1 USING SPSS FOR DATA ANALYSIS

LEARNING OBJECTIVES

In this chapter, you will learn how to:

- Use the Data Editor's Data and Variable Views of a dataset.
- Set options for viewing variable lists.
- Execute commands using SPSS's graphical user interface (GUI).
- Select, print, and save results generated by SPSS.
- Format a table of results.
- Write and save a syntax file (a file that contains SPSS command syntax).
- Use SPSS's help system.

Procedures Used
Analyze ▶ Descriptive Statistics ▶ Frequencies
Edit ▶ Options
File ▶ New ▶ Syntax
File ▶ Open ▶ Data
File ▶ Open ▶ Output
File ▶ Print
Format ▶ TableLooks (in Table Editor)
Help
Utilities ▶ Variables

In this chapter, we take readers on a brief tour of the SPSS program. We describe the main windows that students will encounter: the welcome screen, the Data Editor, and the Viewer (equivalent to the console in other statistical analysis programs). For maximum benefit, practice the steps and procedures we discuss here on your own computer. To get the most out of this book, have SPSS running as you read and do what we demonstrate on your own computer.

SUGGESTED READING IN *ESSENTIALS OF POLITICAL ANALYSIS*

We recommended reading Chapter 1: "The Definition and Measurement of Concepts" in *The Essentials of Political Analysis.*

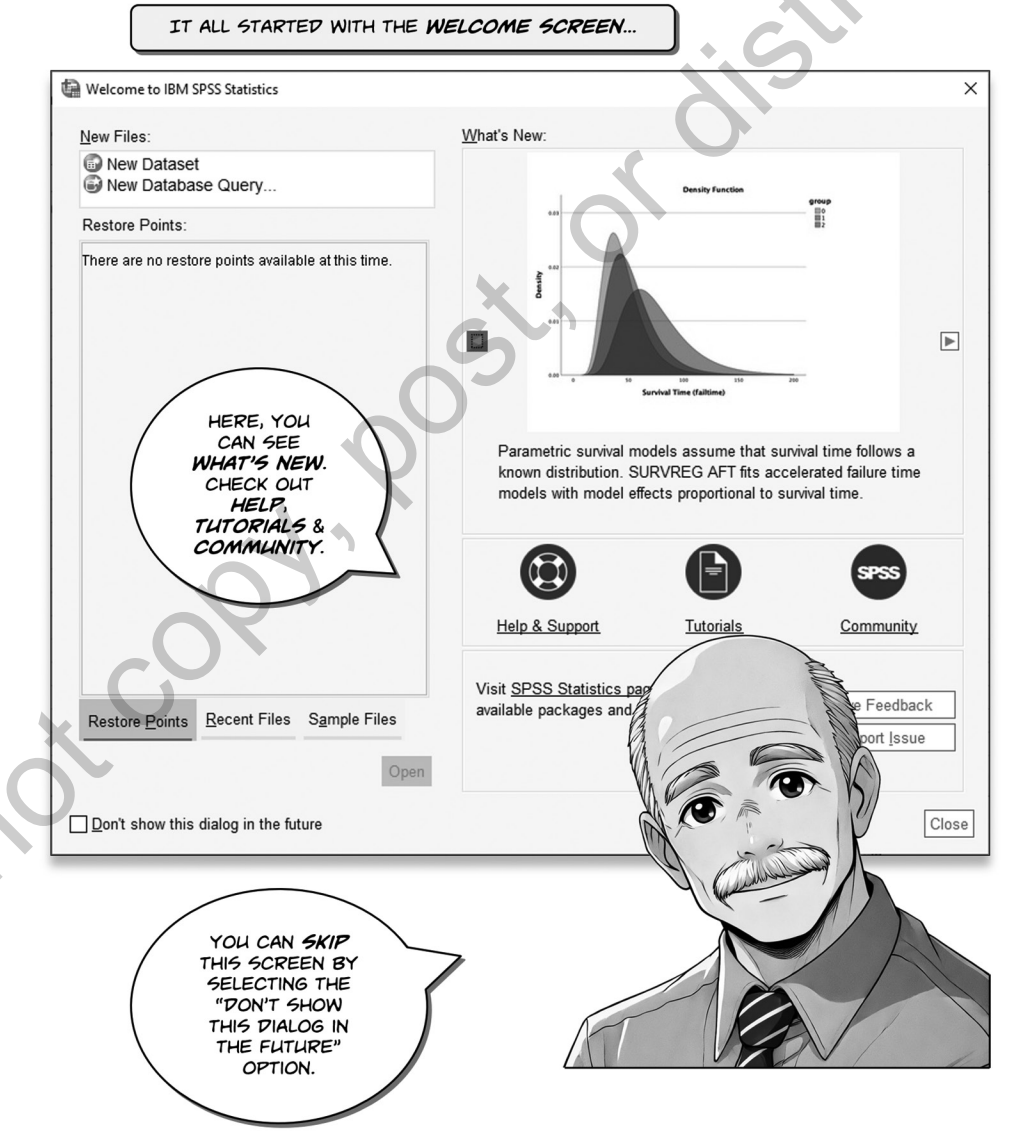
1.1 THE DATA EDITOR

Open the General Social Survey dataset, GSS.sav, to get acquainted with SPSS's **Data Editor**. To open this dataset, locate the GSS.sav file in the folder where you saved it and double-click it. If you already have SPSS running, select File ▸ Open ▸ Data and find the GSS.sav file on your own computer.

Recent versions of SPSS will open several windows at once. You may see a welcome screen (Figure 1-1). You can skip the welcome screen in the future by checking the "Don't show this dialog in the future" option in the lower left corner of the window.[1] Click the "Close" button to close the welcome screen.

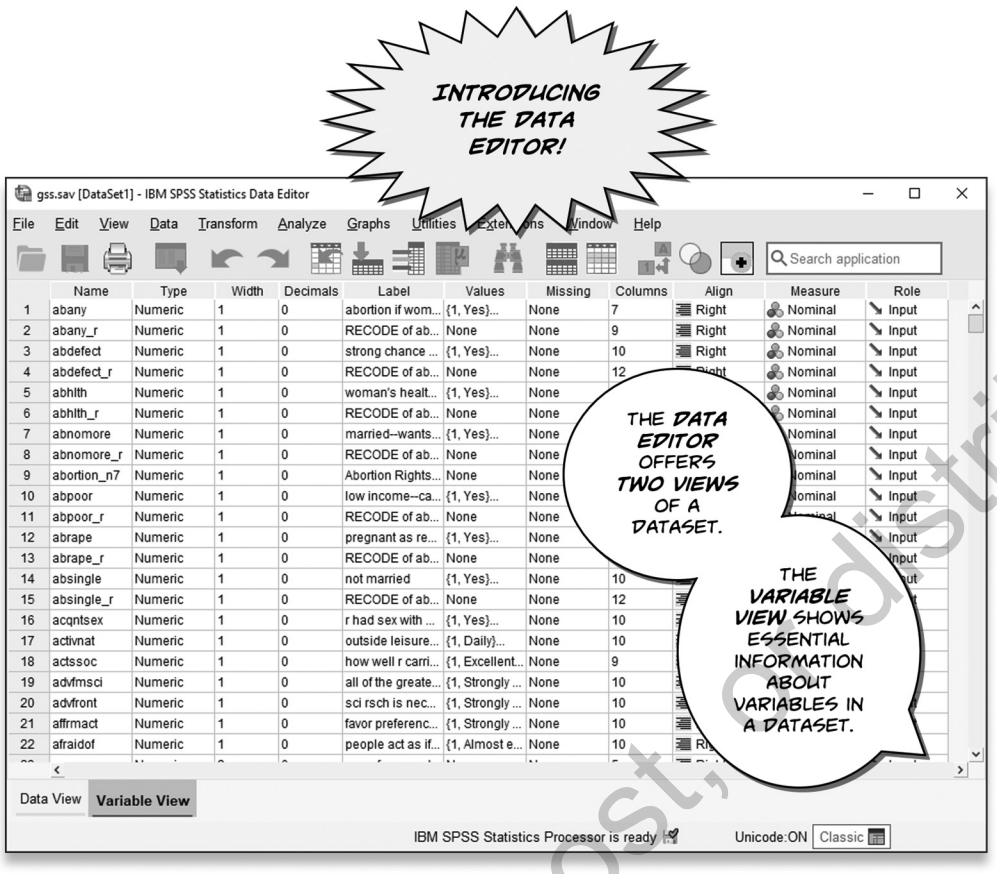
SPSS opens the data file and displays the Data Editor. Notice the two tabs at the bottom of the Data Editor window: **Data View** and **Variable View**. The SPSS Data Editor offers two "views" of a dataset (Figures 1-2 and 1-3). Both views are useful. You can select Data View or Variable View by clicking one of the tabs at the bottom of the Data Editor.

FIGURE 1-1 ■ SPSS "Welcome Screen"



The Variable View shows information about how the GSS variables are coded (Figure 1-2). It shows **metadata**, data about the data. The Variable View, among other helpful features, shows the word labels

FIGURE 1-2 ■ SPSS Data Editor in Variable View



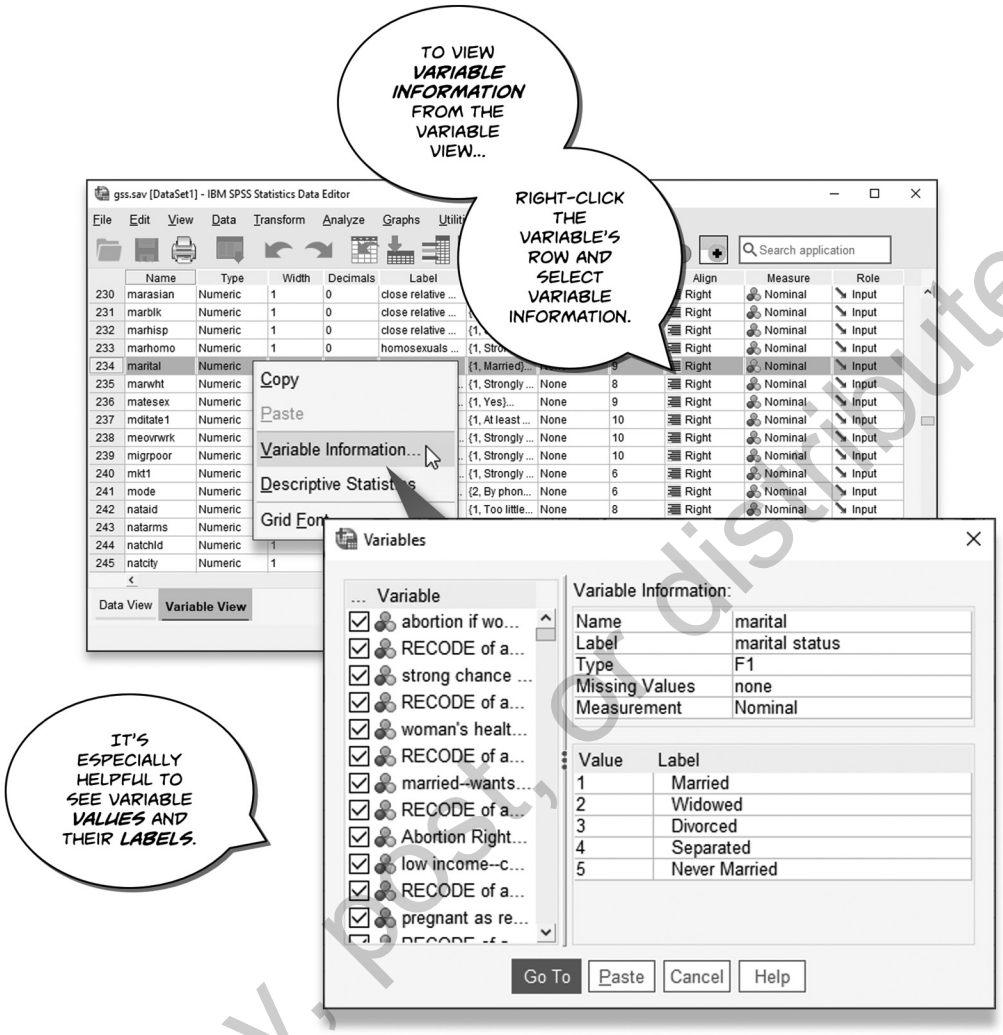
that the researcher has assigned to the numeric codes. This view shows complete information on the meaning and measurement of each variable in the dataset. (You can adjust the width of a column by clicking, holding, and dragging the column border.)

The most frequently used variable information is contained in the Name, Label, Values, and Missing fields. Name is the brief descriptor recognized by SPSS when it does analysis. Names can be up to 64 characters in length, although they need to begin with a letter (not a number). Plus, names must not contain any special characters, such as dashes or commas, although underscores are okay.

Because **variable names** are short and often abbreviated, researchers will use **variable labels**, long descriptors (up to 256 characters are allowed), to provide more detailed information about variables. For example, when SPSS analyzes the GSS variable "granborn," it will look in the Variable View for a Label. If it finds one, then it will label the results of its analysis by using Label instead of Name. So granborn shows up as "How many grandparents born outside U.S."—a bit more descriptive than "granborn."

Just as Label permits a wordier description for Name, Values attaches descriptive labels to a variable's numeric values. We can examine the **value labels** for the gender and race variables to find out what it means when they're coded 1. To find out the value labels for the gender variable, find its row in the Variable View, click the mouse anywhere in the Values cell, and then click the gray button that appears. A Value Labels window opens, revealing the labels that SPSS will attach to the numeric codes of gender. Unless you instruct it to do otherwise, SPSS will apply these labels to its analysis of the gender variable. Repeat this process to find out what value 1 on the race variable signifies or what the numeric codes of the "marital" variable mean (see Figure 1-3). You can see that respondents who have never been married are coded 5 on the variable marital. Click the Cancel button in the Value Labels window to return to the Variable View.

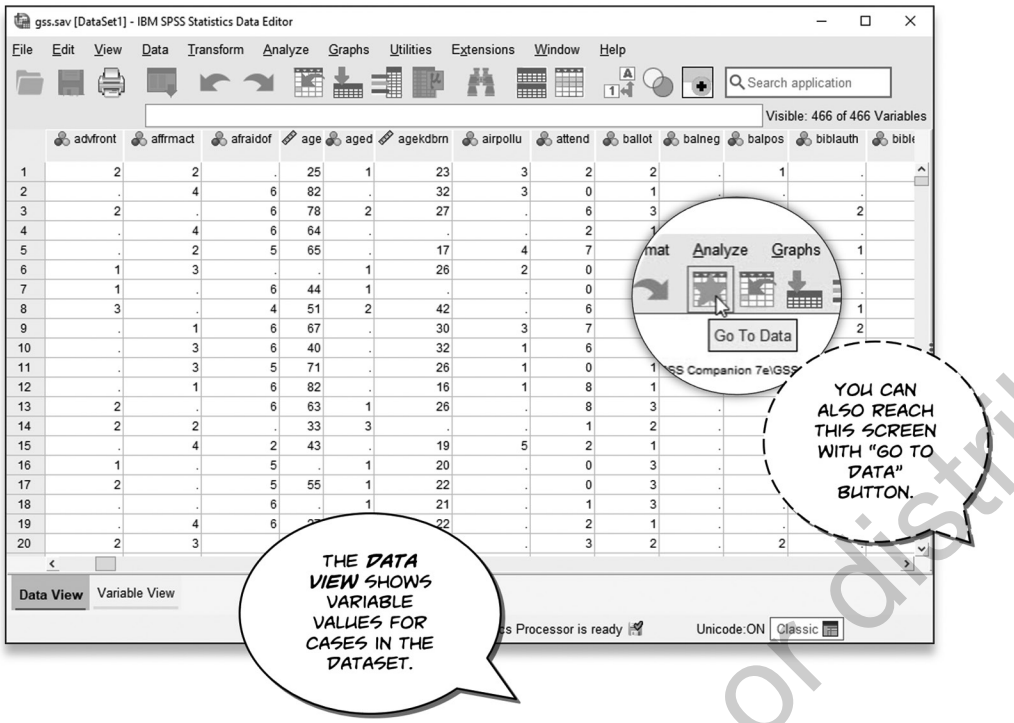
FIGURE 1-3 ■ Value Labels of Variable in the Data Editor



Finally, a word about the Missing column. Sometimes a dataset does not have complete information for all variables and observations. **Missing values** occur for a variety of reasons; researchers may add or remove questions from the survey, some questions may not apply to everyone, or the response may not be clear. In coding the data, researchers typically give special numeric codes to missing values. In coding mobile16, for example, the GSS coders entered a value of 0 for respondents who were not asked the question ("IAP"), 8 for respondents who did not know ("DK"), and 9 when the information was otherwise not available ("NA"). Because these numeric codes have been set to missing (and thus appear in the Missing column), SPSS does not recognize them as valid codes and will not include them in an analysis of mobile16. In many cases, the author has set most missing values in the datasets to *system-missing*, which SPSS automatically removes from the analysis. However, when you use an existing variable to create a new variable, SPSS may not automatically transfer missing values on the existing variable to missing values on the new variable. Later in this volume, we discuss how to handle such situations.

Turn your attention to the Data View. (Make sure the Data View tab is clicked.) The Data View shows how all the cases are organized for analysis (see Figure 1-4). Information for each case occupies a separate row. When you're working with the GSS dataset, each row represents a person who participated in the survey. Numbers in the "id" column record each respondent's case identification number ("caseid"). The variables, given brief yet descriptive names, appear along the columns of the editor.

FIGURE 1-4 ■ SPSS Data Editor in Data View



Scroll right to see information recorded from the first respondent. You can tell that the first respondent in the dataset is 25 years old (see the first value in the "age" column). You can also see that this respondent has a 1 in the "aged" column and a 2 in the "attend" column. Storing information as numbers is efficient, but it's not immediately clear what it means for a person to have aged value of 1 and attend value of 2. To paint a more complete word portrait of this respondent, you need to see how the variables are coded.

