar, *Graphs*, which lists more than a dozen types of graphs that SPSS can create.

Under the *Graphs* menu, select *Legacy Dialogs* and then *Bar*, which will produce various types of bar graphs. We will use bar graphs to display the distribution of the nominal variable MARITAL (marital status of respondent). After clicking on *Bar*, you will be presented with the initial dialog box, with the *Simple* bar graphs option already selected.

Almost all graphics procedures in SPSS begin with a dialog box that allows you to choose exactly the type of chart you want to construct. Many graph types can display more than one variable (the Clustered or Stacked choices). We will keep things simple here, so click on *Simple*, then on *Define*. When you do so, the main dialog box for simple bar graphs opens.

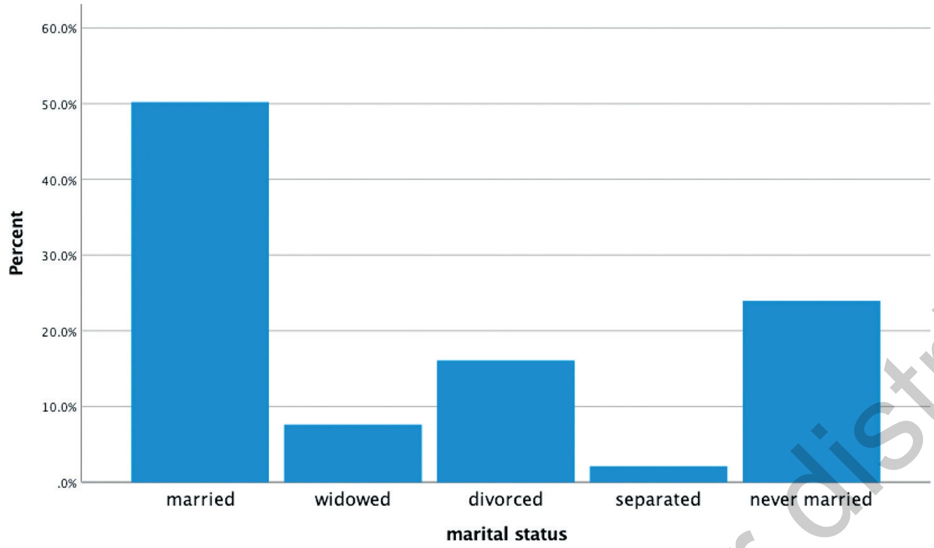
The variable MARITAL should be placed in the box labeled *Category Axis*. In the *Bars Represent* box, click on the *% of cases* radio button. This choice changes the default statistic from the number of cases to percentages, which are normally more useful for comparison purposes.

FIGURE 2.13 ■ Recoding POLVIEWS as RPOLVIEWS

RPolviews Recoded Polviews

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 liberal	482	32.1	32.6	32.6
	2.00 moderate	547	36.5	37.0	69.6
	3.00 conservative	449	29.9	30.4	100.0
	Total	1478	98.5	100.0	
Missing	System	22	1.5		
Total		1500	100.0		

FIGURE 2.14 ■ MARITAL Bar Graph



Click on *OK* to submit your request. (SPSS automatically excludes missing values. You can change this by clicking on *Options*. Click in the box labeled *Display groups defined by missing values* to turn on this choice. Then, click on *Continue*, then on *OK* to submit your request to SPSS.)

The bar graph for MARITAL is presented in an output window labeled SPSS Viewer. You can see in Figure 2.14 that the bar graph for MARITAL has five bars because the only valid responses to this question are (1) "married," (2) "widowed," (3) "divorced," (4) "separated," and (5) "never married."

SPSS graphs can be edited by selecting *Edit*, then *Edit Content, In Separate Window*, which moves the graph to its own window (*Chart Editor*) and displays various editing tools and choices.

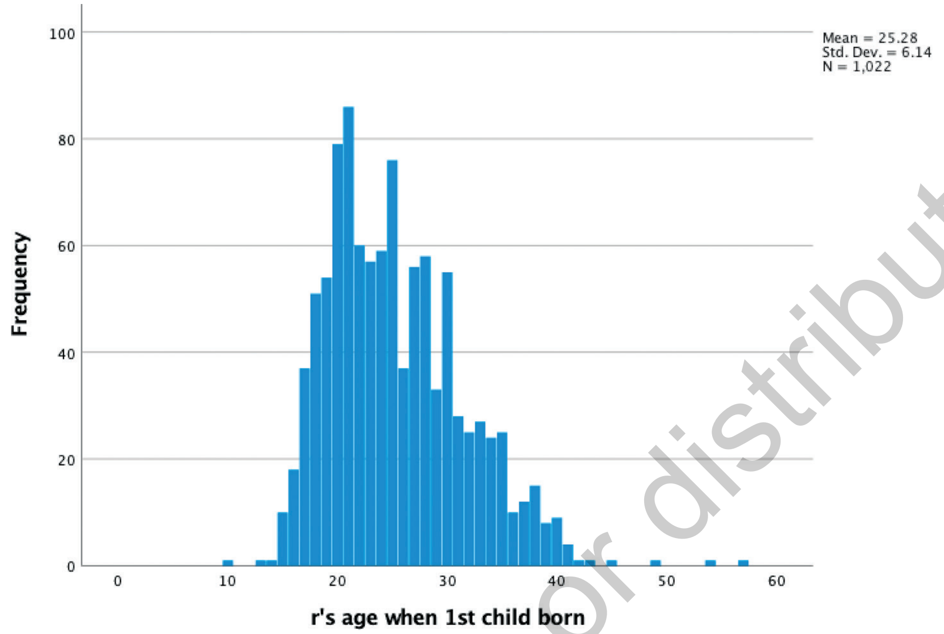
Demonstration 4: Producing a Histogram

Histograms are used to display interval-ratio variables. We'll use the variable AGEKDBRN—respondent's age when first child was born. Under the *Graphs–Legacy Dialogs* menu in SPSS, select the *Histogram*.

Histograms are created for one variable at a time (that's why there was no opening dialog box as for bar graphs). You simply insert (drag) the variable you want to display in the first empty box. You don't need to worry about missing values in histograms; unlike the bar graph default, SPSS automatically deletes them from the display. Notice that SPSS includes icons to indicate the level of measurement for each variable. Interval-ratio variables (or scale variables as SPSS refers to them) is matched with a ruler icon. Click on the *OK* button (on the bottom right-hand corner) to process this request. The resulting histogram is shown in Figure 2.15.