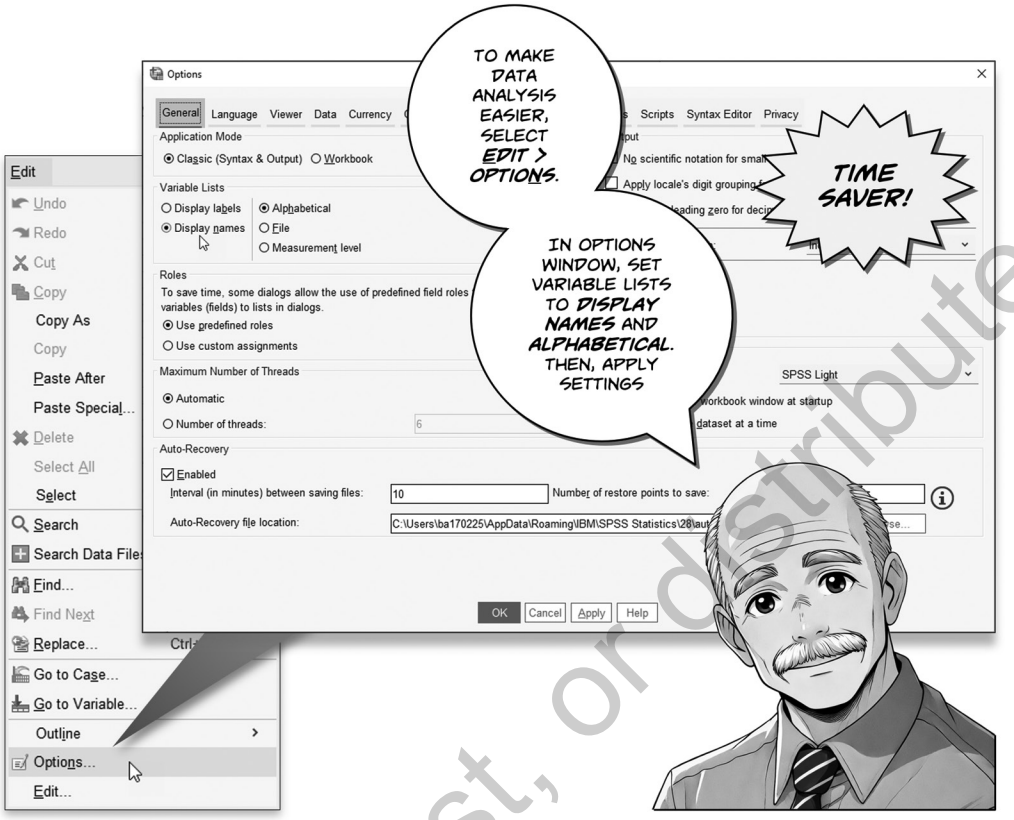
1.2 SETTING OPTIONS FOR VARIABLE LISTS

Now you have a feel for how data are organized and stored in SPSS. Before looking at how SPSS produces and handles output, you must do one more thing. To ensure that all the examples in this workbook correspond to what you see on your screen, you will need to follow the steps given in this section when you open each dataset for the first time.

DO THIS NOW: In the main menu bar of the Data Editor, select Edit ▸ Options. Make sure that the General tab is clicked (see Figure 1-5). If the **Variables Lists** options for Display names and Alphabetical are not already selected, select them (as in Figure 1-5). Click Apply and then OK, returning to the Data Editor. When you open a new dataset for the first time, go to Edit ▸ Options and ensure that Display names/Alphabetical are selected and applied. This will help you find variables to analyze more efficiently. (If the radio button Display names *and* the radio button Alphabetical are already selected when you open the Options menu, you are set to go and can click the Cancel button.)

You may not need to alphabetize variables in the GSS Dataset; the variables have been sorted and saved alphabetically, but you'll work with other datasets and will find it useful to set display options so you can work efficiently.

FIGURE 1-5 ■ Setting Options for Variable Lists



A CLOSER LOOK: VARIABLES UTILITY

Although the names of GSS variables are not terribly informative, SPSS makes it easy to view complete coding information. In the text, we show how you can access information about variables in the Data Editor's Variable View. You can also view detailed information about the variables in a dataset using Utilities ▶ Variables. This selection will yield the Variables window (see Figure 1-3).

Suppose you want to view detailed information about the GSS variable marital to better understand how the dataset records respondents' marital statuses. Scroll through the alphabetical list of variables on the left side of the Variables window until you find "marital" and select it. You'll then see some essential information about the variable, like the text label associated with it ("Marital Status"), along with a breakdown of how different marital statuses are encoded.

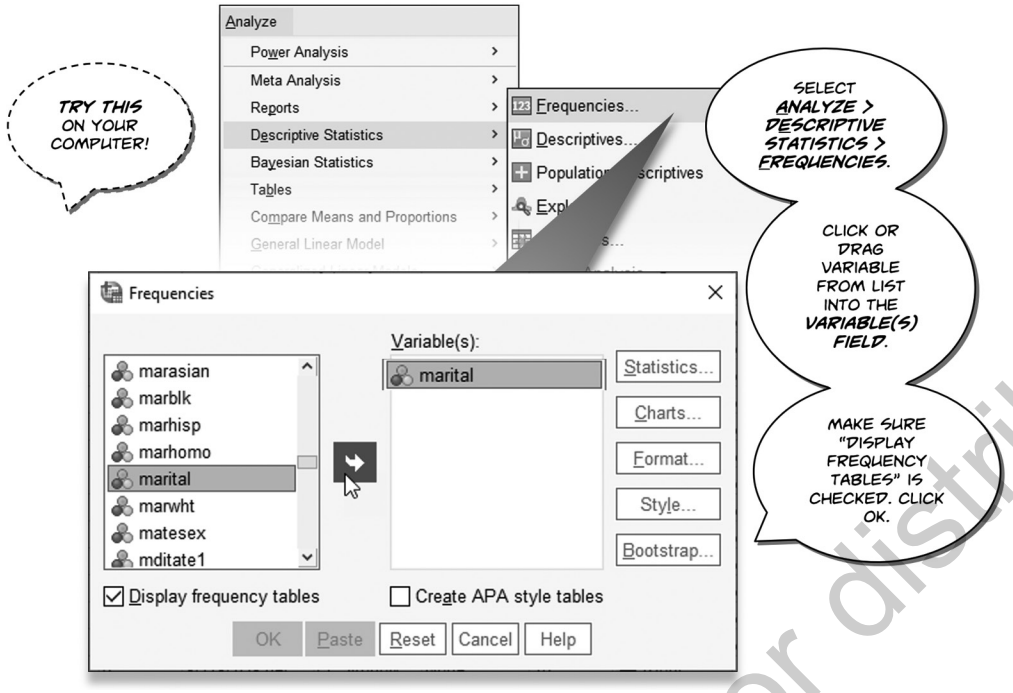
1.3 EXECUTING SPSS COMMANDS

We will run through a brief example to show how SPSS analyzes variables and generates output. You execute most SPSS commands using a graphical user interface (GUI). SPSS's commands for analyzing variables are organized under the "Analyze" tab. You'll start most of your data analysis by clicking the Analyze tab and selecting the type of analysis you wish to perform from its menu of options. The Analyze tab appears at the top of both the **Viewer** and Data Editor. We prefer accessing GUI commands from the Viewer because the results appear in the Viewer, but either way works fine. For this example, select Analyze ▶ Descriptive Statistics ▶ Frequencies. The Frequencies window appears (Figure 1-6).

You'll notice that the Frequencies window has two panels. On the right is the (currently empty) Variable(s) panel. This is the panel where you enter the variables you want to analyze. On the left, you see the names of all the variables in GSS in alphabetical order, just as you specified in the Options menu.[2]

Scroll down the alphabetized list of GSS variables window until you find marital and add "marital" to the Variable(s) panel. (*Hint*: Click anywhere on the variable list and type "m" on the

FIGURE 1-6 ■ Using SPSS's GUI to Analyze Frequencies



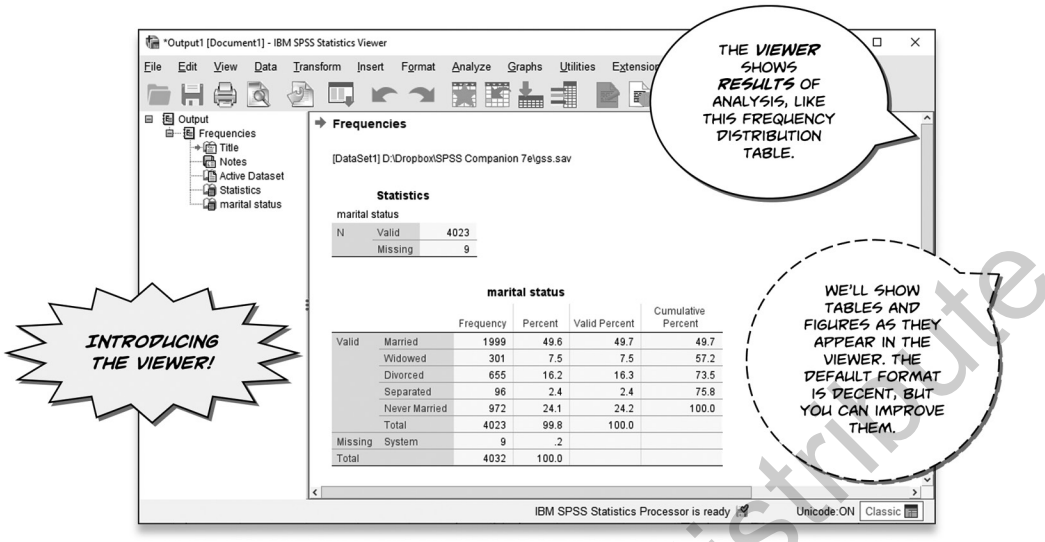
keyboard. SPSS will jump to the first m's in the list.) To add marital to the Variable(s) panel, click on marital and then click the arrow between the panels, or drag and drop marital from the alphabetical list to the Variable(s) panel.[3] Click OK. SPSS runs the analysis and displays the results in the Viewer (Figure 1-7).

A frequency distribution table for the marital statuses of GSS respondents appears in the Viewer. We'll have more to say about frequency distribution tables in the next chapter and discuss many different types of tables in this book. In the future, we'll focus on the tables SPSS generates and won't show the entire Viewer as we do in Figure 1-7. For now, focus on the SPSS workflow: datasets open in the Data Editor, you execute commands with the GUI, and results of analysis appear in the Viewer.

When you execute a procedure using the GUI, SPSS temporarily stores the information you entered so you can return to the same window and adjust your selections. This is particularly useful when you're executing complex commands for the first time, making graphics, or performing the same operation repeatedly on different variables. After running the frequency analysis illustrated in Figures 1-6 and 1-7, for example, you could select Analyze ▸ Descriptive Statistics ▸ Frequencies again and find marital still on the variables list, making it easy to change the command settings or analyze a different variable.

It's helpful to understand some general features of commands and the GUI. Commands are selecting from a menu. As you can see from the Frequencies example, the menu has several levels, starting with a general category of commands like "Analyze," which is then subdivided into commands for "Descriptive Statistics," which includes the "Frequencies" command. When you select a command, SPSS opens a new window for the command. The command's window is organized like a form with fields to fill in and a selection to make. Some commands windows have a button that open more specialized windows to adjust different aspects of the command. The Frequencies command (Figure 1-6), for example, has one field for "Variable(s)," two selection boxes, five buttons on the right side that open more specialized windows, and a set of buttons on the bottom that are common to all commands. We refer to the specialized windows for command suboptions as dialogs. Command windows have "OK" buttons you press to run commands, whereas command dialogs have "Continue" buttons that return to command windows. Command dialogs are not accessed directly from menus; they are suboptions for commands.

FIGURE 1-7 ■ Sample Table Output in the SPSS Viewer



A CLOSER LOOK: KEYBOARD SHORTCUTS

Some SPSS users may find navigating the GUI cumbersome after a while. Those who prefer the keyboard to the mouse will be happy to know that there is an easy way to navigate the GUI using **keyboard shortcuts**. To get to the Frequencies window, hold down the "Alt" key and press "A," "E," and then "F" (you don't need to capitalize the letters or use quotation marks). You can navigate the SPSS menu by pressing Alt followed by the letter(s) corresponding to different branches of the menu. If you look closely at the command notation, Analyze ▶ Descriptive Statistics ▶ Frequencies, you'll see that we've underlined the letters A, e, and F to show the keyboard shortcuts and we follow this convention throughout the book when we introduce new procedures.

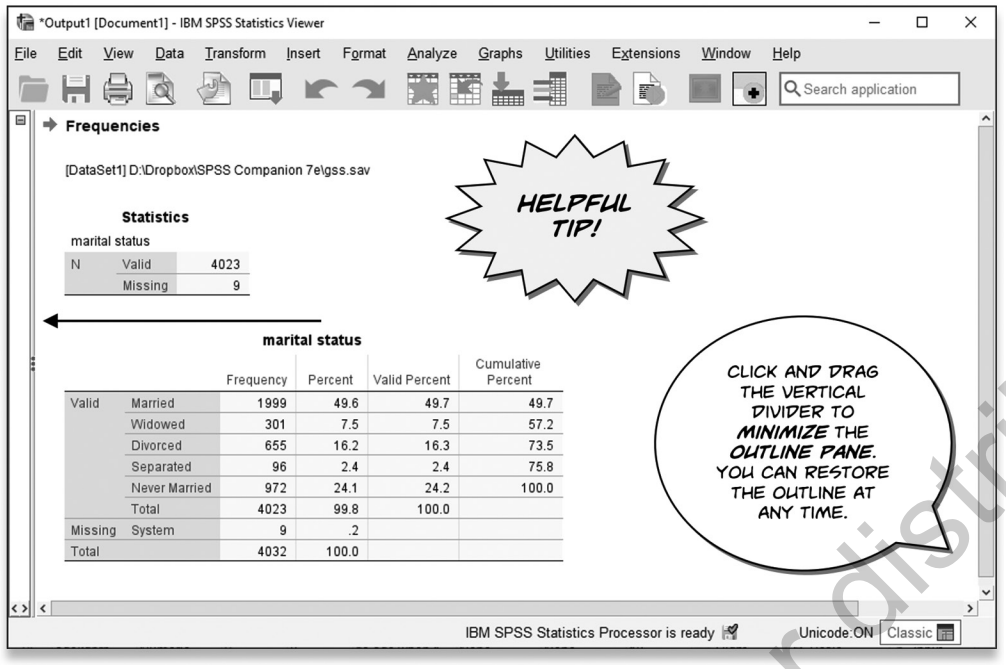
Notice that the SPSS Viewer has two panes. In the **Outline pane**, SPSS keeps a running list of the analyses you are performing. The Outline pane references each element in the Contents pane, which reports the results of your analyses. In this book, we are interested exclusively in the Contents pane.

You can minimize the Viewer's Outline pane by first placing the cursor on the Pane divider. Click and hold the left button of the mouse and then move the Pane divider over to the left-hand border of the Viewer. The Viewer should now look like Figure 1-8. The Contents pane shows you the frequency distribution of the marital variable with value codes labeled. In Chapter 2, we discuss frequency analysis in more detail. Our immediate purpose is to become familiar with SPSS output.

Here are some key points about the Viewer to keep in mind:

- The Viewer is a separate file, created by you during your analysis of the data. The Viewer file, a log of your analysis and output, is completely distinct from the data file. Whereas SPSS data files all have the file extension *.sav, Viewer files have the file extension *.spv. The output can be saved, under a name that you choose, and then reopened later in SPSS. You can't open a *.spv file in another program, like a word processor (e.g., Microsoft Word); if you want to use SPSS output in a word processing document, follow the directions below for exporting graphics and copying tables.
- Output from each succeeding analysis does not overwrite the Viewer's *.spv file. Rather, it appends new results to the Viewer file. If you were to run another analysis for a different variable, SPSS would dump the results in the Viewer below the analysis you just performed.

FIGURE 1-8 ■ SPSS Viewer with the Outline Pane Minimized



- The quickest way to return to the Data Editor is to click the starred icon on the menu bar, as shown in Figure 1-8. And, of course, Windows accumulates icons for all open files along the bottom Taskbar.
- The "Analyze" tab appears at the top of the Viewer and the Data Editor, so you can start your data analysis from either SPSS window. In fact, all tabs in the Data Editor window appear in the Viewer window (along with two tabs that appear only in the Viewer: Insert and Format). Because the results of your analysis appear in the Viewer, it makes sense to start your analysis there, but you can get the same results starting from the Data Editor.

As we discuss in the next two sections, you may select any part of the output file, format it, print it, or copy and paste it into a word processing program.

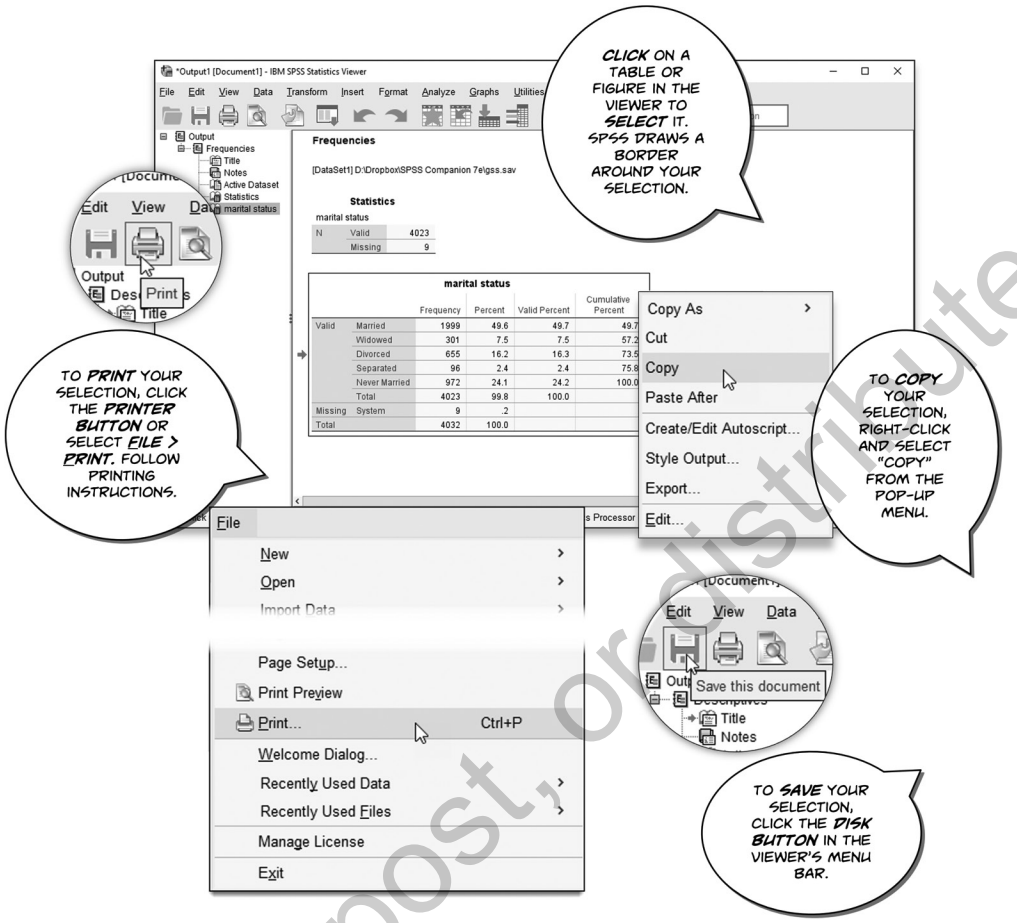
1.4 SELECTING, PRINTING, AND SAVING OUTPUT

Many of the exercises in this workbook will ask you to print the results of your SPSS analyses or incorporate results in a document, so let's cover these procedures. We'll also address a routine necessity: saving output.

Printing desired results requires, first, that you select the output or portion of output you want to print. A quick and easy way to select a single table or chart is to place the cursor anywhere on the desired object and click once to select that object. For example, if you want to print the marital frequency distribution table produced in the preceding section, place the cursor on the frequency table and click. A red arrow appears in the left-hand margin next to the table (Figure 1-9). Click the Printer icon on the Viewer menu bar, select File ▸ Print, or right-click the object and select print. (There is usually more than one way to do something in SPSS.) The Print window opens. In the window's Print Range panel, the radio button next to "Selected output" should already be clicked. Clicking OK sends the frequency table to the printer.

To select more than one table or graph, hold down the Control key (Ctrl) while selecting the desired output with the mouse. Thus, if you want to print the frequency table and the statistics table, first click

FIGURE 1-9 ■ Selecting, Printing, and Saving Output



on one of the desired tables. While holding down the Ctrl key, click on the other table. SPSS will select both tables.

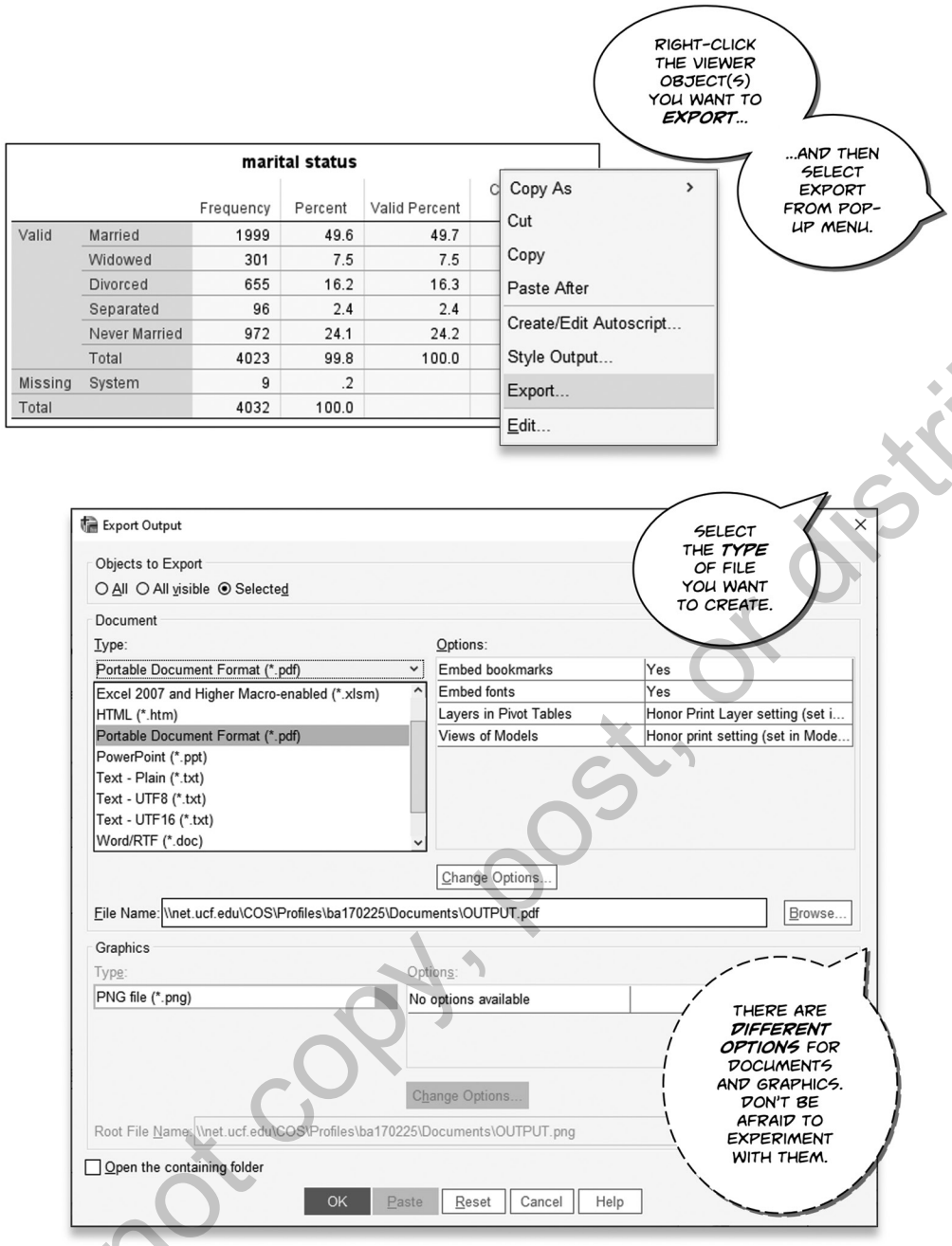
To copy your Viewer output to your computer's clipboard to paste into another document, simply select the table(s) you want to copy, right-click, and select the "Copy" option (see Figure 1-9). Recent versions of SPSS have greatly improved table formatting over prior versions, and the table copied from the Viewer now looks decent in a word processing document. (For additional table-formatting tips, see Section 1.5.)

To save your Viewer output, simply click the familiar Save icon on the Viewer menu bar (refer to Figure 1-9 again). Browse for an appropriate location. Invent a file name (but preserve the ".spv" extension), such as "chap1.spv," and click Save. SPSS saves all of the information in the Viewer to the file chap1.spv. Saving your output protects your work. Plus, the output file can always be reopened later. Suppose you are in the middle of a series of SPSS analyses and you want to stop and return later. You can save the Viewer file, as described here, and exit SPSS. When you return, you start SPSS and load a data file (like GSS.dta) into the Data Editor. In the main menu bar of the Data Editor, you select File ▸ Open ▸ Output, find your .spv file, and open it. Then you can pick up where you left off.

If you want to save a table or graphic that appears in the Viewer but don't want to save all of your Viewer output, select the Viewer object you want to save and right-click it. Select "Export . . ." from the pop-up dialog window. If you've selected a table, you can export it to a variety of document types, such as a .pdf document. If you've selected a graphic, you can create a variety of different types of image files.

When you execute a command in SPSS, the program tends to generate all the results you might want to see. Some will be more relevant to you than others are. You may need to sift through tables in the Viewer to find the items that interest you the most as if you're eating at a buffet. SPSS is user-friendly and does excellent analysis, but when it comes to printing, saving, and sharing the results of your

FIGURE 1-10 ■ Exporting Viewer Output

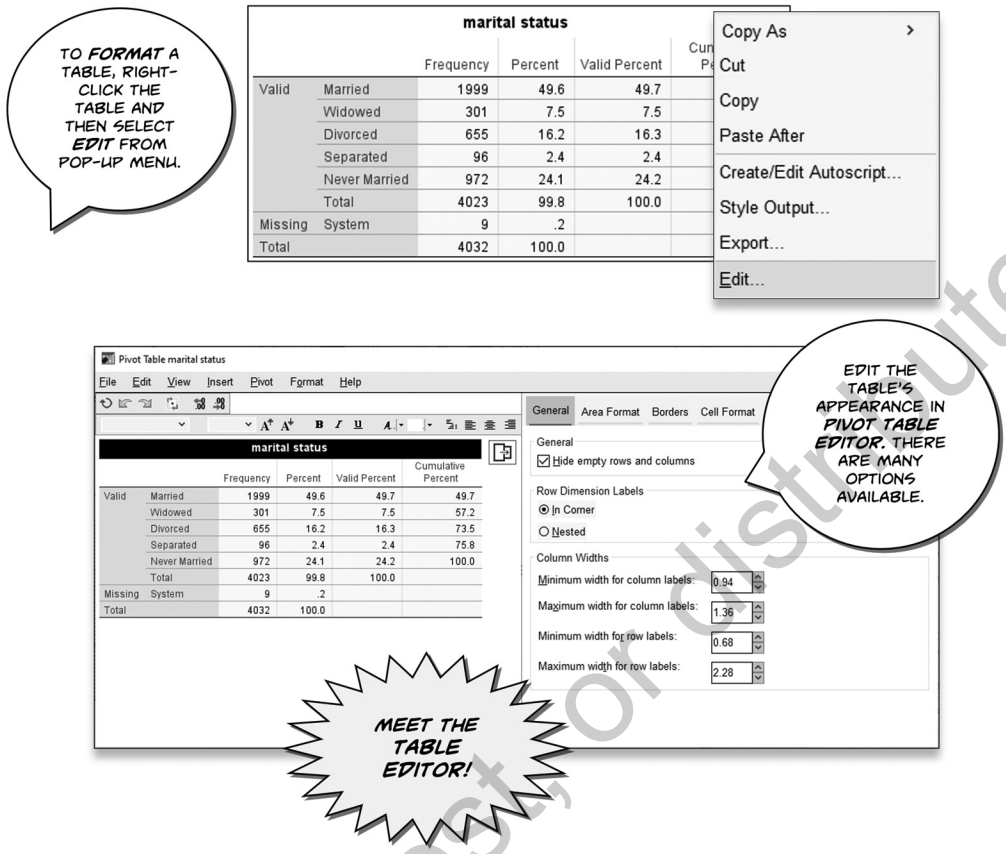


analysis, you should be selective and focus on what's relevant in the situation. SPSS has been around for a relatively long time, and to be completely honest, its tables and figures have a dated appearance. Fortunately, there are some easy ways to format and style SPSS tables and figures, as we discuss in the next section.

1.5 HOW TO FORMAT AN SPSS TABLE

When you analyze political science data, you'll produce a lot of tables of results. You want the tables you create to communicate the results of your analysis effectively. No one wants to try to decipher results from a mess of numbers. Fortunately, SPSS offers some easy-to-use options for formatting tables.

FIGURE 1-11 ■ Formatting a Table



To access the table formatting tools, select the table you want to format and right-click it (like you would to copy or export it). Now, select the "Edit" option. You'll see your table in a new, separate window with formatting tools (see Figure 1-11). SPSS's **Table Editor** allows you to change the look and feel of your table; you can change the colors, shading, fonts, alignments, and more. For example, you can widen the far-right column of the marital status frequency distribution table so the heading "Cumulative Percent" stays on one line. Keep in mind what you're attempting to communicate. If you're conducting serious analysis for an academic paper, a tropical color theme isn't the best choice.

Here's a suggestion to help you quickly create professional-style tables. The **Academic-style tables** are good for academic purposes. In the separate table editing window, select Format ▸ TableLooks (this procedure is only available in the table editing window) (Figure 1-12). TableLooks is a predefined set of table styles. There are multiple defined styles, but the "Academic" style is a solid choice for most of your political analysis. Select "Academic" from the list of TableLook files and click OK.

You'll see your table with your TableLook selection applied in the separate table editing window. (You can also use Edit > Options in Table Editor to limit the number of digits displayed after decimal points in tables.) When you close the separate table editing window, you should see your freshly formatted table in the Viewer. If you selected the Academic look for the marital variable's frequency distribution table, your table should look like Figure 1-13.

In this book, we'll show tables in the SPSS default style to make it easier to follow our examples, but you can make the Academic-style table your default format by selecting Edit ▸ Options and then the Pivot Tables tab, selecting Academic from the TableLook options, and then clicking OK. Tables in this predefined SPSS table format look a lot like the tables one sees in top political science journals. It's a good look for your political analysis.

FIGURE 1-12 ■ Creating Academic-Style Tables

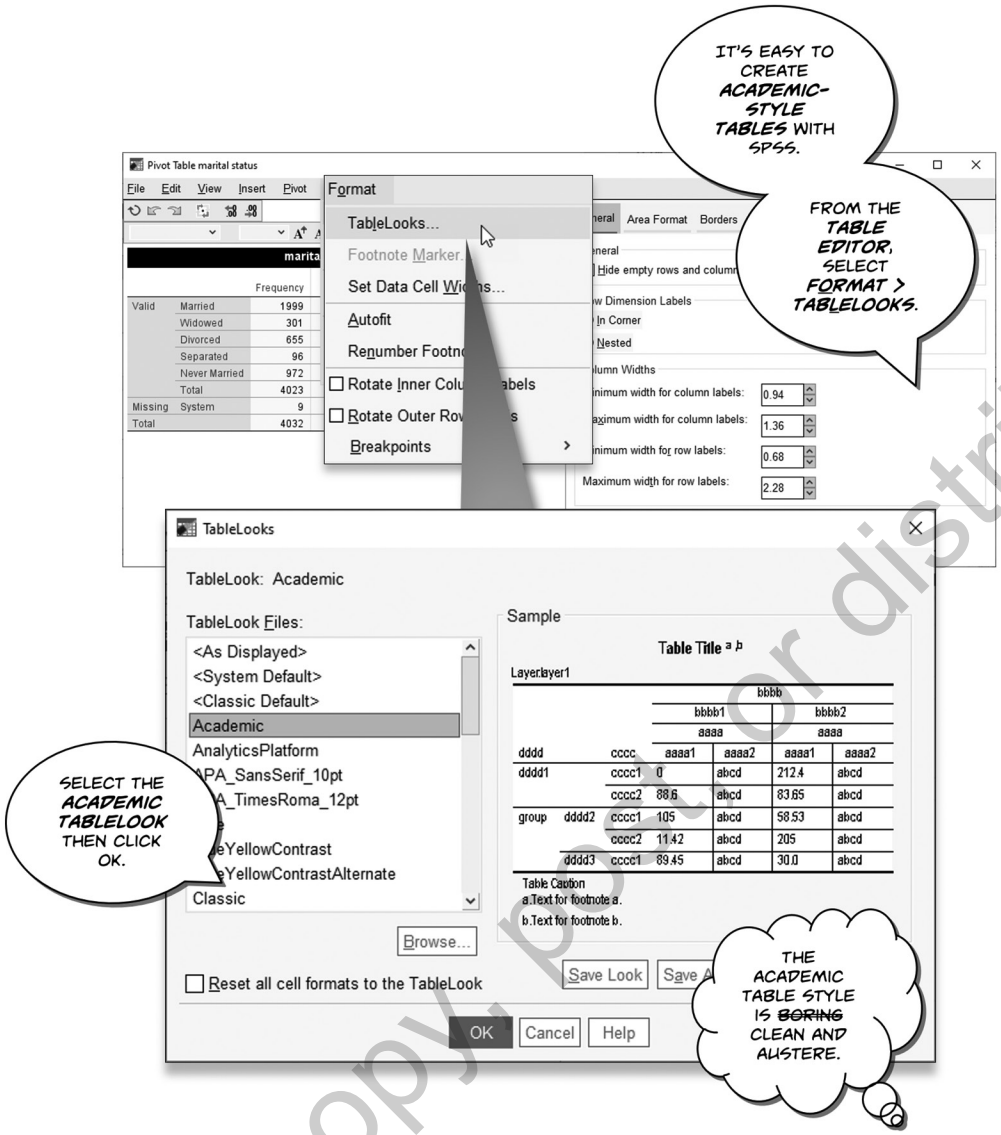


FIGURE 1-13 ■ Example of an Academic-Style Table

marital status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Married	1999	49.6	49.7	49.7
	Widowed	301	7.5	7.5	57.2
	Divorced	655	16.2	16.3	73.5
	Separated	96	2.4	2.4	75.8
	Never Married	972	24.1	24.2	100.0
	Total	4023	99.8	100.0	
Missing	System	9	.2		
Total		4032	100.0		

1.6 SAVING COMMANDS IN SYNTAX FILES

In this book, we show how to implement essential political science research methods using SPSS's graphical user interface. SPSS's GUI offers a straightforward and consistent approach to analyzing

data. In some situations, however, you may want to document your data analysis to enable others to see what you did and replicate your analysis. These situations call for making a syntax file.

A **syntax file** is a text document with the *.sps file extension that records the series of commands used to perform data analysis. The procedures you execute using SPSS's GUI can also be stated as text commands.

SPSS can display the text equivalent of your commands in the Viewer. Some versions of SPSS will display the syntax for GUI commands by default; SPSS version 29 no longer does this, but there is an option to display syntax for commands. To see syntax for commands, select Edit ▸ Options and then the Viewer tab. Check the "Display commands in the log" option and then click OK. Here is the text-equivalent of the frequencies command that generated the output in Figure 1-8:

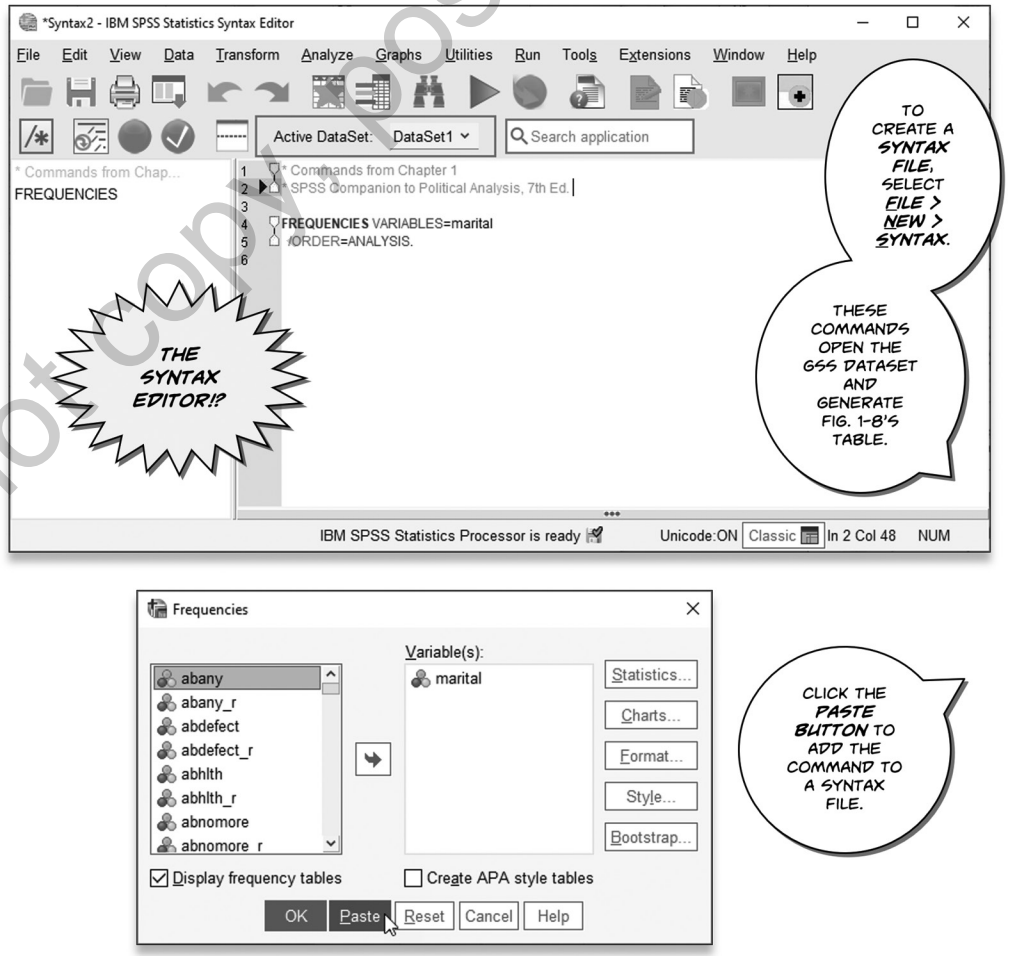
```
FREQUENCIES VARIABLES=marital
  /ORDER=ANALYSIS.
```

This text-command equivalent of the frequency analysis demonstrated above, executed from a syntax file, would yield the same results as using the GUI, enabling someone else to replicate the analysis.