SPSS automatically decided the appropriate width for each interval based on the range of the variable and the optimal number of bars to be displayed on a screen. The histogram also includes the calculation for mean and standard deviation, which will be discussed in Chapters 3 and 4.

FIGURE 2.15 ■ AGEKDBRN Histogram



SPSS PROBLEMS [GSS21SSDS-A]

S1. Use the SPSS Frequencies command to produce a frequency table for the variable CONJUDGE (confidence in the United States Supreme Court).

- **a.** What is the number of valid respondents?
- **b.** Based on the valid number of respondents, what *percentage* of the sample reported a great deal of confidence in the supreme court? What *proportion* of the sample reported hardly any confidence?
- **c.** What is the best way to graphically display these data? Explain.

S2. The GSS included a series of questions on respondents' attitudes about abortion if the woman wants an abortion for any reason (ABANY), is married and doesn't want any more children (ABNOMORE), and is low income and can't afford any more children (ABPOOR).

- **a.** Run frequencies for all the three variables.
- **b.** Prepare a general statement summarizing your results from the three frequency tables. Identify the level of measurement for each variable. How would you describe respondents' attitudes about abortion?

S3. Produce the frequency table for IMMLIMIT (America should limit immigration to protect our way of life).

- **a.** What is the level of measurement for this variable?
- **b.** Identify two independent variables that may affect how someone responded to IMMLIMIT. Explain the relationship between these variables and IMMLIMIT.
- **c.** What is the best way to graphically display these data? Explain.

S4. The GSS2021 respondents were asked to report their highest year of school (EDUC). Using GSS21SSDS-A, run the frequency table for this variable. Collapse this interval-ratio variable into an ordinal measure (omitting those who did not respond to the question). How many categories do you have?

a. Prepare a frequency and cumulative percentage table of your recoded EDUC variable.

b. Prepare a graphic presentation of your recoded EDUC variable.

S5. Determine how best to represent the following variables graphically:

a. PARTYID (respondent's political party identification)

b. SEXORNT (sexual orientation)

c. CONAGE (extent of conflict between young and old)

d. TVHOURS (television viewing hours per week)

Note: Before selecting/constructing your graph, you may want to review the variable by first using the *Frequencies* or *Utilities—Variables* command. The levels of measurement for several variables might be mislabeled in SPSS. If you are using the *Utilities—Variables* option to review each variable and its level of measurement, you should confirm the level of measurement by reviewing the variable's frequency table (*Analyze—Descriptive—Frequencies*).

EXCEL DEMONSTRATIONS [GSS21SSDS-E]

To access the data set for this chapter and view tutorial videos for using Excel, visit edge.sagepub.com/frankfort10e.

Before we begin this chapter's Excel demonstrations, it is important to know that you can paste any tables, graphs, and charts you produce in Excel into a Word document by selecting the object in Excel you've created and then clicking on *Edit* and then *Copy*. You will be able to *Paste* (*Edit* and then *Paste*) your object into a document, email, and so on.

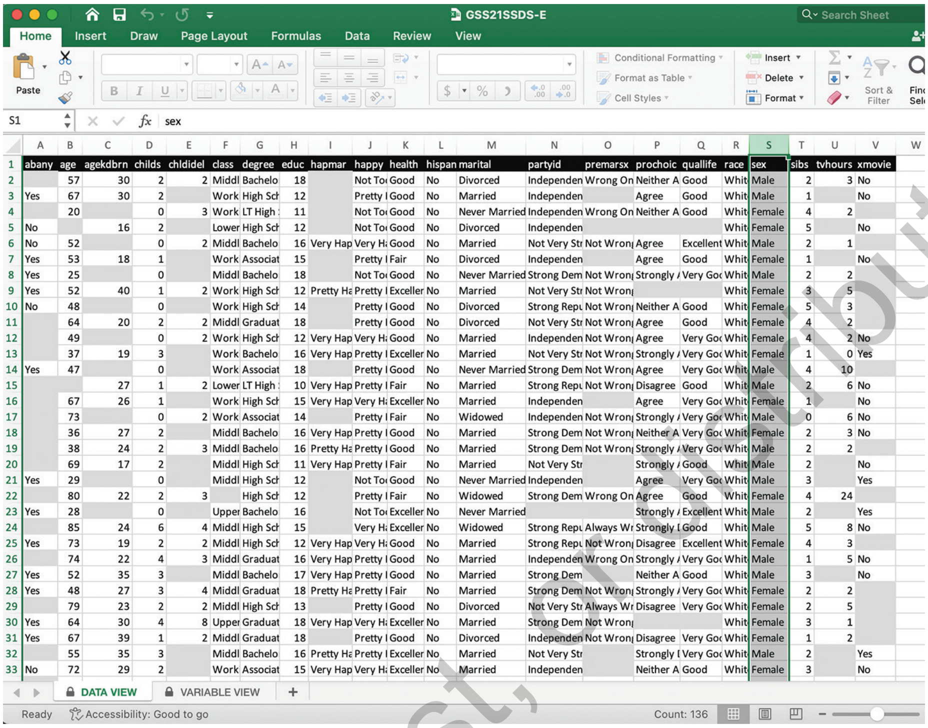
Demonstration 1: Producing a Frequency Distribution

In Excel, you can produce a frequency distribution for one variable at a time. Open the GSS21SSDS-E data set. In the data view sheet of Excel, use your mouse to highlight the column "SEX," including all of the respondents (up to and including row 136). Your screen should look like Figure 2.16.

From the main Excel toolbar, click on *Edit* and then *Copy*. This will copy the SEX column data. On the bottom of the Excel window, next to the sheets *DATA VIEW* and *VARIABLE VIEW*, click on the "+" sign. Doing this will open a blank sheet that's capable of editing. Click on the empty A1 cell, and then select *Edit* and then *Paste*. The SEX column data have been reproduced in your new sheet (see Figure 2.17). You are now ready to produce a frequency distribution and any other graphs or charts you may desire. (*Note:* This copy-and-paste procedure of the data is an extra step because, as explained in the Chapter 1 Excel Demonstration, both the *DATA VIEW* and *VARIABLE VIEW* sheets are in "Protect" mode to avoid any accidental changes to the data set. If you prefer, you can turn off the "Protect" mode and work directly from the *DATA VIEW* sheet. To do so, on the Excel Home Tab, click on *Format* and under *Protection* select *Unprotect Sheet*.)

Click on cell A1 (labeled "SEX"). On the "Insert" Excel tab, select "PivotTable." A Create PivotTable window will open. You will be able to select the range of data you want to analyze.

FIGURE 2.16 ■ Selecting SEX in GSS21SSDS-E



Now select "OK" to continue. Your copied and pasted data should automatically be selected if you clicked on cell A1 as directed before selecting the "PivotTable" option. Under "Choose where to place the PivotTable," select "Existing worksheet," and then click on an empty cell in this sheet. We will click on cell C2 (see Figure 2.18).

A "PivotTable Fields" window will open on the right side of the Excel sheet. In the "Field Name" box, you should see "SEX." Click on "SEX" and drag it to "Rows." Then, drag "SEX" to the "Values" box. Repeat this step three times. Doing so will easily allow you to produce frequencies, percentages, and cumulative percentages. As you follow these steps, you will see Excel begin to construct a frequency distribution that starts in cell C2 and builds out. Click on "Row Labels" in C2. Rename "Row Labels" to "SEX." Doing this will make our frequency distribution more presentable. Click on the little arrow in the lower right-hand corner of C2. A window will open and allow you to uncheck missing data under "Filter" so that the missing data are no longer displayed in your frequency distribution. Now you are ready to double-click on D2. A PivotTable Field window will open. In the "Field name" box, replace "Count of SEX" with "Frequency." Make sure in the "Summarize by" option that "Count" is highlighted. Select *OK.* Double-click on E2. In the "Field name" box, replace "Count of SEX2" with "%." Make sure the option "Count" is highlighted in "Summarize by." In "Show data as," select "% of Column Total" in the dropdown box. Select *OK.* Double-click on F3. In the "Field name" box, replace "Count of SEX3" with "C%" for cumulative percentage. Make sure the option "Count" is highlighted in "Summarize by." In "Show data as," select "% of Running Total In" in the dropdown box. Select *OK.* Your frequency distribution is ready for interpretation (see Figure 2.19).

FIGURE 2.17 ■ Reproducing SEX in a New Excel Sheet

The first column of the frequency table contains the categories (Female/Male) of your variable (SEX). The second column contains the frequency for each category. Notice how there are 133 respondents (79 females and 54 males) in this data set. We've already excluded missing data in this variable, so we can easily interpret the percentage as presented. For example, we could say that our data set is made up of just over 59% females and nearly 41% males. Given our variable is nominal, it wouldn't be useful to interpret the cumulative percentage distribution we created.

FIGURE 2.18 ■ Creating a PivotTable for SEX in an Existing Worksheet

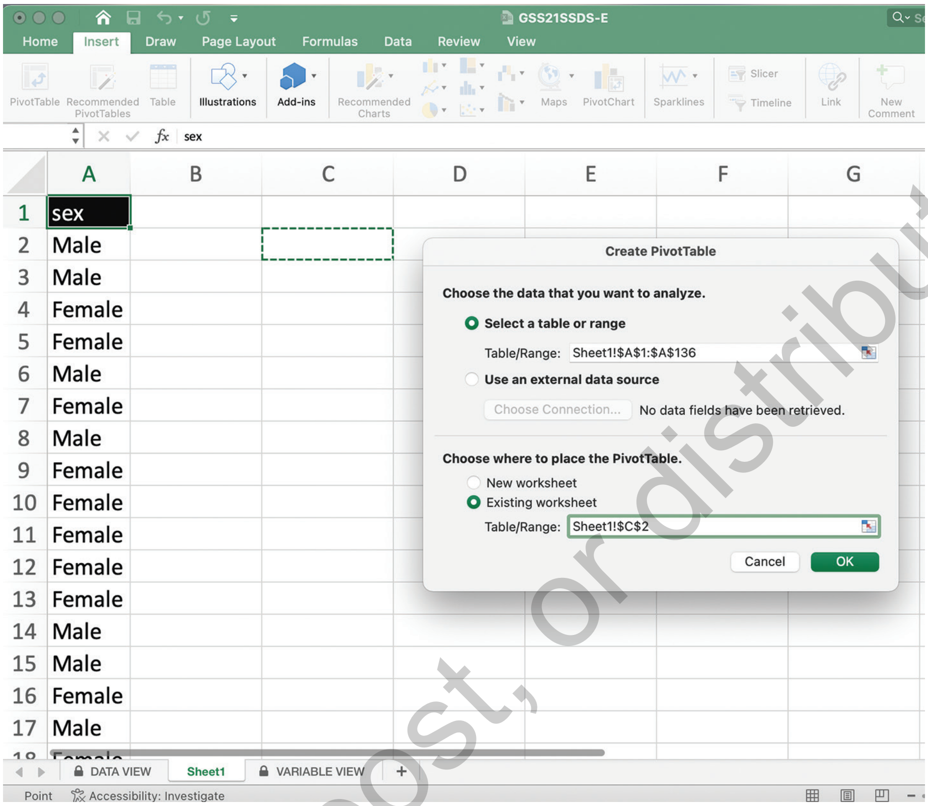
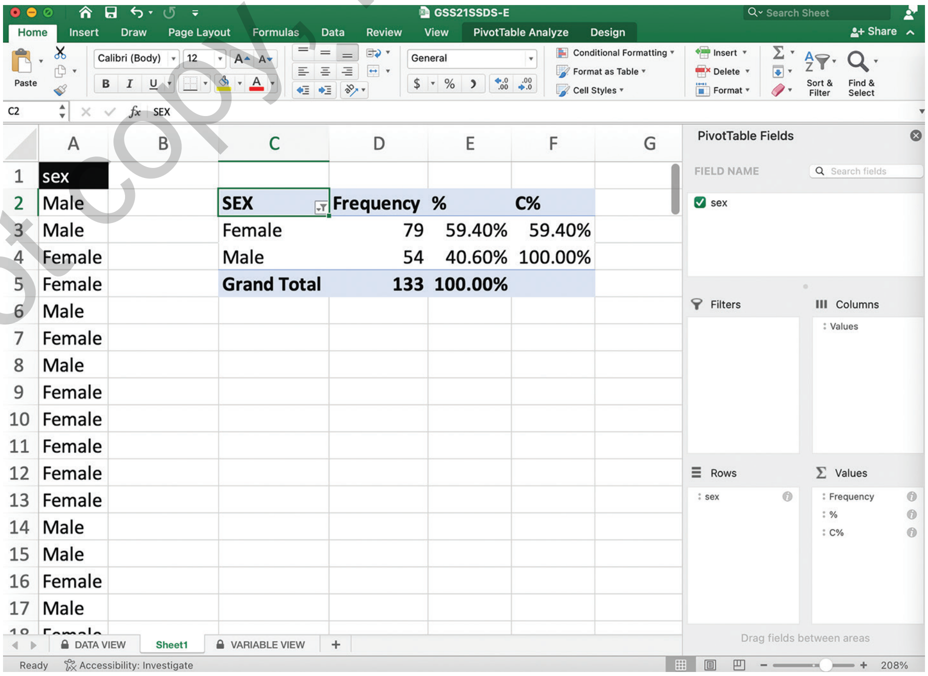