This book demonstrates the essentials of political analysis without using the esoteric SPSS command language. For most data analysis tasks, the GUI works fine and will allow you to start applying core concepts much sooner than encoding commands. Thankfully, SPSS makes it easy for users to "reverse engineer" a syntax file for replication purposes.

To create a syntax file, select File ▸ New ▸ Syntax. This selection will call up a new window, the Syntax Editor (see Figure 1-14, which shows commands executed in this chapter). As we've seen, SPSS prints the text-command equivalent of procedures implemented through its GUI in the Viewer. You

FIGURE 1-14 ■ The Syntax Editor



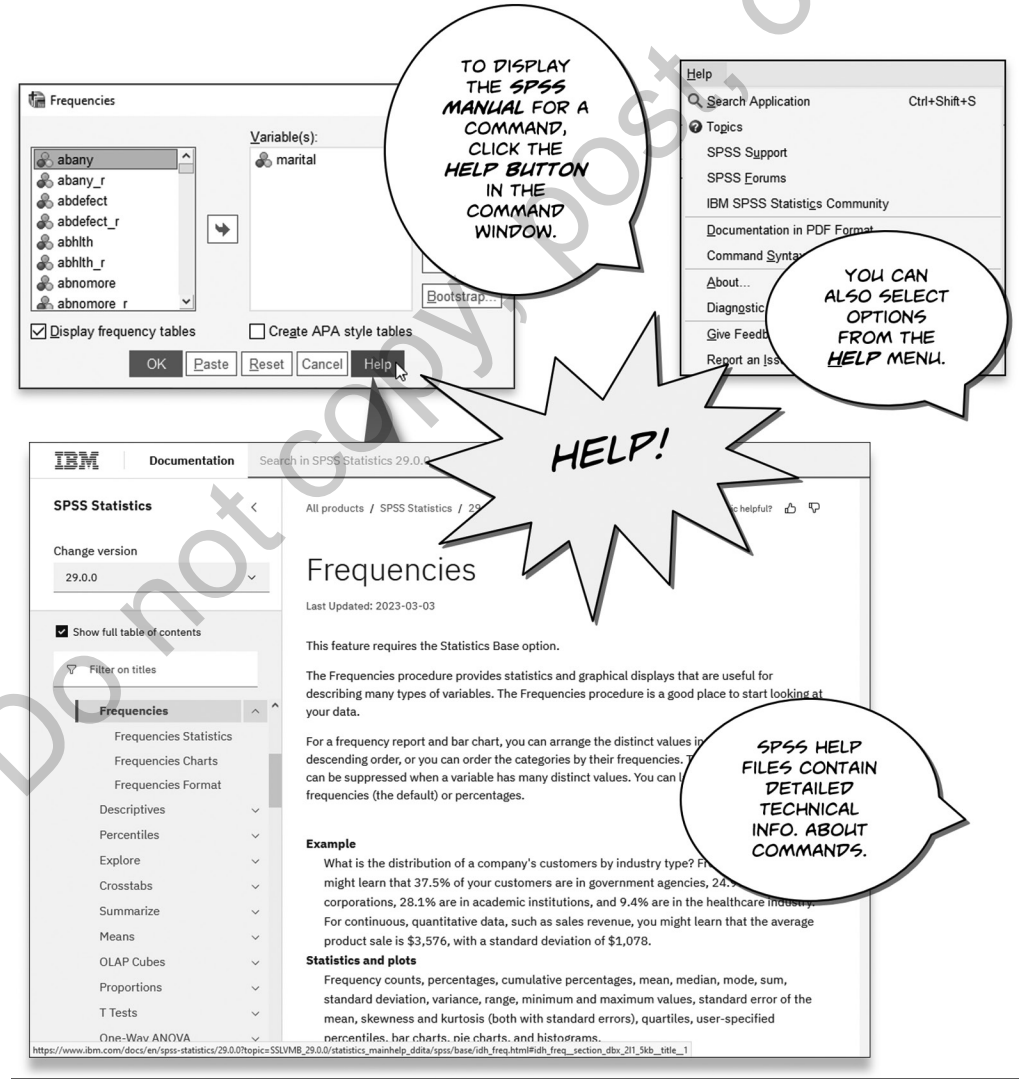
can copy and paste these text-equivalent commands into a syntax file. To create a complete syntax file, which replicates analysis from start to finish, you can also copy and paste the commands to get the GSS dataset, set viewing options, and execute the analysis. The grayed-out lines that start with * are comments (lines in the syntax file to be read by human users rather than executed by SPSS).

Another way to save your commands in a syntax file is the "Paste" button that appears in the GUI windows that execute commands. Take another look at Figure 1-6, the procedure we used to generate a frequency distribution table for the marital variable. Next to the OK button, you'll see the "Paste" button. If you press this button, SPSS will paste the syntax for the procedure at the end of your syntax file (if you don't have a syntax file open, SPSS will open a new one). For completeness, you may want to insert commands to open the dataset and some user-friendly comments to the syntax file. The Paste button is convenient feature for generating syntax files that replicate your analysis.

1.7 GETTING HELP

To view formal documentation for any SPSS procedure, you can click the "Help" button from the GUI window that executes that procedure. For example, if you want to see detailed instructions on the frequency analysis procedure you used earlier in this chapter, you could click the Help button in the Frequencies window (see Figure 1-6). SPSS retrieves the technical manual information and displays it in a web browser (Figure 1-15).

FIGURE 1-15 ■ SPSS Help Manual



You can access SPSS technical documentation through the Help menu. You can also get help by watching screencasts the authors of this book produced to show students how to execute the procedures discussed in this book. We've included links to these screencasts throughout this book. Simply point your smartphone's camera at the chapter QR codes and you'll be directed to a set of streaming videos for that chapter. You can find a complete list of screencasts on the SAGE Edge website for this book: edge.sagepub.com/pollock.

1.8 CHAPTER REVIEW

In this chapter, you have explored the essential features and functions of SPSS for data analysis. Let's review the key lessons to ensure you have met the learning objectives.

- **Using the Data Editor:** In Section 1.1, you learned how to navigate the SPSS Data Editor's Data and Variable Views. This knowledge is crucial for understanding how to enter, view, and manage datasets. Make sure you can switch between these views and understand the different types of information each provides.
- **Setting Options for** Variable Lists**:** Setting options for viewing variable lists helps you customize how variables are displayed, making it easier to manage large datasets. Do you know how to access these settings and adjust them to suit your preferences? See Section 1.2 to review.
- **Executing Commands Using SPSS's Graphical User Interface (GUI):** In Section 1.3, we used SPSS's GUI to execute a command and read results in the Viewer. Executing commands requires navigating menus, using dialog boxes, and understanding the basic interface. Practicing these skills will help you perform data analyses efficiently without needing to write syntax commands.
- **Selecting, Printing, and Saving Results:** In Section, 1.4, you learned the steps to select, print, and save results generated by SPSS. This is important for documenting your work and sharing your findings. Ensure you know how to perform these actions and can select, print, and/or save your results when needed.
- **Formatting a Table of Results:** Formatting tables is essential for presenting results clearly. Review the process of editing table appearances, covered in Section 1.5, to ensure your results are professional and easy to understand.
- **Writing and Saving Syntax Files:** In Section 1.6, you learned how to write and save syntax files, which contain SPSS command syntax. This skill is valuable for replicating analyses and automating repetitive tasks. Make sure you can create, save, and execute SPSS syntax files.
- **Using SPSS's Help System:** Lastly, in Section 1.7, you explored SPSS's help system. Knowing how to access and use this resource is helpful for troubleshooting and learning more about SPSS features. Practice using the help system to find answers to questions and expand your SPSS knowledge.

By mastering these skills, you will be well prepared to use SPSS for data analysis in your research projects. Review each section, replicate our example, practice political analysis by completing chapter exercises, and ensure you understand the key concepts and processes introduced in this chapter.

KEY TERMS

- **Academic-style tables:** the style of tables used in scholarly publications; limited embellishment, simple borders, typically black-and-white with no shading or coloring.
- **Data Editor:** an SPSS component that allows users to view, enter, edit, and manage data in a spreadsheet-like format.

- **Data View:** part of the Data Editor; individual data entries are displayed in rows and variables in columns, showing the actual data values.
- **Keyboard shortcuts:** key combinations that perform specific commands or functions in SPSS to expedite data analysis and other tasks.
- **Outline pane:** a section of SPSS display that outlines the contents or sequence of analysis, allowing users to navigate between different parts of output.
- **Metadata:** descriptive information about dataset contents, including variable names, labels, types, and other attributes.
- **Missing values:** dataset entries that are absent or undefined, which may occur due to nonresponses, data entry errors, or other reasons.
- **Syntax file:** text files with saved SPSS commands and comments; allow users to save commands and use work efficiently; have a .sps file extension.
- **Table Editor:** an SPSS tool that allows users to customize the appearance and layout of tables, including editing titles, formatting cells, and adjusting column widths.
- **Value labels:** text labels associated with each value/category of a nominal or ordinal variable.
- **Variable labels:** text that describes a variable or an attribute of a variable; may be incorporated into tables and plots.
- **Variable lists:** a section of many SPSS command windows that lists dataset variables; variables moved from list into fields of command window to execute analysis.
- **Variable names:** short labels assigned to variables in dataset to distinguish one variable from another; shorter than variables labels or descriptions.
- **Variable View:** part of the Data Editor where users can define and modify the properties of variables, such as names, types, labels, measurement levels, and formats.
- **Viewer:** a window where the results of statistical analyses, including tables, graphs, and output, are displayed after executing commands in SPSS.

CHAPTER ONE EXERCISES

1. The ANES Dataset contains a variable named terrorism_worry. The variable records ANES respondents' answers when researchers asked how worried they are about a terrorist attack in the near future.

A. Open the ANES Datatset (file name: ANES.sav), navigate to the Data Editor's Variable View, and find the terrorism_worry variable. What's the numeric code for respondents who are extremely worried about a terrorist attack in the near future?

B. Apply the Analyze ▸ Descriptive Statistics ▸ Frequencies procedure to the ANES Dataset's terrorism_worry variable. What percentage of respondents are extremely worried about a terrorist attack in the near future?

C. According to results, how many respondents have missing values on terrorism_worry because they didn't have a postelection interview, refused to answer, or had some other reason for not responding?

2. Obtain a frequency distribution table for the ANES variable leader_compromise by running the Analyze ▸ Descriptive Statistics ▸ Frequencies procedure. Copy and paste the table into a blank word processing document.[4] Edit the table for appearance and readability.[5] Submit the finished table.

3. Suppose that you have just opened the World, States, or ANES Dataset for the first time. The first thing you do is select Edit ▸ Options and consider the Variable Lists panel of the General tab. You must make sure that which two choices are selected and applied? (select two)
   - Display labels
   - Display names
   - Alphabetical
   - File
   - Measurement level
4. Create a syntax file for the commands you executed to complete Exercises 1 and 2.[6] Run your syntax file to verify it works correctly. When you run the syntax file, it should reproduce all the frequency distribution tables you created for Exercises 1 and 2. Submit the text of the syntax file.
5. SPSS's File ▸ Open command works with different types of files. These file types include datasets that record information about sample observations, output files of SPSS results, and syntax files used to replicate SPSS commands. Each of these different file types has a unique, three-character file name extension. Complete the table below to help you remember the file name extension for each file type.

SPSS File Type	File Name Extension
Dataset	
Output	
Syntax	

Additional Exercises Available

Instructors can go to edge.sagepub.com/pollockspss7e for more exercises on fillable PDF forms.

ENDNOTES

1. SPSS uses the welcome screen to promote some extensions you can add on to the program along with support resources. We won't delve into these extensions and resources, but they're worth exploring on your own.
2. If you don't see an alphabetized list of variable names in the Frequencies window, follow the DO THIS NOW instructions in the "Setting Options for Variable Lists" section above. Setting correct options for variable lists will make it easier for you to execute SPSS commands.
3. You can also access variable information within this dialog. Put the mouse pointer on the variable, marital, and right-click. Then click on Variable Information. SPSS retrieves and displays the variable's name (marital), label (Marital Status), and, most usefully, the value labels for the marital variable's numeric codes. (To see all the codes, click the drop-down arrow in the Value Labels box.)
4. See Section 1.4 for reference on selecting, copying, and saving SPSS tables.
5. See Section 1.5 for suggested edits.
6. See Section 1.6 for guidance on syntax files.

2 DESCRIPTIVE STATISTICS

LEARNING OBJECTIVES

In this chapter, you will learn how to:

- Access information about a variable in a dataset.
- Identify a variable's level of measurement.
- Describe nominal-level variables using tables and figures.
- Describe ordinal-level variables and evaluate their dispersion.
- Describe interval-level variables with descriptive statistics and figures.
- Use SPSS's Chart Editor to modify a graph.
- Obtain case-level information about observations.

Procedures Used
Analyze ▶ Descriptive Statistics ▶ Frequencies
Analyze ▶ Reports ▶ Case Summaries
Data ▶ Weight Cases
Edit ▶ Options
File ▶ Open ▶ Data

Descriptive statistics are the most basic—and sometimes the most informative—form of analysis you will do. Before you analyze why something varies, it's critical to understand how it is measured and how much it varies. In this chapter, you will learn how to use SPSS to obtain descriptive statistics for variables in datasets.

Descriptive statistics are most often used to convey two attributes of a variable: its typical value (central tendency) and its spread (degree of dispersion or variation). You will find it helpful to describe variables using specialized terminology, tables of numbers, and graphics. Most empirical research begins with a description of the variables of interest.

The precision with which we can describe central tendency for any given variable depends on the variable's **level of measurement**. The level of measurement of a variable refers to the way in which the variable's values are measured and can be used to categorize cases. For nominal-level variables, we can identify the *mode*, the most common value of the variable. For ordinal-level variables, those whose categories can be ranked, we can find the mode and the median—the value of the variable that divides the cases into two equal-size groups. For interval-level variables, we can obtain the mode, median, and arithmetic *mean*, the sum of all values divided by the number of cases.

Finding a variable's **central tendency** is ordinarily a straightforward exercise. Simply read the computer results and report the numbers. Describing a variable's degree of **dispersion** or **variation**, however, often requires informed judgment.[1]

Central tendency and variation work together in providing a complete description of any variable. Some variables have an easily identified typical value and show little dispersion. For example, suppose you were to ask a large number of U.S. citizens what sort of political system they believe to be the best: democracy, authoritarianism, or monarchy. What would be the modal response, the political system preferred by most people? Democracy. Would there be a great deal of dispersion, with large numbers of people choosing the alternatives, authoritarianism or monarchy? Probably not.

If you ask many citizens a different question, you may find that one value of a variable has a more tenuous grasp on the label "typical." And the variable may exhibit more dispersion, with the cases more evenly spread out across the variable's other values. For example, suppose a large sample of voting-age adults were asked, in the weeks preceding a presidential election, how interested they are in the campaign: very interested, somewhat interested, or not very interested. Among your own acquaintances, you probably know a number of people who fit into each category. So even if one category, such as "somewhat interested," is the median, there are likely to be many people at either extreme: "very interested" and "not very interested." This would be an instance in which the amount of dispersion in a variable—its degree of spread—is essential to understanding and describing it.

SUGGESTED READING IN *ESSENTIALS OF POLITICAL ANALYSIS*

To learn how variables are measured and described in political science, we recommend reading *The Essentials of Political Analysis*:

- 6th edition, pages 34–55 of Chapter 2
- 7th edition, all of Chapter 2

In this chapter, you will use the Analyze ▸ Descriptive Statistics ▸ Frequencies procedure to obtain appropriate measures of central tendency, and you will learn to make informed judgments about variation. With the correct prompts, the Frequencies procedure also provides valuable graphic support—bar charts and (for interval variables) histograms. These tools are essential for distilling useful information from datasets with thousands of anonymous cases, such as the American National Election Study (ANES) or the General Social Survey (GSS). For smaller datasets with aggregated units, such as the States and World Datasets, SPSS offers an additional procedure: Analyze ▸ Reports ▸ Case Summaries. The Case Summaries procedure lets you see firsthand how specific cases are distributed across a variable that you find especially interesting.

2.1 HOW SPSS STORES INFORMATION ABOUT VARIABLES

Understanding how SPSS stores information about variables helps you use SPSS effectively. To develop this understanding, consider an example: Suppose you were hired by a telephone polling firm to interview a large number of respondents. Your job is to find out and record three characteristics of each person you interview: their age, political ideology, and newspaper reading habits. The natural human tendency would be to record these attributes in words. For example, you might describe a respondent this way: "The respondent is 22 years old, is ideologically moderate, and reads the newspaper about once a week." This would be a good thumbnail description, easily interpreted by another person. To SPSS, though, these words would not make sense.

Whereas people excel at recognizing and manipulating words, SPSS excels at recognizing and manipulating numbers. This is why researchers devise a **coding system**, a set of numeric identifiers for the different values of a variable. For one of the above variables, age, a coding scheme would be

straightforward: Simply record the respondent's age in number of years, 22. To record information about political ideology and newspaper reading habits for data analysis, however, a different set of rules is needed. For example, the GSS applies the following coding schemes for political ideology (polviews) and newspaper reading habits (news):

Variable Name (GSS)	Response in Words	Numeric Code
Political ideology (polviews)	Extremely liberal	1
	Liberal	2
	Slightly liberal	3
	Moderate	4
	Slightly conservative	5
	Conservative	6
	Extremely conservative	7
Newspaper reading habits (news)	Every day	1
	A few times a week	2
	Once a week	3
	Less than once a week	4
	Never	5

Thus, the narrative profile "the respondent is 22 years old, is ideologically moderate, and reads the newspaper about once a week" becomes "22 4 3" to SPSS. SPSS doesn't really care what the numbers stand for. As long as SPSS has numeric data, it will crunch the numbers—telling you the mean age of all respondents or the modal level of newspaper reading. The variable's values are encoded as numbers, and the labels associated with those numbers tell us what the numbers mean in practical terms. It is important, therefore, to provide SPSS with labels for each code so that the software's analytic work makes sense to the user.

2.2 IDENTIFYING LEVELS OF MEASUREMENT

There are three main levels of measurement: nominal, ordinal, and interval.[2] Correctly identifying the level of measurement of a variable is important because level of measurement determines the types of statistical analyses and graphical representations that can be applied to the data. In this section, we review levels of measurement to help you identify the level of measurement of variables you encounter.