FIGURE 2.19 ■ Summarizing SEX Data in a Frequency Distribution



If you are working with an ordinal or interval-ratio variable, interpreting the cumulative percentage would be beneficial. But, you would need to make sure the categories of the variable are placed in order from low to high or high to low. You can easily order categories in any frequency distribution created in Excel by adding a "1," "2," or "3" in front of each category. Then reorder the categories by clicking on the little arrow in the variable (in our example, this is cell C2). A window will appear. Click on *Sort*, then choose either *Ascending* or *Descending*. Your categories should be ordered from either low to high (ascending) or high to low (descending).

Demonstration 2: Producing a Bar Graph

We will now produce a bar graph using the SEX variable we were working with in Demonstration 1. On the "Insert" Excel tab, next to "Recommend Charts," click on the drop-down image of the bar graph. There are many bar graphs to choose from. Let's choose the first one under "2-D Column" (see Figure 2.20).

In the "PivotChart Fields" window on the right, under "Values," select "Frequencies" and drag it to the "SEX" window at the top. Note: If you accidentally closed the PivotTable Fields panel, click on the Field List button in the PivotTable Analyze tab. This will remove "Frequencies" from the frequency distribution as well as the bar graph we created. Do the same for "C%." You can delete the legend in the bar graph by selecting it and clicking delete on your keyboard. At the top of the bar graph, you should see "Total" as a tentative title. We've clicked on the tentative title and retitled it to read "Respondent's Sex, GSS21SSDS-E" (see Figure 2.21). Excel graphs can be edited by clicking on the bar graph you created and the selecting the Excel "Design" tab. From this menu, you can change the colors of the graph you created and add a chart element (a title, a legend, and more). We suggest you experiment with the many different design options Excel offers its users.

FIGURE 2.20 ■ Creating a Bar Graph for SEX

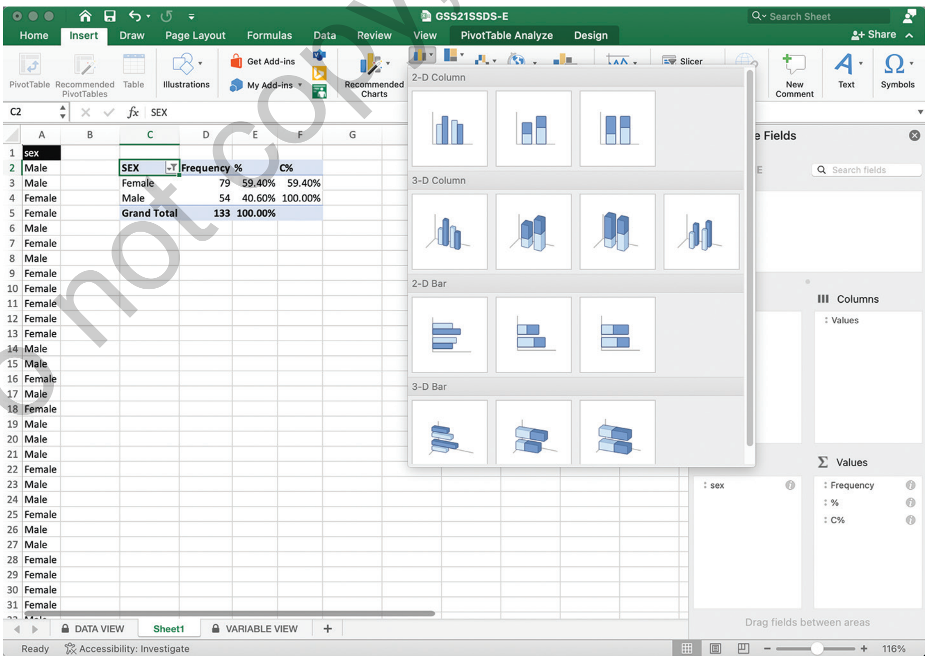
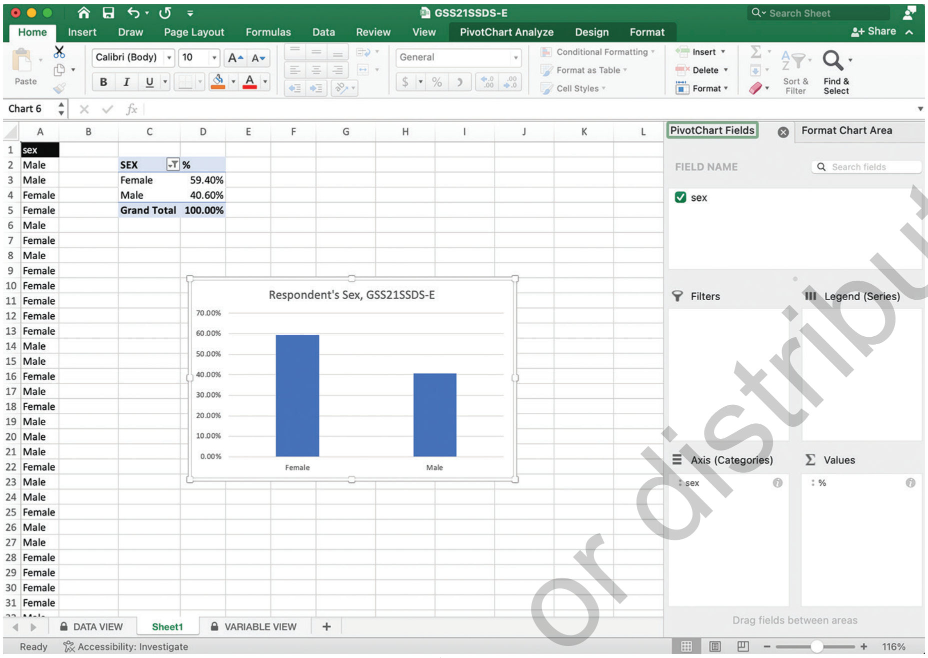


FIGURE 2.21 A Bar Graph of SEX



Demonstration 3: Producing a Pie Chart

Let's produce a pie chart with Excel, again using our SEX variable. On the "Insert" Excel tab, next to "Recommend Charts," click on the dropdown image of the pie chart. There are many pie charts to choose from. Let's choose the first one under "2-D Pie" (see Figure 2.22).

Excel will create a pie chart for the variable SEX. You can retitle the tentative title as we've done to read "SEX, GSS21SSDS-E." You can further edit the pie chart by navigating to the "Design" tab. We've chosen one of the designs Excel offers that includes percentages for the pie slices (see Figure 2.23). As with the bar graph, we suggest you experiment with the many different design options Excel offers its users.

Demonstration 4: Producing a Histogram

We will create a histogram in Excel using the variable CHILDS (number of children). Although the "8 or more" category of the variable technically makes the variable ordinal, we will treat it as interval-ratio for the sake of this demonstration. In the data view sheet of Excel, use your mouse to highlight the column "CHILDS," including all of the respondents (up to and including row 136). From the main Excel toolbar, click on *Edit* and then *Copy*. This will copy the CHILDS column data. On the bottom of the Excel window, next to the sheets *DATA VIEW* and *VARIABLE VIEW*, click on the "+" sign. Doing this will open a blank sheet that's capable of editing. Click on the empty A1 cell, and then select *Edit* and then *Paste*. The CHILDS column data have been reproduced in your new sheet. You are now ready to ask Excel to create a histogram. On the "Insert" Excel tab, next to "Recommend Charts," click on the dropdown image of the histogram. There are three histograms to choose from. Let's choose the first one under "Histogram" (see Figure 2.24).