Some variables have *qualitative* values. For example, when we ask someone where they were born, their response is a place, like Kansas, Atlanta, or Mexico. Everyone was born somewhere, and a variable like birthplace simply identifies the place. Birthplace is a **nominal-level** variable. Similarly, when we ask someone their political ideology, their response is a phrase, like "ideologically moderate," that expresses the value of this varying characteristic in words. Political ideology, like birthplace, is qualitative information, but its values can be ordered, making it a variable measured at the **ordinal level**. Some people are ideologically moderate, some are extremely liberal, and others are extremely conservative. We could ask people to identify their political ideology along a spectrum that runs from extremely liberal on one side to moderate in the middle to extremely conservative on the other side.

Some variables, like someone's age in years, provide *quantitative* information. Variables measured at the **interval level** provide precise, numerical information about the units of analysis. We can describe the central tendency and dispersion of any variable, but the higher the variable's level of measurement,

the larger our toolkit for describing it. When a variable's values are meaningful numbers, we can analyze the variable's values with math in ways that are not possible when a variable's values are words.

We can describe the central tendency and dispersion of any variable, but the tools and terminology used to describe a variable depend on the variable's level of measure. To reiterate: *How you describe a variable depends on the variable's level of measurement.* The lower the level of measure, the more limited our toolkit for describing central tendency and dispersion. Nominal- and ordinal-level variables simply record qualitative information about the units of analysis. The methods available to describe **categorical variables** are relatively limited. When a variable's values quantify characteristics of the units of analysis with numbers, there are more tools available to describe the variable.

When you look at variables in the Data Editor's Variable View (see Figure 1-2), you may notice fields for "Type" and "Measure." These fields, along with variable names and labels, offer additional information about variables in a dataset, similar to what one finds in the dataset's codebook. In some cases, the variable's type and measure will accurately identify the level of measurement of the variable, but not always. This information can be helpful but should not be relied upon to identify a variable's level of measurement.

These and other points are best understood by working through some guided examples. In the next section, we'll show you how to use SPSS to describe a nominal-level variable (the lowest level of measurement).

2.3 DESCRIBING NOMINAL VARIABLES

For this and the next few analyses, you will use the World Dataset. Open the World Dataset by double-clicking the World.sav file or, if you already have SPSS running, select File ▸ Open ▸ Data and locate World.sav. Before you start analyzing this dataset, select Edit ▸ Options in the Data Editor and then click on the General tab. Just as you did when analyzing a dataset in Chapter 1, make sure that the radio buttons in the Variable Lists area are set to "Display names" and "Alphabetical." (If these options are already set, click Cancel. If they are not set, select them, click Apply, and then click OK. Now you are ready to go.)

First, you will obtain a **frequency distribution table** and **bar chart** for a nominal-level variable in the World Dataset. The region variable identifies the region of each country in the dataset. Select Analyze ▸ Descriptive Statistics ▸ Frequencies. Scroll down the left-hand side of the variables list until you find region. Click region into the Variable(s) panel. To the right of the Variable(s) panel, click the Charts button (Figure 2-1). The Frequencies: Charts Dialog appears. In Chart Type, select "Bar charts." In Chart Values, be sure to select "Percentages." Click Continue, which returns you to the main Frequencies window. Make sure "Display frequency tables" is checked in the Frequencies window. Click OK. SPSS runs the analysis.

SPSS produces two items of interest in the Viewer: a frequency distribution table of the region variable's values and a bar chart of the same information. (The small "Statistics" table isn't of much interest to us, but we include it here so you'll see what's in this book in your Viewer.) First, examine the frequency distribution table (Figure 2-2).

The value labels for each region appear in the leftmost column, with Africa occupying the top row of numbers and Western Europe occupying the bottom row. There are four numeric columns: Frequency, Percent, Valid Percent, and Cumulative Percent. Each column provides distinct information about the region variable's observed values.

- The **Frequency** column shows raw frequencies, the number of countries within each region.
- **Percent** is the percentage of *all* countries, including any missing region values, in each category of the variable. Ordinarily, the Percent column can be ignored, because we generally are not interested in including missing cases in our description of a variable. Valid Percent is the column to focus on.

FIGURE 2-1 ■ Obtaining Frequencies and a Bar Chart (for Nominal Variable)

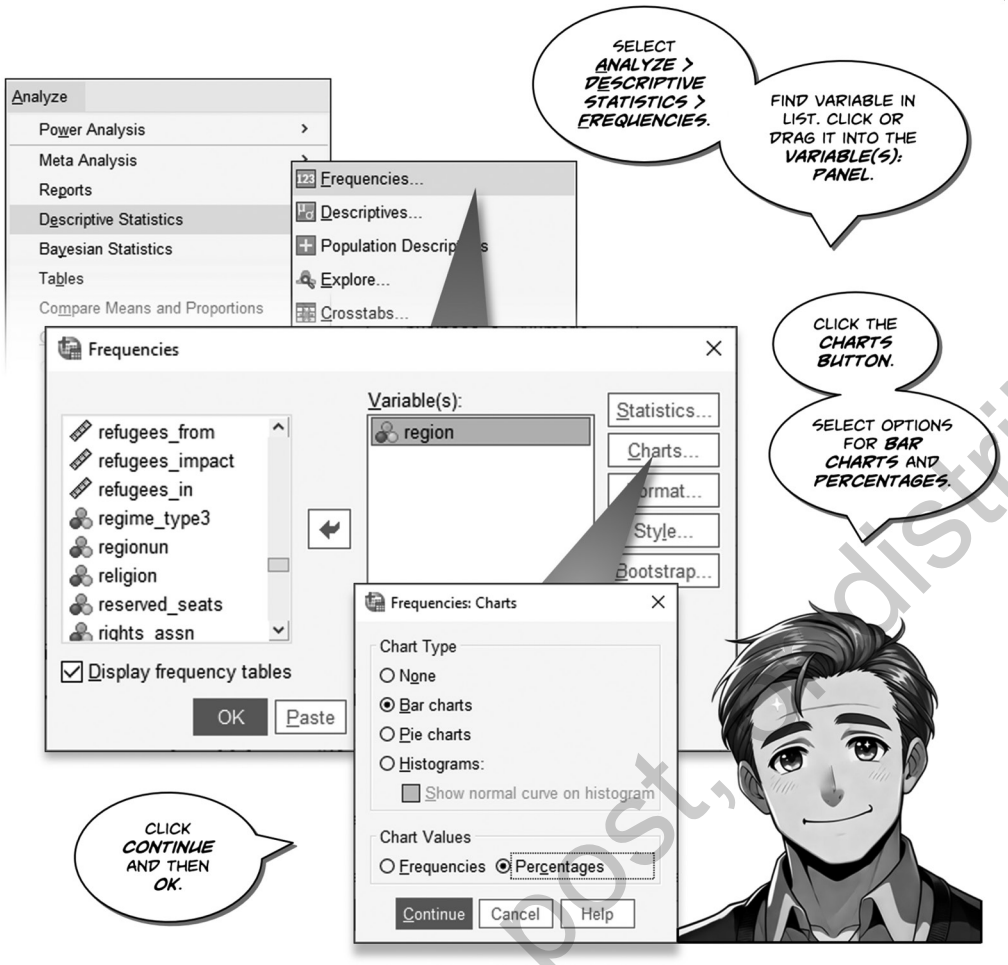


FIGURE 2-2 ■ Frequency Distribution Table for Nominal-Level Variable

Statistics

Region name

N	Valid	169
	Missing	0

Region name

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Africa	48	28.4	28.4	28.4
	Asia-Pacific	27	16.0	16.0	44.4
	C. Asia & E. Europe	28	16.6	16.6	60.9
	Middle East	19	11.2	11.2	72.2
	North America	3	1.8	1.8	74.0
	South America	24	14.2	14.2	88.2
	Scandinavia	5	3.0	3.0	91.1
	Western Europe	15	8.9	8.9	100.0
	Total	169	100.0	100.0	

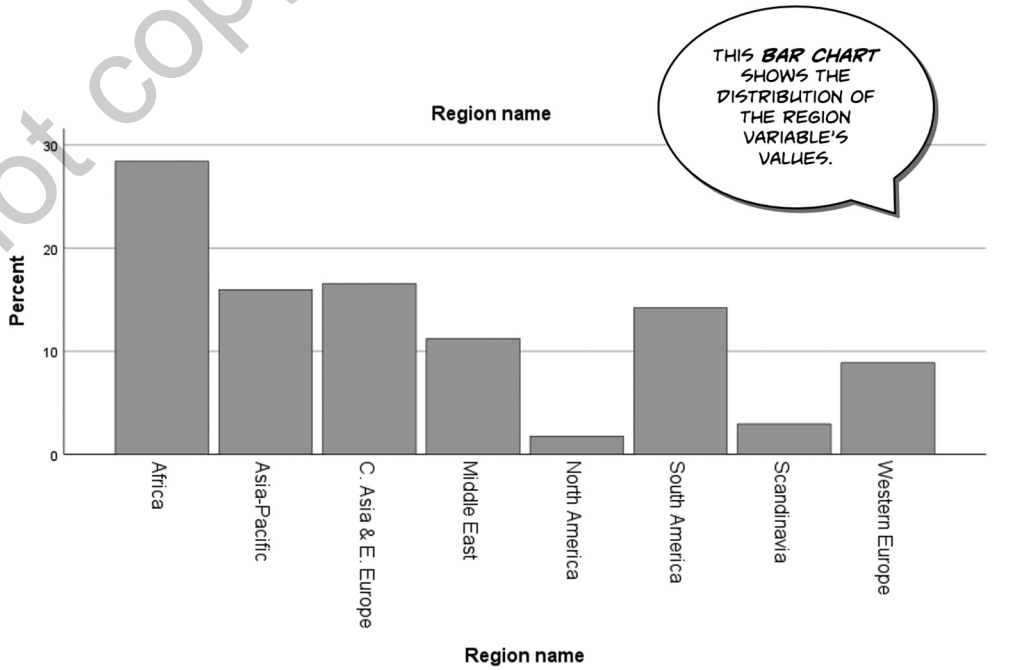
- **Valid Percent** tells us the percentage of nonmissing responses in each value of region. (The Percent and Valid Percent values happen to be the same for the region variable, but this will not always be the case.)
- The **Cumulative Percent** column reports the percentage of cases that fall in *or below* each value of the variable. For ordinal or interval variables, as you will see, the Cumulative Percent column can provide valuable clues about how a variable is distributed. But for nominal variables like region, which cannot be ranked, the Cumulative Percent column provides no information of value.

Consider the values in the Valid Percent column more closely. Scroll between the frequency distribution table and the bar chart, which depicts the region variable's values in graphic form (Figure 2-3). What is the mode, the most common value of region? For nominal variables, the answer to this question is (almost) always an easy call: Simply find the value with the highest percentage of cases. Africa is the mode. When it comes to describing the central tendency of nominal-level variables, like the World Dataset's region variable, our toolkit is limited to identifying the variable's mode.

Let's turn to describing this nominal-level variable's dispersion. Here is a general rule that applies to nominal- and ordinal-level variables: A variable has no dispersion if all the cases—states, countries, people, or whatever—fall into the same value of the variable. A variable has maximum dispersion if the cases are spread evenly across all values of the variable. The number of cases in one category equals the number of cases in every other category. A nominal- or ordinal-level variable's level of dispersion falls somewhere on a continuum between no dispersion and maximum dispersion. One describes the level of dispersion qualitatively, because there is no quantitative measure available.

No dispersion	A little dispersion	Moderate dispersion	A lot of dispersion	Maximum dispersion
All cases have same the value of variable.	Most cases fall into one category.	Most cases fall into very few categories.	Each category has roughly the same percentage of cases.	Each category has the same percentage of cases.

FIGURE 2-3 ■ Bar Chart of a Nominal-Level Variable's Values



Does the region variable have little dispersion or a lot of dispersion? Study the Valid Percent column and the bar chart. Are most of the countries concentrated in Africa, or are there many countries in each region? Africa is the modal value, but most countries are not in Africa. If countries were spread out evenly across eight regions, we would observe 1/8th of countries in each region (12.5 percent in each region), but we can see that the bars in Figure 2-3 have different heights. The variable has a lot of dispersion, but not maximum dispersion. Looking at the bar chart of region in Figure 2-3, it may be tempting to say the distribution is highest on the low/left side and decreases at higher values, as one moves from left to right, but remember that the order of nominal-level values is essentially arbitrary. SPSS defaulted to displaying values in alphabetical order, but the downward slope this generates is not a true feature of a region's dispersion.

2.4 DESCRIBING ORDINAL VARIABLES

In this section, you will analyze and describe two ordinal-level variables in the ANES Dataset, one of which has little variation and the other of which is more spread out. The American National Election Study (ANES) is an important survey of Americans' political opinions, beliefs, and attitudes that is conducted every two years. When we analyze the public opinion survey data like the ANES or GSS Datasets, it is appropriate to weight observations in the sample to better represent the population of interest. Weighting observations in the ANES Dataset helps the sample of respondents better represent the American public. To learn how to weight observations, see *A Closer Look*: Weighting Observations in the GSS and ANES Datasets.

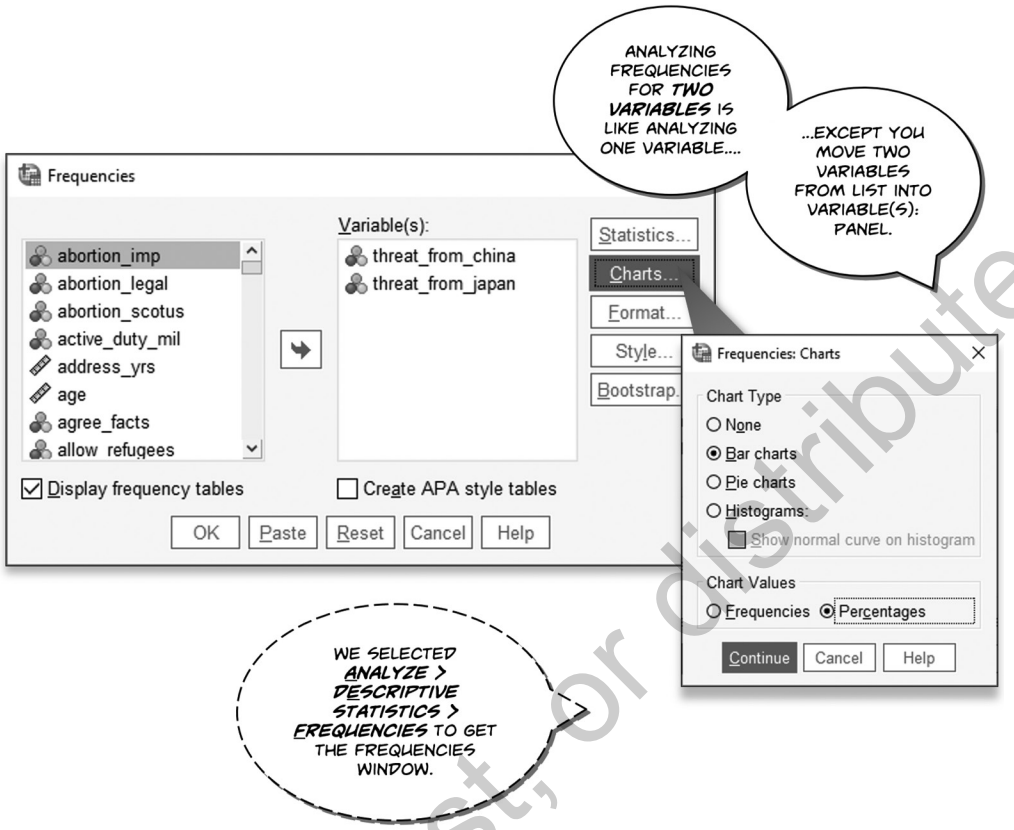
The ANES Dataset contains the variable threat_from_china, which measures the extent to which Americans think the United States is threatened by China. Similarly, the variable threat_from_japan measures how much Americans think their country is threatened by Japan. Both variables have five possible values: not at all, a little, a moderate amount, a lot, and a great deal. Both variables are measured at the ordinal level. The variables' values are qualitative descriptions of perceived threat level, not quantitative measures, but the values are ordered and identify increasing amounts of perceived threat.

We will use the same function we used to describe a nominal-level variable, so click the Analyze tab on the top menu bar of the Viewer and select Analyze ▸ Descriptive Statistics ▸ Frequencies. Scroll through the ANES Dataset variable list until you find the variable "threat_from_china" and click on it so it is added to the Variable(s) list. While you have the Frequencies Dialog open, find "threat_from_japan" in the list of variables and add it. (See Figure 2-4.) We can describe two variables at the same time almost as easily as describing one variable at a time; let's practice that skill too. SPSS should retain your earlier settings for Charts, so accompanying bar charts will appear in the Viewer.[3] Click OK.

SPSS produces descriptive statistics for the threat_from_china and threat_from_japan variables. SPSS generates two frequency distribution tables, one for each variable. To better understand how Americans think about China and Japan, we'll look at the variables' frequency distribution tables (see Figure 2-5) and bar charts (see Figures 2-6 and 2-7).

First, consider the frequency distribution table and bar chart that summarize Americans' perception of threat from China. How would you describe this variable's *central tendency*? Because threat_from_china is an ordinal variable, we can report both its mode and its **median**. Its mode is the response "5. A great deal," the option chosen by 30.6 percent of the sample. What about the variable's median value? This is where the Cumulative Percent column of the frequency distribution comes into play. *The median for any ordinal (or interval) variable is the category below which 50 percent of the cases lie.* Is the first category, "1. Not at all," the median? No, this code contains fewer than half the cases. How about "2. A little"? No, again. According to the cumulative percent column, only 17.2 percent of the cases fall in or below this response category. It is not until we notch up in rank, to "4. A lot," that the cumulative percentage exceeds the 50 percent mark. Because more than

FIGURE 2-4 ■ Generating Descriptive Statistics for Two Ordinal-Level Variables



50 percent of the cases fall in or below "4. A lot" (the cumulative percentage is equal to 69.4 percent), "4. A lot" is the median.

How would you describe the *variation* in Americans' perception of threat from China? If threat_from_china has a high level of variation, then the percentages in each category would be about equal. The modal value is a "great deal" of threat but less than half of respondents (30.6 percent) gave this response. Many respondents said the threat level was "moderate" (28.6 percent) or "a lot" (23.7 percent). Most cases fall into these three categories, but there are only five possible categories. This variable has a lot of dispersion (Figure 2-6).

Now examine the table and chart for the threat_from_japan variable. More than half of respondents (52.4 percent) fall into the modal value: "1. Not at all." In contrast, relatively few respondents said "4. A lot" (4.6 percent) or "5. A great deal" (3.0 percent). If threat_from_japan had no dispersion, then all the observations would fall into one category. That is, one value would have 100 percent of the cases, and each of the other categories would have 0 percent. This variable has a low level of dispersion; most respondents fall into one category (Figure 2-7). Based on descriptive statistics, particularly dispersion, it appears that the public is more divided—more widely dispersed—on perceptions of China than it is on perceptions of Japan.

When a variable is measured at the ordinal level, we can describe the range of observed values as well as the interquartile range (IQR) of an ordinal-level variable (more on range and IQR in Section 2.4). While it is possible to identify the range and IQR of an ordinal-level variable, they do not greatly enhance our descriptions of dispersion of ordinal-level variables.

Terms like *a little*, *a lot*, or *moderate* are loose descriptions of dispersion, but when it comes to nominal- and ordinal-level variables, our toolkit for measuring dispersion is limited. As you'll see in the next section, with a higher level of measurement, we can describe a variable's dispersion more precisely.

FIGURE 2-5 ■ Frequency Distribution Tables for Two Ordinal-Level Variables

Statistics

		POST: How much is China a threat to the United States	POST: How much is Japan a threat to the United States
N	Valid	7218	7200
	Missing	1062	1080

POST: How much is China a threat to the United States?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Not at all	408	4.9	5.7	5.7
	2. A little	834	10.1	11.6	17.2
	3. A moderate amount	2061	24.9	28.6	45.8
	4. A lot	1709	20.6	23.7	69.4
	5. A great deal	2205	26.6	30.6	100.0
	Total	7218	87.2	100.0	
Missing	System	1062	12.8		
Total		8280	100.0		

POST: How much is Japan a threat to the United States?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Not at all	3775	45.6	52.4	52.4
	2. A little	1551	18.7	21.5	74.0
	3. A moderate amount	1329	16.1	18.5	92.4
	4. A lot	332	4.0	4.6	97.0
	5. A great deal	213	2.6	3.0	100.0
	Total	7200	87.0	100.0	
Missing	System	1080	13.0		
Total		8280	100.0		

FIGURE 2-6 ■ Bar Chart of an Ordinal Variable with High Dispersion

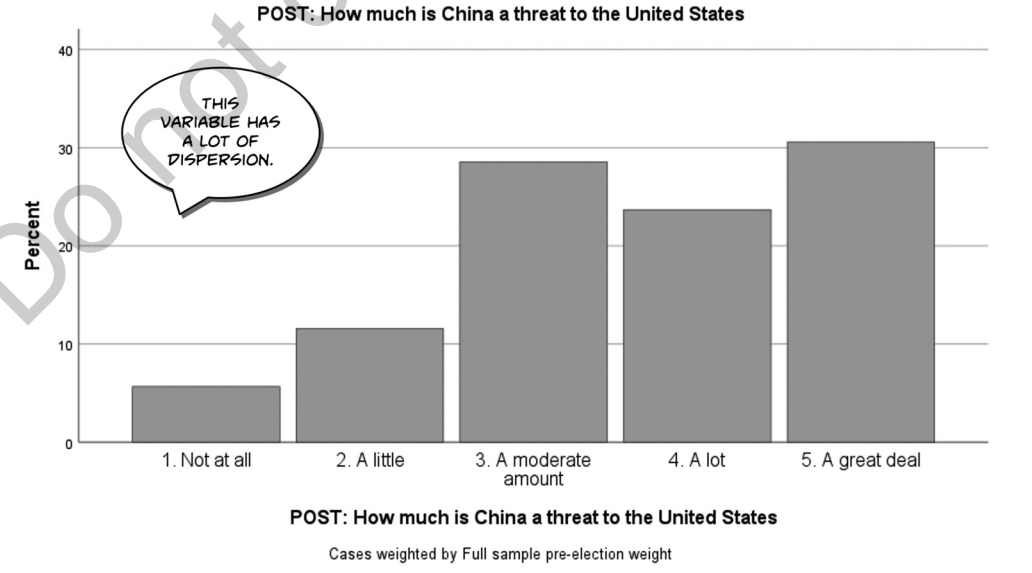
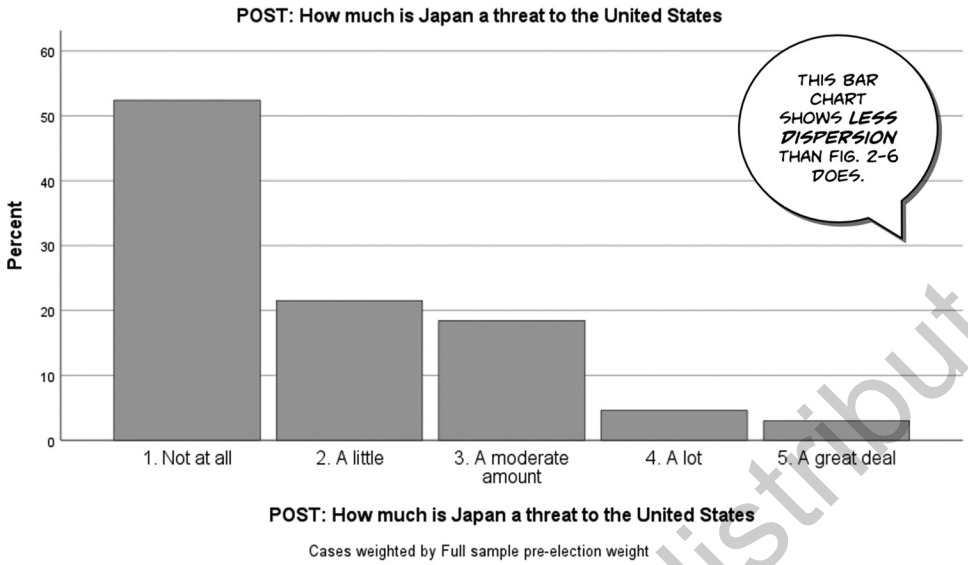


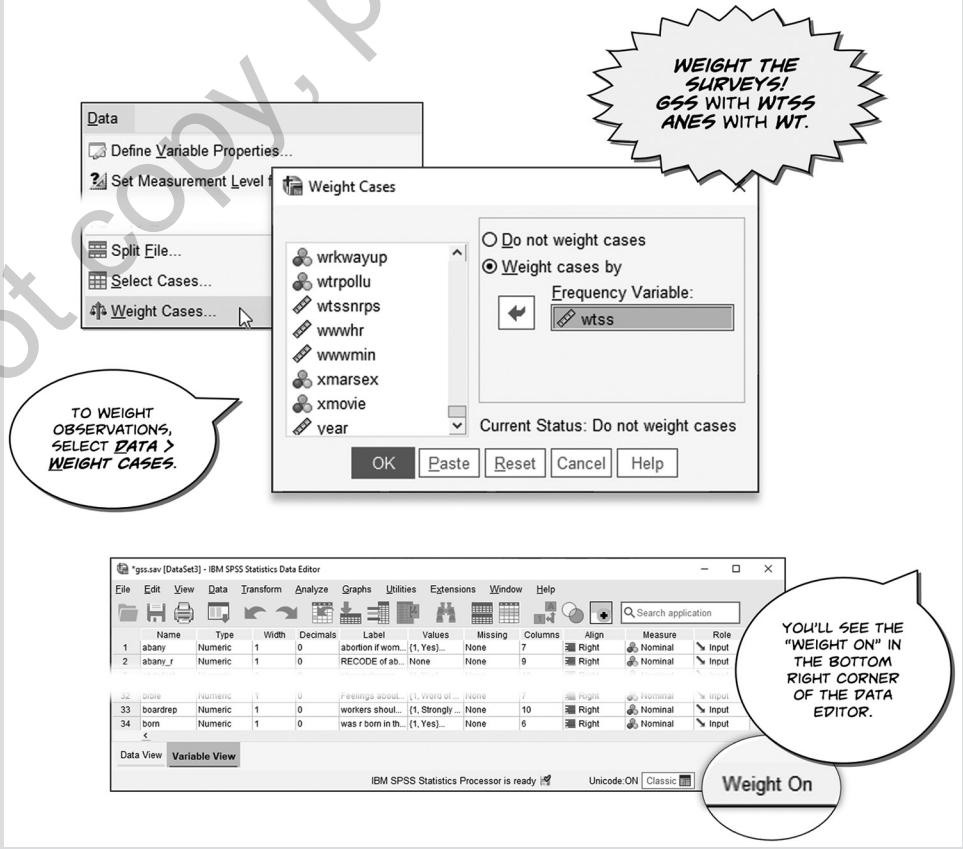
FIGURE 2-7 ■ Bar Chart of an Ordinal Variable with Low Dispersion



A CLOSER LOOK: WEIGHTING OBSERVATIONS IN THE GSS AND ANES DATASETS

Many of this book's guided examples and exercises use the two survey datasets: the General Social Survey (GSS) and the American National Election Survey (ANES). Before proceeding, you need to learn about a feature of these datasets that will require special treatment throughout the book.

FIGURE 2-8 ■ Weighting Observations in a Dataset



In raw form, the GSS and ANES Datasets are not completely representative of all groups in the population. This lack of representativeness may be intentional (e.g., the American National Election Study purposely oversampled Latino respondents so that researchers could gain insights into the attitudes of this group) or unintentional (e.g., some income groups are more likely to respond to surveys than are other groups). For some SPSS commands, this lack of representativeness does not matter. For most SPSS commands, however, analyzing raw survey data leads to incorrect results.

Fortunately, survey designers included the necessary corrective in the ANES and GSS Datasets: a weight variable. A **weight variable** adjusts for the distorting effect of sampling bias and calculates results that accurately reflect the makeup of the population. If a certain type of respondent is underrepresented in a sample, like young people in a survey conducted by dialing random landline phone numbers, that group's responses are weighted more heavily to make up for being underrepresented. In contrast, if a certain type of respondent is oversampled, that group's responses are weighted less heavily.

To obtain correct results, *you must specify the weight variable whenever you analyze the GSS or ANES Datasets.* Otherwise, your analysis will be biased. To weight observations in these datasets to produce nationally representative results, select Data ▸ Weight Cases. When analyzing the GSS, you will specify the weight variable, *wtss* (Figure 2-8). For the ANES dataset, the weight variable is *wt*. When cases are weighted, you'll see the notification "weights on" in the lower right corner of the Viewer. Conveniently, SPSS will then weight cases in subsequent analysis and not require you to set weights as an option in command windows.

2.5 DESCRIBING INTERVAL VARIABLES

Let's now turn to the descriptive analysis of interval-level variables. An interval-level variable represents the most precise level of measurement. Unlike nominal variables, whose values stand for categories, and ordinal variables, whose values can be ranked, the values of an interval variable *tell you the exact quantity of the characteristic being measured.*

Because interval variables have the most precision, they can be described more completely than can nominal or ordinal variables. We have a relatively large toolkit available for describing variables measured at the interval level. For any interval-level variable, you can report its mode, median, and arithmetic average, or *mean*. You can also make more sophisticated judgments about variation. Additionally, you can determine if an interval-level distribution is *skewed* and measure its *kurtosis*.

A step-by-step analysis of a GSS Dataset variable, age, will clarify these points. When we analyze the GSS Dataset, we should weight observations so the sample is more representative of American adults generally. To learn how to weight observations, see A Closer Look: Weighting Observations in the GSS and ANES Datasets.

2.5.1 Central Tendency

To generate descriptive statistics for the GSS Dataset's age variable, select Analyze ▸ Descriptive Statistics ▸ Frequencies. Click age into the Variable(s) list. So far, this procedure is the same as in the analysis in Sections 2.3 and 2.4.

When you are running a frequencies analysis of an *interval-level* variable, you need to adjust the settings for the Frequencies command to get proper results. Click the Statistics button in the Frequencies window, as shown in Figure 2-9. The Frequencies: Statistics window appears. In the Central Tendency panel, click the boxes next to Mean, Median, and Mode. Click Continue, returning to the main Frequencies window.

While you're in the main Frequencies window, here's something you may want to do before you execute the analysis: *Uncheck* the box next to "Display frequency tables" in the bottom left of the window. For interval-level variables, like age, that has many categories, a frequency distribution table can run several output pages and is not very informative.[4] Unchecking the "Display frequency tables" box suppresses the frequency distribution. Click OK.

FIGURE 2-9 ■ Requesting Central Tendency Statistics for an Interval Variable

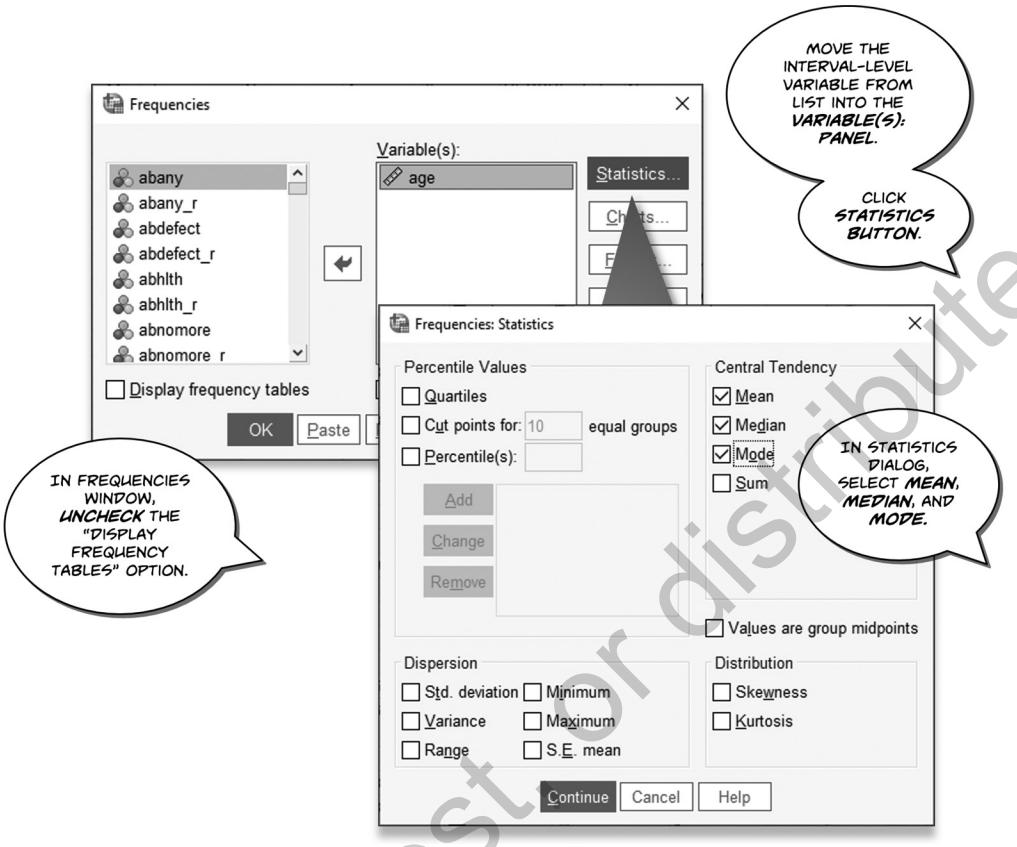


FIGURE 2-10 ■ Central Tendency Statistics for Interval-Level Variable

Statistics

age of respondent

N	Valid	3689
	Missing	343
Mean		47.96
Median		47.00
Mode		29

SPSS analyzes the age variable and outputs the requested statistics into the Viewer. Most of the entries in the Statistics table are familiar to you: valid number of cases, number of missing cases, and mean, median, and mode (Figure 2-10).

The mean age of GSS respondents, 47.96, is on display. The median age, the 50th percentile of the age variable, is 47. This variable's mean and median are similar; mean and median offer a consistent account of the age variable's central tendency. The median of the GSS Dataset's age variable is higher than median age of all Americans (38.9 years) because GSS respondents are all adults.[5]

The age variable's mode is 29. That's the most commonly observed age of GSS respondents. Is it a good measure of the variable's central tendency? When you're analyzing an interval-level variable, like the age of respondents, be skeptical about the reported mode. The values that appear to be modes are often artificial by-products of rounding the quantity of interest to the nearest integer.

Rounded to nearest year, respondents report the same age, like 29 years old, but they do not really have the age. If age is measured with greater precision, like age in years, months, and days, few respondents will report the same age. When a variable has discrete, integer values, like a count of something that occurs infrequently, the mode might provide useful information. However, if you're analyzing a continuous variable, one that can be reported to several decimal places, the mode usually doesn't help describe the variable's central tendency.

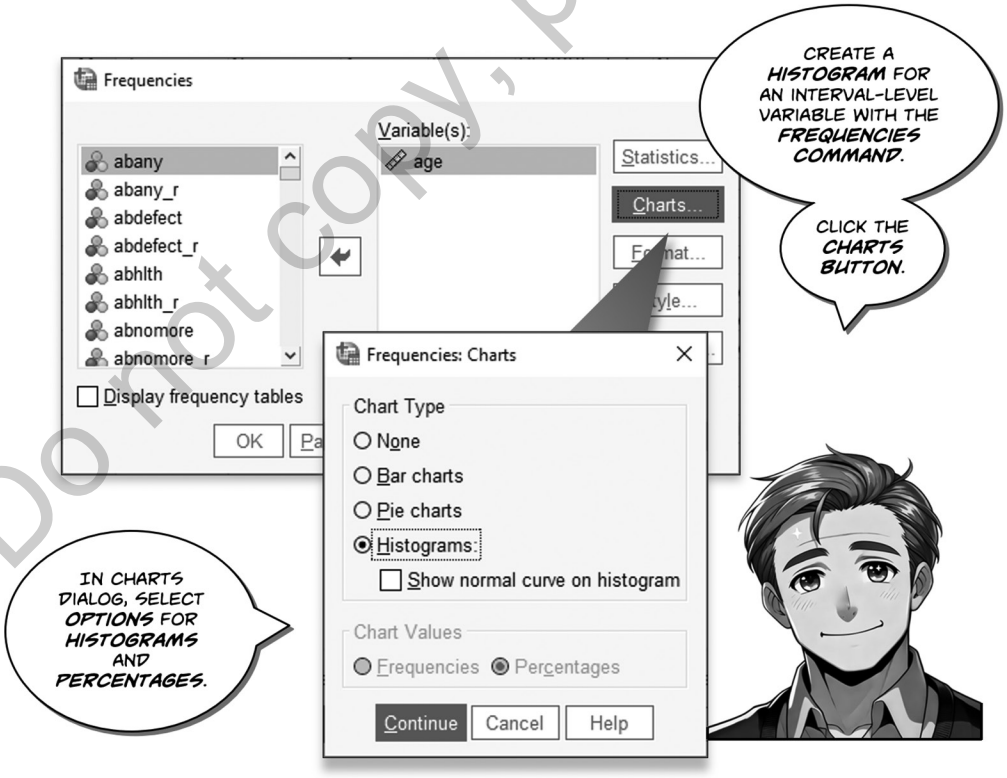
2.5.2 Creating Histograms

All the guided examples thus far have used bar charts for graphic support. For nominal and ordinal variables, a bar chart should always be your choice. For interval variables, however, you may want to ask SPSS to produce a **histogram** instead.