FIGURE 2.22 ■ Creating a Pie Chart for SEX

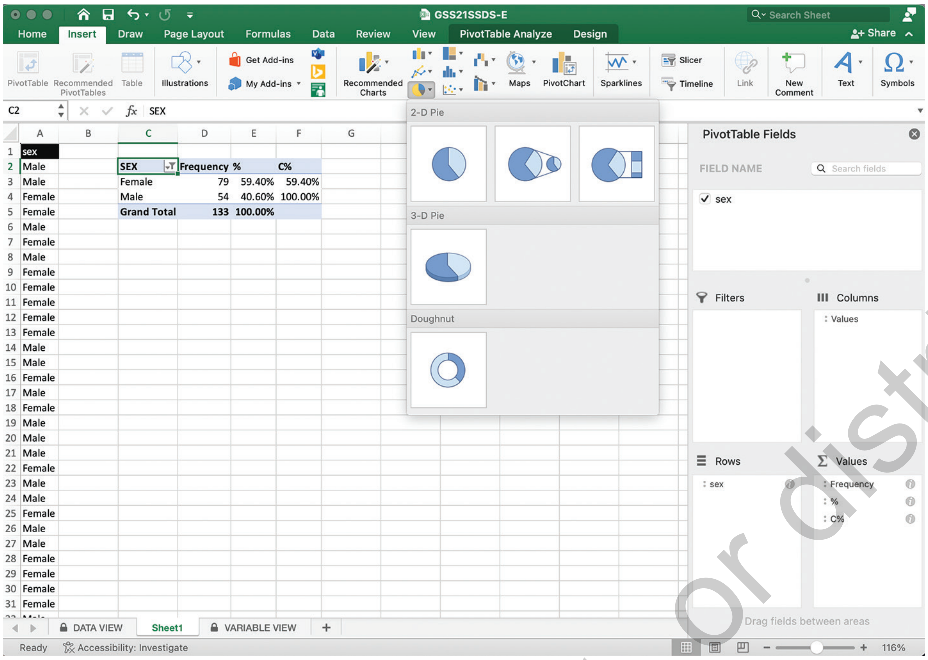


FIGURE 2.23 ■ A Pie Chart of SEX

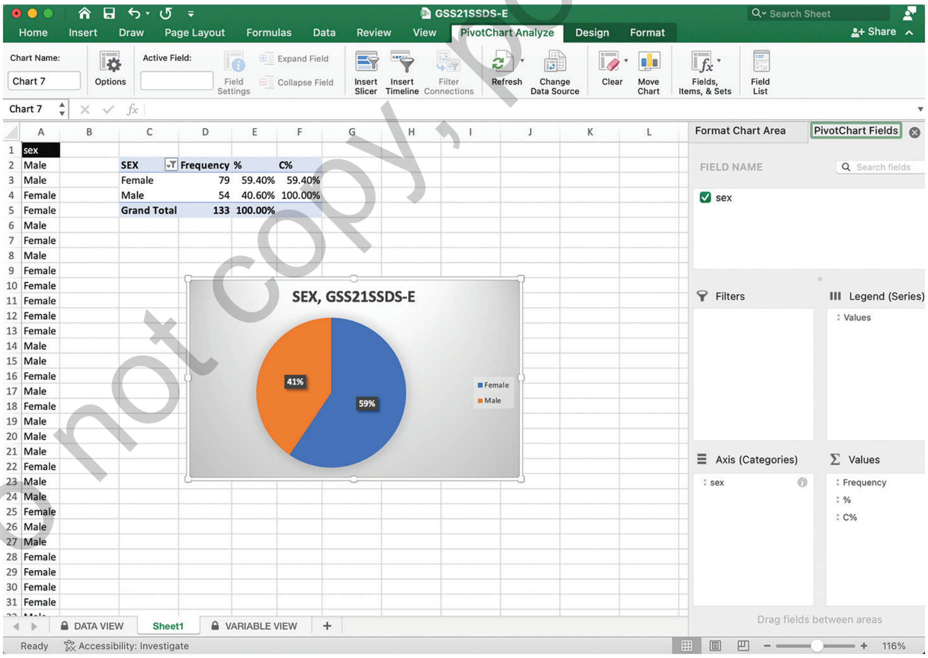


FIGURE 2.24 ■ Creating a Histogram for CHILDS

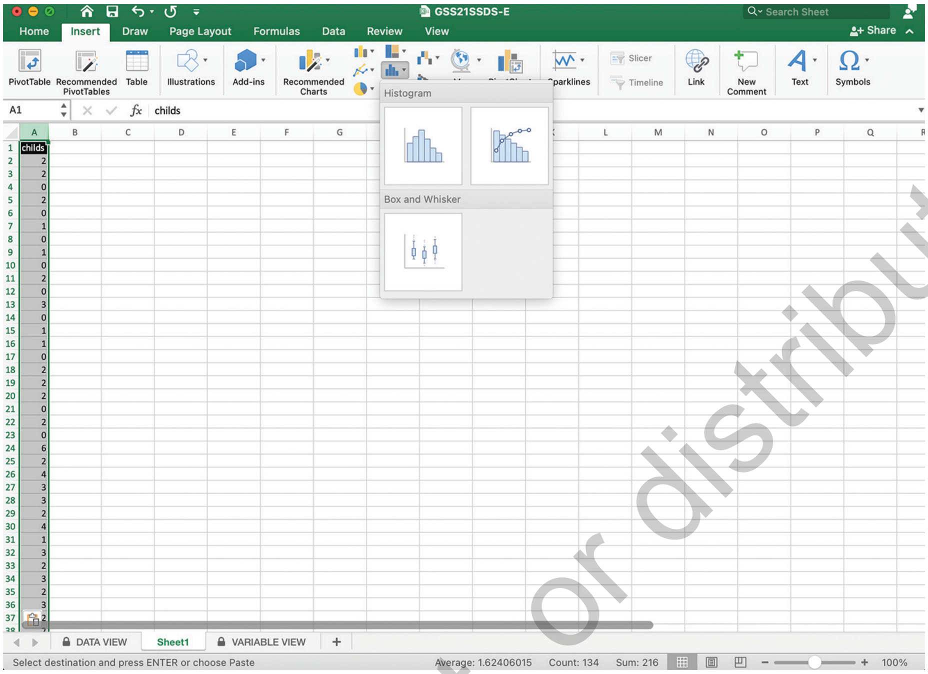
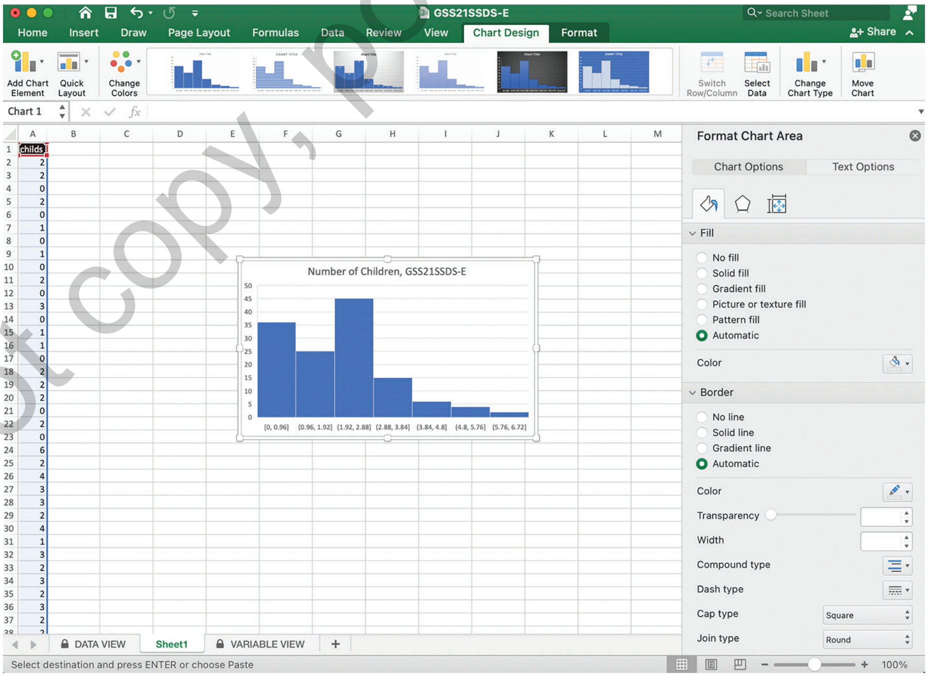


FIGURE 2.25 ■ A Histogram of CHILDS



Excel will create a histogram for CHILDS. You can retitle the tentative title as we've done to read "Number of Children, GSS21SSDS-E." You can further edit the histogram by navigating to the "Chart Design" tab (see Figure 2.25).

EXCEL PROBLEMS [GSS21SSDS-E]

E1. We will use Excel to examine the highest degree our respondents have earned.

- **a.** Produce a frequency distribution for the variable DEGREE. Be sure to order the categories in the frequency distribution from lowest to highest.
- **b.** What percentage of respondents reported their highest degree earned was high school?
- **c.** How many respondents reported their highest degree earned was a bachelor's degree?
- **d.** Considering the level of measurement for DEGREE, have Excel create an appropriate graph or chart. (*Hint*: consider level of measurement.) Be sure to name the variable in the title of the graph/chart, and also identify the data source.

E2. Use Excel to create a histogram of any interval-ratio variable in GSS21SSDS-E. Do not use CHILDS since we used that variable in Demonstration 4 to demonstrate how to produce a histogram in Excel.

- **a.** How many respondents are represented in your chosen variable?
- **b.** In two to three sentences, what does the histogram you've created show us about your chosen variable?

E3. For the following variables in GSS21SSDS-E, identify whether a bar graph, pie chart, and/or histogram is the most appropriate tool to visually represent each variable.

- **a.** RACE (respondent's race)
- **b.** HAPMAR (happiness of marriage)
- **c.** PROCHOIC (consider one's self pro-choice)
- **d.** TVHOURS (hours per day watching TV)

CHAPTER EXERCISES

C1. 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 Table 2.15:

- **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?

TABLE 2.15 ■ Race, Class, and Trauma

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

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

C2. 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?

C3. 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?

C4. Using the data from Exercise 1, construct appropriate graphs showing percentage distributions for race, class, and trauma.