What is the difference between a bar chart and a histogram? A bar chart displays the percentage (alternatively, the raw number) of cases that fall into each category of a variable. A histogram is similar to a bar chart, but instead of displaying each category of a variable, it collapses categories into ranges (called bins), resulting in a compact display. Histograms are sometimes more readable and elegant than bar charts. For interval variables with many unique values, a histogram is the graphic of choice.

So that you can become familiar with histograms, run the analysis of age once again—only this time ask SPSS to produce a histogram instead of a bar chart. Select Analyze ▸ Descriptive Statistics ▸ Frequencies. Make sure age is still in the Variable(s) list. Click Statistics, and then *uncheck* all the boxes: Mean, Median, and Mode. Click Continue. Click the Charts button, and then, under the Chart Type heading, select the Histograms radio button. Click Continue. For this analysis, we do not need a frequency distribution table. In the Frequencies window, uncheck the "Display frequency tables" box. (Refer to Figure 2-11.) Click OK.

FIGURE 2-11 ■ Creating a Histogram-Style Chart for an Interval Variable



This is a bare-bones run of the Frequencies command with the option to create a histogram. SPSS reports its obligatory count of valid and missing cases, plus a histogram of the age variable's values (Figure 2-12). On the histogram's horizontal axis, notice the tick marks, which are spaced at 20-year intervals. SPSS has compressed the data so that each bar represents about 2 years of age rather than 1 year of age. Histograms smooth out the choppiness you'd see in a bar chart of an interval-level variable, while still capturing the essential qualities of a variable's distribution. Examining the histogram, you can see the central tendency of the variable, dispersion of the variable's values, and other features of the distribution. Notice, for example, how the distribution of age cuts off abruptly on the left side at age 18, while on the right side, the distribution trails off gradually.

2.5.3 Standard Deviation, Variance, Quantiles, and Range

With variables measured at the interval level, we can measure dispersion more precisely than we can with nominal- or ordinal-level variables. To obtain descriptive statistics that quantify dispersion, select Analyze ▸ Descriptive Statistics ▸ Frequencies. To replicate our example, make sure age is still in the Variable field. Click the Statistics button in the Frequencies Dialog, as shown in Figure 2-13. The Frequencies: Statistics window appears. We will request a battery of dispersion measures for demonstration purposes. In the Percentile Values panel, check **Quartiles**. In the Dispersion panel, click the boxes next to standard deviation ("Std. deviation"), variance, range, minimum, and maximum. Additionally, in the distribution section of the Statistics Dialog, check options for skewness and kurtosis. Click Continue, returning to the main Frequencies window, and then OK.

SPSS returns a table full of statistics that describe the age variable's distribution (Figure 2-14). When it comes to conducting your own political analysis or solving homework problems, you probably won't request all these statistics, but it gives us a chance to see what's readily available in SPSS. We generated descriptive statistics and a histogram for the GSS Dataset's age variable in three parts to pause and consider each task, but you can do all the analysis at once with the Frequencies command.

The two most common measures of the dispersion of interval variables are **variance** and **standard deviation**. Both variance and standard deviation are reported in the table of descriptive statistics. These statistics measure the typical amount of variation one observes in the values of a variable. The mean age of GSS respondents is 47.96 years, but we know that individual respondents are younger or older than 47.96 years.

FIGURE 2-12 ■ Histogram of the Interval-Level Age Variable

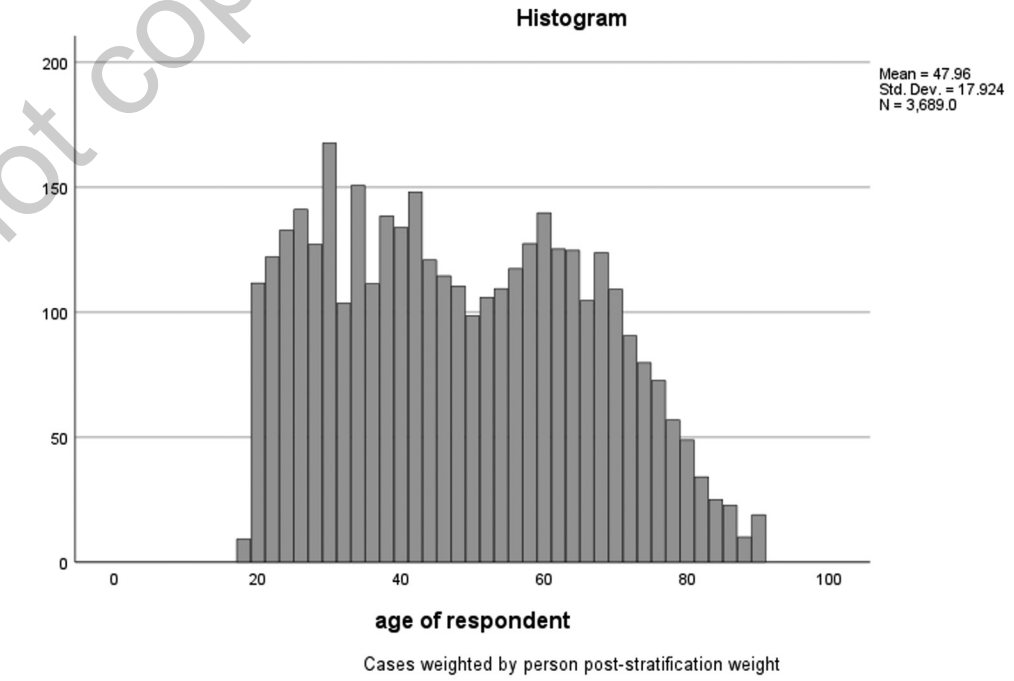


FIGURE 2-13 ■ Requesting Dispersion and Distribution Statistics for an Interval Variable

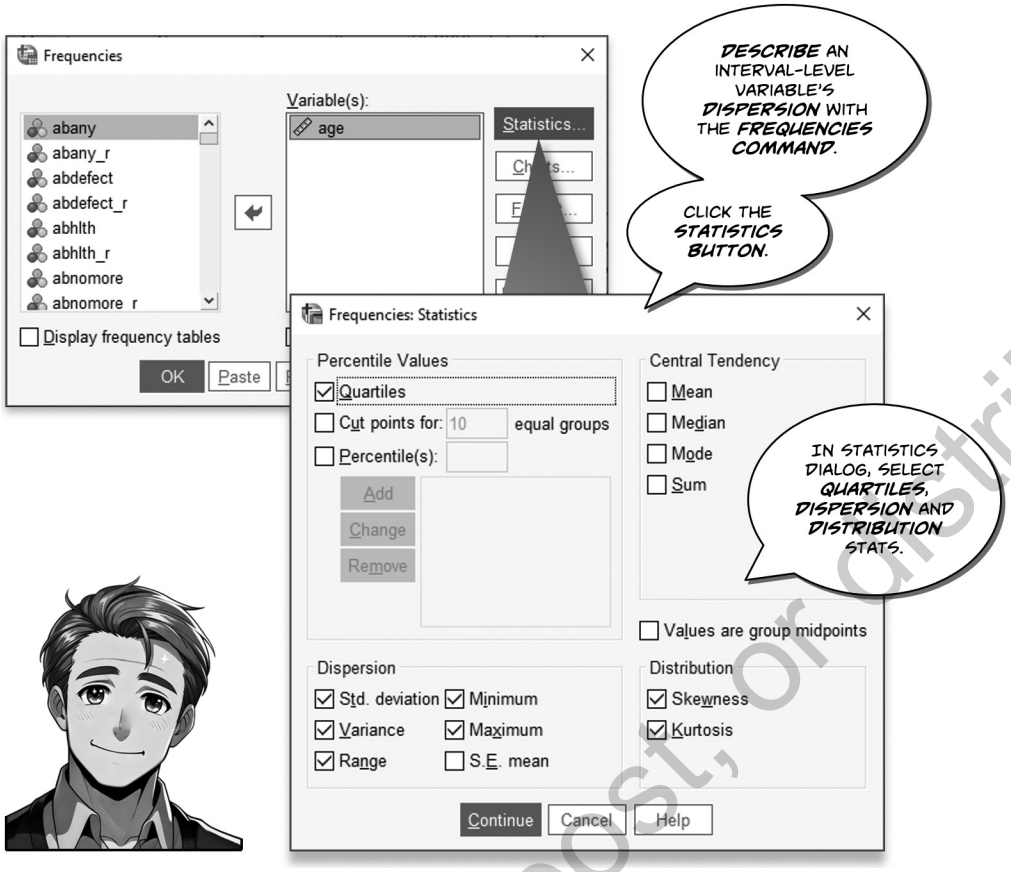


FIGURE 2-14 ■ Dispersion and Distribution Statistics for Interval Variable

Statistics

age of respondent

N	Valid	3689
	Missing	343
Std. Deviation		17.924
Variance		321.263
Skewness		.192
Std. Error of Skewness		.040
Kurtosis		–1.020
Std. Error of Kurtosis		.081
Range		71
Minimum		18
Maximum		89
Percentiles	25	33.00
	50	47.00
	75	62.52

Variance is the typical amount of squared deviation from the variable's mean. SPSS calculates variance by summing all the squared deviations and dividing that sum by the sample size minus one ($n - 1$). Variance is not exactly mean squared deviation, but it's close to it, especially with large samples. The age variable's variance is 321.26.

Standard deviation is the square root of variance; it tells you how much deviation from the mean is typically observed. Standard deviation is not exactly the mean absolute deviation; it's a little bit larger than the mean absolute deviation. Standard deviation is the most widely reported and readily understood measure of dispersion. The age variable's standard deviation is 17.92.

Another way to measure an interval-level variable's dispersion is its **range**. An interval-level variable's range is equal to the difference between its maximum and minimum observed values. The range of the GSS Dataset's age variable is 71, which is equal to the difference between its maximum value, 89, and its minimum value, 18.

Selecting the quartiles option permits us to determine the age variable's **interquartile range** (IQR). IQR is the range of the "middle half" of a variable's distribution: the spread between the top of the lowest quartile ("25%") and bottom of the highest quartile ("75%"). For the current example, we can see that the middle half of the distribution of observed age values falls between 33 and 62.52 years.[6] Thus, IQR for the age variable is 29.52, and this is another way to measure the variable's dispersion. In Chapter 5, we introduce another type of graph, a box plot, which shows a variable's IQR, among other things.

2.5.4 Skewness and Kurtosis

What does a skewness statistic tell us about the distribution of an interval variable's values? **Skewness** refers to how symmetrical a distribution is. If a distribution is not skewed, the cases tend to cluster symmetrically around the mean of the distribution. If a distribution is skewed, by contrast, one tail of the distribution is longer and skinnier than the other tail.

- When a distribution is perfectly symmetrical—no skew—it has *zero skew*. The distribution tapers off evenly on both sides.
- A distribution with a longer, skinnier right-hand tail has a *positive skew*. Many social science variables, like income and education, have positive skew because they have true zero points but no real upper limit.
- A distribution with a longer, skinnier left-hand tail has a *negative skew*. Test scores, for example, have a negative skew when the mean is close to an upper limit like 100%, with some scores falling far below the mean.

For the age variable, the skewness statistic is .192, a positive number. This suggests that the distribution has a skinnier right-hand tail—a feature that is confirmed by the shape of the histogram (see Figure 2-11). Note also that the mean (47.96 years) is slightly higher than the median (47 years), a situation that often—although not always—indicates a positive skew.[7]

Skewness affects a variable's mean value. A positive skew tends to "pull" the mean upward; a negative skew pulls it downward. However, skewness has less effect on the median. Because the median reports the middlemost value of a distribution, it is not tugged upward or downward by extreme values. *For badly skewed distributions, it is a good practice to use the median instead of the mean in describing central tendency.*

DOING YOUR OWN POLITICAL ANALYSIS

Studying descriptive statistics is a great way to start a political science research project. We showed you how to generate descriptive statistics for a few variables from our *Companion* datasets, but there are many more variables in the datasets, covering a wide range of topics. Try adapting our examples to describe variables in the datasets that you find interesting. Practice identifying levels of measurement, applying appropriate methods of analysis with SPSS, and interpreting the results you obtain.

The age variable's values are positively skewed, but .192 is not a particularly high skewness statistic; it is relatively close to zero. You have to exercise judgment, but in this case, it would not be a distortion of reality to use the mean instead of the median to describe the central tendency of the distribution.[8] Indeed, the variable's mean and median values are less than 1 year apart.

SPSS's descriptive statistics table also includes a statistic called **kurtosis**. Kurtosis measures whether the tails of a distribution are heavier (positive kurtosis values) or lighter (negative kurtosis values) than normal. However, we won't do much with kurtosis in this book.

SPSS supplements the skewness and kurtosis statistics by reporting their respective standard errors. Standard errors quantify how much a sample statistic can be expected to vary due to random sampling error; we will explore standard errors further in Chapter 8.

2.6 USING THE CHART EDITOR TO MODIFY GRAPHICS

SPSS permits the user to modify the content and appearance of any graph it produces in the Viewer using the **Chart Editor**. The user invokes the Chart Editor by double-clicking on a graph in the Viewer, edits the graphic using the Chart Editor tool, and then closes the Chart Editor to return to the Viewer. Changes made to a graphic in the Chart Editor are recorded automatically in the Viewer.

In this section, we show how you can use the Chart Editor to edit a bar chart. We demonstrate by editing a bar chart of the GSS Dataset's helpsick variable. The helpsick variable is an ordinal-level measure of public opinion about the government's responsibility for helping sick people.

To create the basic, initial chart, we follow the procedure shown in Figure 2-1. After creating the basic bar chart, we use the Chart Editor to change labels on the chart as well as the style and color of the bars. To launch the Chart Editor from the Viewer, place the cursor anywhere on the bar chart and double-click the graph (see Figure 2-15).

The Chart Editor recognizes the elements that make up the bar chart. It recognizes some elements as text. These elements include the axis titles and the value labels for the categories of helpsick. It recognizes other parts of the bar chart as editable elements, such as the bars in the bar chart. First, we'll edit a text element, the title on the vertical axis. Then we will modify a graphic element, the color of the bars.

Place the cursor anywhere on the title, which is "Should Govt Help Pay For Medical Care?" in this example, and single-click it. SPSS selects the chart title. With the cursor still placed on the title, single-click again. SPSS moves the text into editing mode inside the chart (see the left side of Figure 2-16). The default title is serviceable, but we can improve it. Edit the title so it reads "Should the Government Help People Pay for Medical Care?" We want the bar chart to communicate its information as clearly as possible.

You can also use the Text Style menu in the **Properties** window that pops up automatically when you double-click a graphic element in the Chart Editor (see the right side of Figure 2-16) to change the font style of the main title and *x*-axis labels from the plain, default sans serif font to something more stylish. If you're going to use an SPSS graphic in a paper or presentation, you may want to match the font used in the graphic with the font used in the paper or presentation. When you change the font of a text element, change the font of all the other text elements to match so your graphic doesn't become a hodgepodge of fonts. Apply your changes to the chart text.

Now click on one of the vertical bars. The editor selects all the bars. As before, you can double-click an element in the Chart Editor to summon the associated Properties Dialog. Alternatively, you can select the element and press the "Show Properties Window" button located near the upper-left corner of the Chart Editor (as shown in Figure 2-17). This opens the Properties window, the most powerful editing tool in the Chart Editor's arsenal.