C5. GSS 2021 respondents were asked whether a women could have an abortion for any reason. Results are provided in Table 2.16 for the percentage in each category by political party affiliation. Do these data support the statement that more Republicans do not support abortion under these circumstances than Democrats? Why or why not?

TABLE 2.16 ■ Support for Abortion for Any Reason by Political Party Affiliation

	Democrat (%)	Independent (%)	Republican (%)
Yes	80%	48%	36%
No	20%	52%	64%
Total	100.0	100.0	100.0

Source: GSS 2021

C6. How many hours per week do people spend on email? Data are presented in Table 2.17 for a GSS sample of 99 men and women, who each reported the number of hours they spent per week on email.

- **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 email?
- **c.** What proportion of the sample spent 6 or more hours per week on email?
- **d.** Construct a graph that best displays these data. Explain why the graph you selected is appropriate for these data.

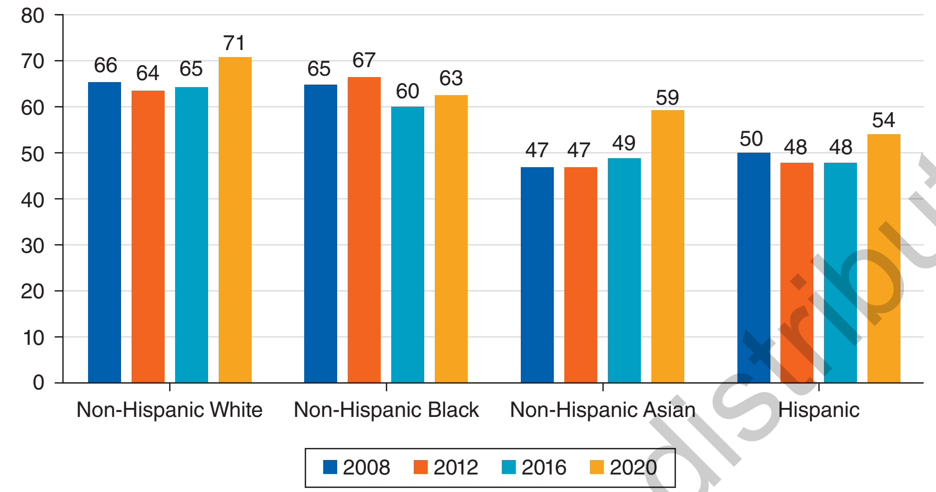
TABLE 2.17 ■ Email Hours per Week

Email 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

Source: GSS 2021

C7. The bar graph in Figure 2.26 displays trends for presidential election voting rates by race and Hispanic origin for 2008–2020. U.S. Census researcher Jacob Fabina (2021) reported that turnout rates in 2020 were higher than in the 2016 election for non-Hispanic white, non-Hispanic Black, non-Hispanic Asian, and Hispanic race and origin groups. Describe the variation in voting rates in 2020 for the four racial and Hispanic origin groups and then compare that variation with that presented for the four groups from 2008 onward.

FIGURE 2.26 ■ Voter Turnout by Race and Hispanic Origin, Presidential Elections 2008–2020



Source: Jacob Fabina, *Despite Pandemic Challenges, 2020 Election had the Largest Increase in Voting Between Presidential Elections on Record*, 2021, https://www.census.gov/library/stories/2021/04/record-high-turnout-in-2020-general-election.html.

C8. According to the Pew Research Center (2020), the country of origin for immigrant populations residing in the U.S. have shifted since the passage of the 1965 Immigration and Naturalization Act. The percent of the U.S. immigrant population by region of origin from 1960 to 2018 are reported in Table 2.18. Write a statement describing the change over time in the percentage of immigrants from the four world regions.

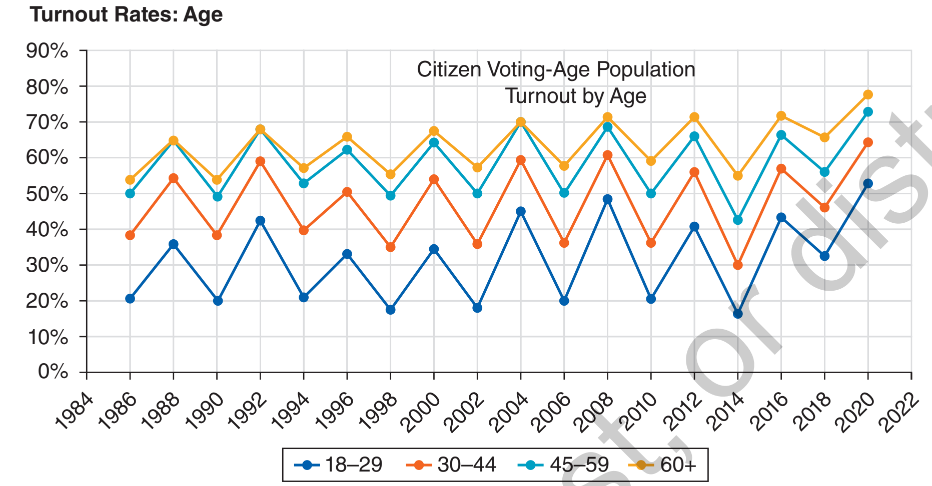
TABLE 2.18 ■ Percent of U.S. Immigrant Population by Region of Origin, 1960–2018

Year	Europe/Canada and other North America	Asia	Other Latin America	Mexico
1960	84%	4%	3%	6%
1970	68%	7%	11%	8%
1980	42%	16%	16%	16%
1990	26%	22%	21%	22%
2000	19%	23%	22%	29%
2010	14%	25%	24%	29%
2015	13%	27%	24%	27%
2018	13%	28%	25%	25%

Source: Abby Budiman, Christine Tamir, Lauren Mora, and Luis Noe-Bustamante, *Facts on U.S. Immigrants, 2018*, 2020, https://www.pewresearch.org/hispanic/2020/08/20/facts-on-u-s-immigrants/#fb-key-charts-language-education.

C9. Older Americans are often described as more politically engaged than younger Americans. One measure of political engagement is election voting. The United States Election Project (2021) presented the following time-series chart (Figure 2.27) of voting rates by age in elections from 1986 to 2020. Based on this time-series chart, are older Americans more likely to vote than younger Americans? (Before you answer, define which age groups are younger vs. older.)

FIGURE 2.27 ■ Citizen Voting—Age Population, Turnout by Age, 1986–2020



Source: United States Election Project, *Voter Turnout Demographics*, 2021, http://www.electproject.org/home/voter-turnout/demographics.

C10. The U.S. Bureau of Justice reports the imprisonment rates of U.S. residents based on sentenced prisoners under state and federal jurisdictions by sex and age as of December 31, 2020. Rates per 100,000 U.S. residents within each group are presented in Table 2.19.

a. Calculate a cumulative percentage distribution for males.

b. Calculate a cumulative percentage distribution for females.

c. Which type of graph would best display the original data in the table?

TABLE 2.19 ■ Imprisonment Rates for Males and Females

Age	All Males	All Females
18–19	0.6	0.4
20–24	7.5	6.9
25–29	14.4	15.8
30–34	16.0	19.5
35–39	15.6	18.5
40–44	13.0	13.3
45–49	10.1	9.7

(Continued)

(*Continued*)

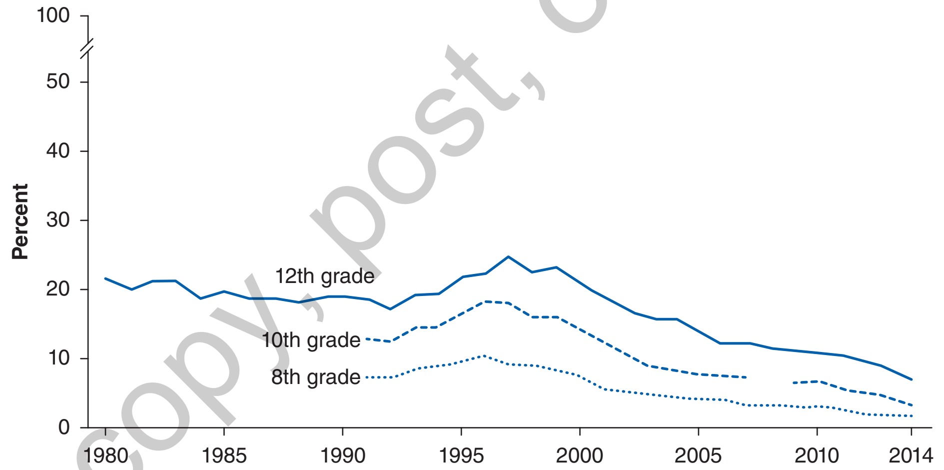
TABLE 2.19 ■ Imprisonment Rates for Males and Females

Age	All Males	All Females
50–54	8.2	7.0
55–59	6.6	4.8
60–64	4.2	2.4
65+	3.7	1.8

Source: E. Ann Carson, *Prisoners in 2020-Statistical Tables*, Bureau of Justice Statistics NCJ 302776, December 2021.

C11. In this exercise, we examine the percentage of 8th-, 10th-, and 12th-grade students who reported smoking cigarettes daily in the past 30 days as reported by the Monitoring the Future survey. Data are reported in this time-series graph (Figure 2.28) from 1980 to 2014. Describe the smoking trends for each group of students.

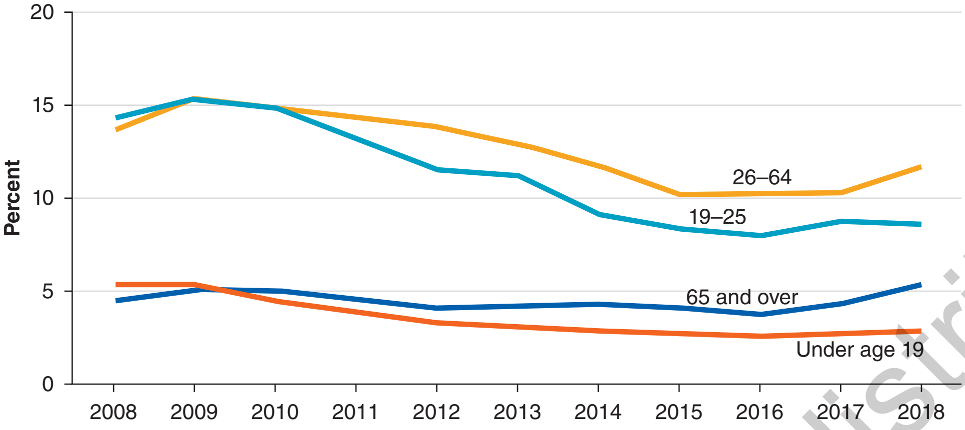
FIGURE 2.28 ■ Regular Cigarette Use by Grade, 1980-2014



Source: Federal Interagency Forum on Child and Family Statistics, *America's Children: Key National Indicators of Well-Being, 2015,*. 2015, https://www.childstats.gov/pdf/ac2015/ac_15.pdf.

C12. Out of pocket costs may be a barrier to obtaining health care in the United States. The National Health Interview Survey measures whether there was any time in the past 12 months when the person needed medical care but didn't get it because the person couldn't afford it or medical care was delayed because of concerns about the cost. The time series chart for 2008–2018 (Figure 2.29) reports the percentage of delay or nonreceipt of needed medical care by four age groups (under age 19, 19–25, 26–64, and 65 and over). Describe the difference in the delay or nonreceipt of needed medical care between these age groups.

FIGURE 2.29 ■ Delay or Nonreceipt of Needed Medical Care in the Past 12 Months Due to Cost, by Age; United States, 2008–2018



Source: National Center for Health Statistics, *Health, United States, 2019*, 2021.

C13. Respondents from Australia, Canada, Mexico, Russia, and United States were asked by the World Values Survey 2017–2020 to indicate their level of confidence in the police on a four-point scale. The percentage of responses per category are reported in Table 2.20. Due to rounding not all row totals will equal 100%.

a. What is the level of measurement for confidence in police?

b. Describe how confidence in police varies by country.

c. Construct a graph that best displays these data.

TABLE 2.20 ■ Confidence in Police in Selected Countries, 2017–2020

	A great deal	Quite a lot	Not very much	None at all
Australia	29.3%	51.3%	15.5%	3.8%
Canada	16.9%	51.6%	25.3%	6.2%
Russia	12.1%	39.6%	30.2%	18.1%
United States	20.8%	47.9%	24.7%	6.5%

Source: World Values Survey Wave 7, 2017–2020.