FIGURE 2-15 ■ Launching the Chart Editor

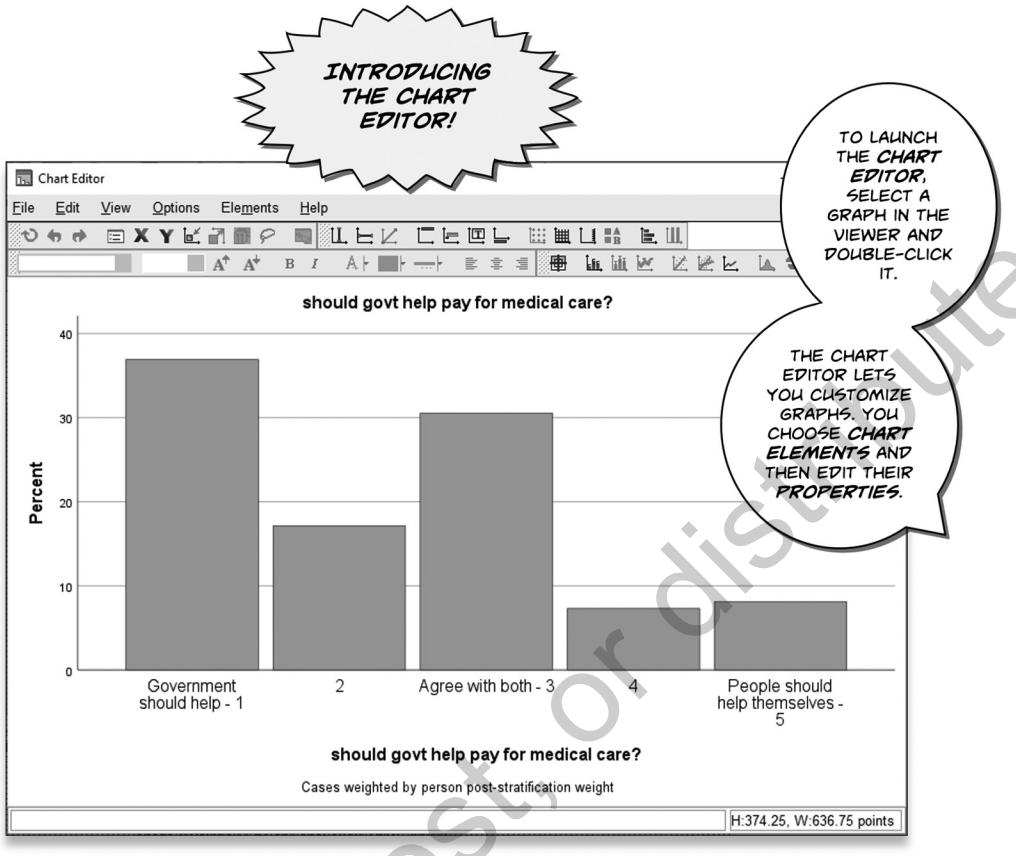


FIGURE 2-16 ■ Editing Bar Chart Title with Chart Editor

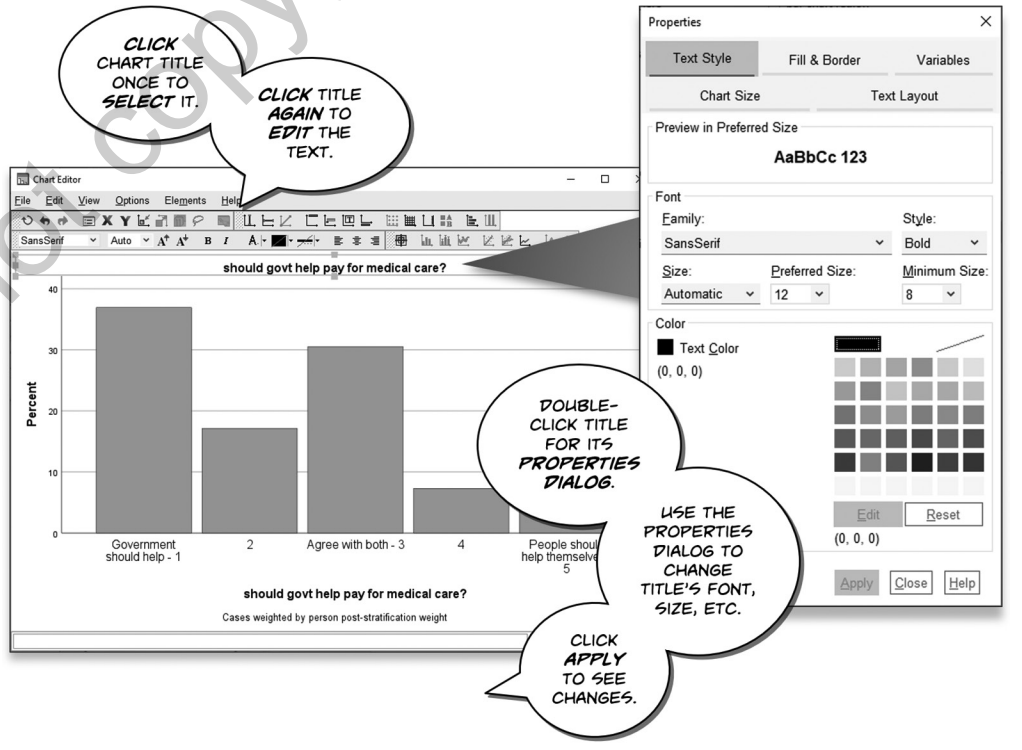
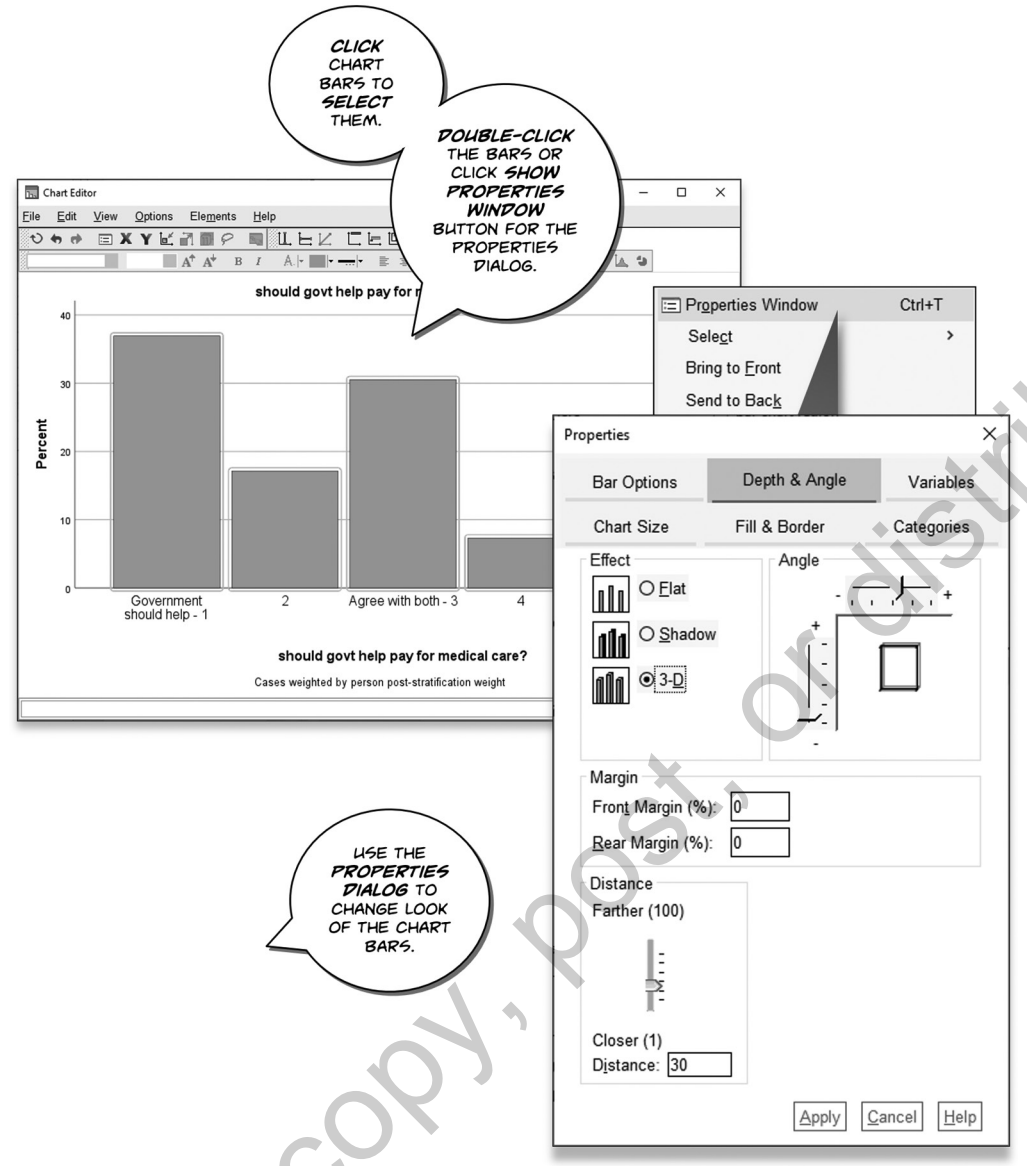


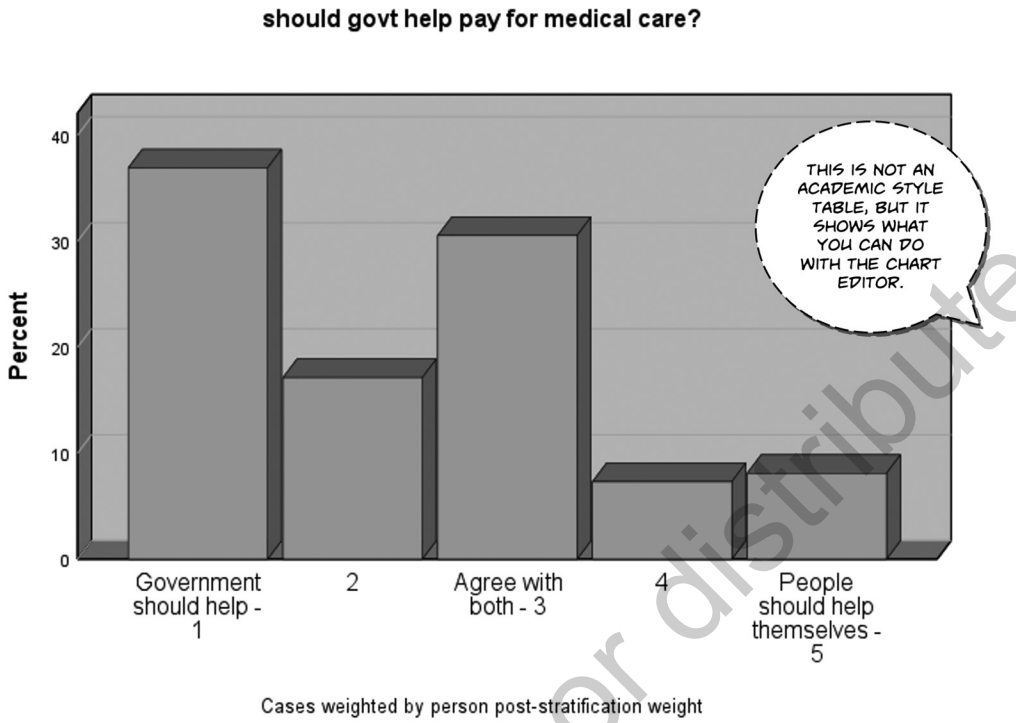
FIGURE 2-17 ■ Using the Properties Window to Edit Bars in a Bar Chart



If you plan to do a lot of editing, it is a good idea to open the Properties window soon after you enter the Chart Editor. Each time you select a different text or graphic element with the mouse, the Properties window changes, displaying the editable properties of the selected element.

The options for editing graphical elements, like the bars of a bar chart, are plentiful. You can change bars' color, adjust their order, and make them bigger or smaller. The "Depth & Angle" tab of the bar properties provides an option that dramatically transforms the humble bar chart into a visually interesting graphic: a 3-D effect. Select the 3-D option (see Figure 2-17) and apply it to the bar chart. You'll see the difference this option makes in the Chart Editor. If you close the Chart Editor, the finished product appears in the Viewer (Figure 2-18).

FIGURE 2-18 ■ Edited Bar Chart in the Viewer



2.7 OBTAINING CASE-LEVEL INFORMATION WITH CASE SUMMARIES

When you analyze a large survey dataset, like the GSS or ANES Datasets, you generally are not interested in how Respondent 42 or Respondent 155 answered a particular question; they're anonymous people randomly picked to participate in the survey. Rather, you want to know how the entire sample of respondents answered a question (for this purpose, their randomness is vitally important). Sometimes, however, you gather data on particular cases because the cases are themselves noteworthy and interesting to discuss.

When you work with the States Dataset (States.sav) and the World Dataset (World.sav), you may want to describe cases beyond the relative anonymity of Frequencies analysis and find out where particular states or countries "are" on an interesting variable. To obtain case-level information, select Analyze ▸ Reports ▸ Case Summaries. This SPSS procedure is readymade for such elemental insights.

Suppose after examining the how many countries are in each region of the world (the descriptive analysis from Section 2.3), you are interested in identifying countries that have the highest and lowest population densities. Perhaps there is some regional variation in the density of countries around the world. To begin this guided example, open the World Dataset. With the dataset open, click Analyze ▸ Reports ▸ Case Summaries.

The World Dataset contains a variable named population_density. This variable records the number of people per square kilometer. Which countries have the most inhabitants per square kilometer? Which countries have the fewest? Where does your country fall on the list? Case Summaries can quickly answer questions like these. SPSS will sort observations based on a "grouping variable" (in this example, population_density) and then produce a report telling you which countries are in each group.

To conduct the desired analysis, you need to do three things in the Summarize Cases window (see Figure 2-19):

1. Click the variable containing case names into the Variables window. In the World Dataset, this variable is named country, which is simply the name of each country.

2. Click the variable you are interested in analyzing, population_density, into the Grouping Variable(s) window. The grouping variable is the criteria by which observations are sorted.
3. Uncheck the "Limit cases to first . . ." option. If this box is left checked when you analyze the World Dataset, SPSS will limit the analysis to the first 100 countries in the dataset and produce an incomplete analysis.[9]

After completing the three steps illustrated in Figure 2-19, click OK and consider the output. SPSS sorts the cases on the grouping variable, population_density in this example, and tells us which country is associated with each value of population_density. For example, Mongolia, with 1.98 inhabitants per square kilometer, has the lowest population density in world. Which country has the highest population density? Scroll to the bottom of the tabular output (Figure 2-20). With 7,915.73 inhabitants per square mile, Singapore has the highest population density in the world.

FIGURE 2-19 ■ Obtaining Case Summaries

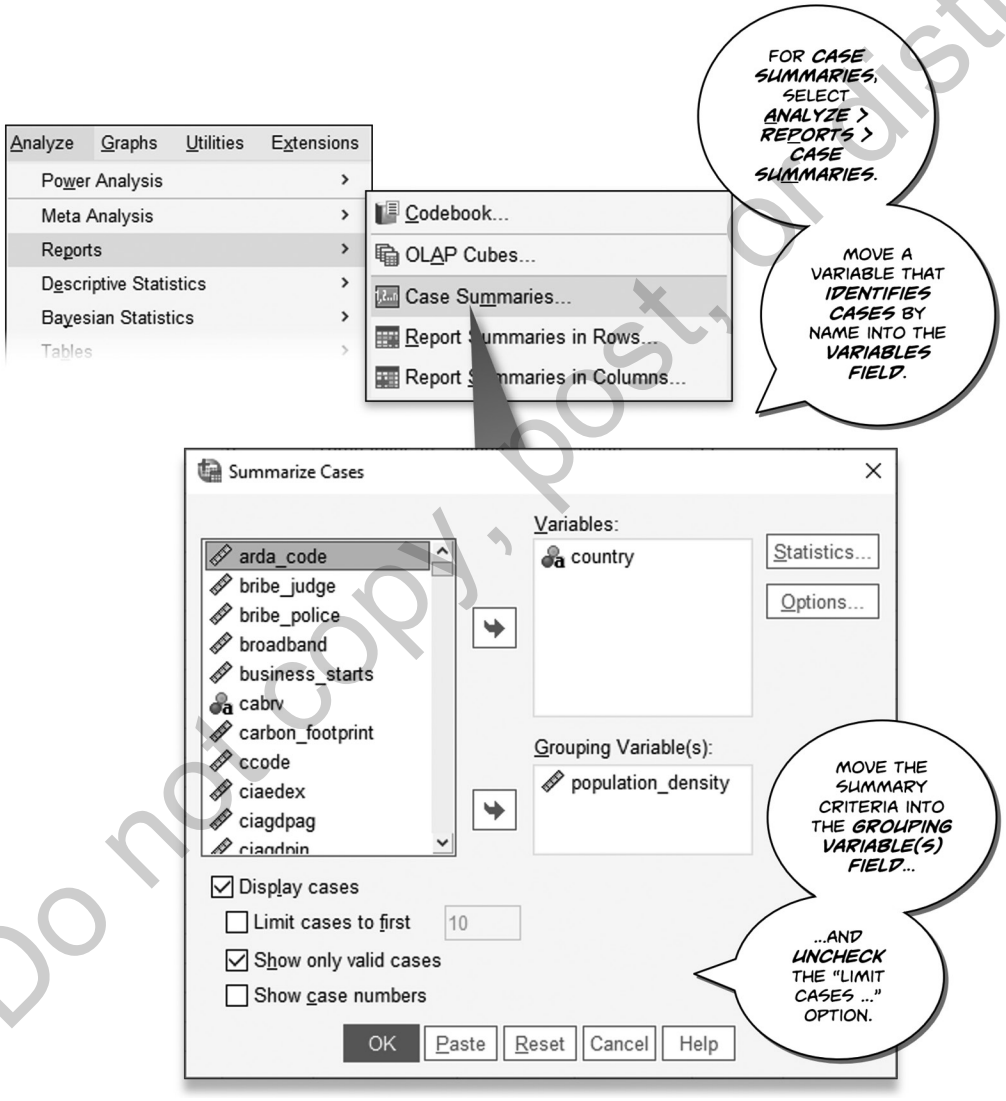


FIGURE 2-20 ■ Case Summary Report about Population Density by Country

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Country/territory name * Number of people per square kilometer	164	97.0%	5	3.0%	169	100.0%

Case Summaries

				Country/territory name
Number of people per square kilometer	1.98	1		Mongolia
		Total	N	1
	3.08	1		Namibia
		Total	N	1
	3.20	1		Australia
		Total	N	1
	3.40	1		Iceland
		Total	N	1
	3.61	1		Suriname
		Total	N	1
	3.62	1		Libya
		Total	N	1
	3.95	1		Guyana
		Total	N	1
	4.04	1		Canada
		Total	N	1

Note: Table rows intentionally omitted

	1265.04	1		Bangladesh
		Total	N	1
	1454.04	1		Malta
		Total	N	1
	1454.43	1		Maldives
		Total	N	1
	1935.91	1		Bahrain
		Total	N	1
	7915.73	1		Singapore
		Total	N	1
	Total	N		164

2.8 CHAPTER REVIEW

In this chapter, we explored the key concepts and techniques for describing and summarizing different types of variables. Let's review the essential lessons to ensure you have met the learning objectives.

- **Access Information about a Variable in a Dataset:** In Section 2.1, you learned how SPSS stores information about variables. This knowledge helps you understand the contents of a dataset and locate relevant information. Make sure you can access and interpret information about any variable in a dataset.

- **Identify a Variable's Level of Measurement:** Section 2.2 focused on identifying levels of measurement, which is crucial for selecting appropriate statistical methods. Ensure you can differentiate between nominal-, ordinal-, and interval-level variables.
- **Describe Nominal-Level Variables:** In Section 2.3, you learned how to describe nominal-level variables using tables and figures. Practice creating frequency tables and bar charts to effectively summarize categorical data.
- **Describe Ordinal-Level Variables:** Section 2.4 covered the description of ordinal-level variables and methods to evaluate their dispersion. Ensure you can create appropriate tables and charts and calculate measures of dispersion like the range and IQR.
- **Describe Interval-Level Variables:** In Section 2.5, you learned how to describe interval-level variables using various descriptive statistics and figures. Review how to calculate central tendency measures (mean, median, mode), create histograms, and assess standard deviation, variance, quantiles, range, skewness, and kurtosis.
- **Use SPSS's Chart Editor to Modify a Graph:** Section 2.6 introduced you to SPSS's Chart Editor, a powerful tool for customizing graphs. Practice modifying graphs to ensure they effectively describe variable values.
- **Obtain Case-Level Information about Observations:** Lastly, in Section 2.7, you learned how to obtain case-level information with case summaries. This skill is valuable for examining individual observations within your dataset.

By mastering these skills, you will be well prepared to use descriptive statistics in your research projects. Review each section, replicate the examples provided, and practice the exercises to ensure you understand the key concepts and processes introduced in this chapter. This solid foundation in descriptive statistics will enhance your ability to analyze and interpret data effectively.

KEY TERMS

- **Bar chart:** a visual depiction of the relative distribution of a nominal- or ordinal-level variable's values.
- **Categorical variables:** nominal- and ordinal-level variables; values define categories.
- **Central tendency:** measures that describe the center of a data distribution, such as mean, median, or mode.
- **Chart Editor:** tool within SPSS that allows users to modify and customize the appearance of graphs and charts.
- **Coding system:** a set of rules or guidelines used to categorize and assign numeric values to categories for analysis purposes.
- **Cumulative percent:** a running tally of percentages; in a frequency distribution table, shows the percentage of observations up to and including that row.
- **Descriptive statistics:** methods that summarize and describe the main features of variables, including measures of central tendency and dispersion.
- **Dispersion:** the extent to which data values in a dataset are spread out or clustered together.
- **Frequency:** count of observations in a category or interval, like the number of countries with a particular value of a variable.
- **Frequency distribution table:** shows the relative frequency of a nominal- or ordinal-level variable's values in percentages or proportions; may also show counts.

- **Histogram:** chart that shows the distribution of an interval-level variable's values. Each vertical bar represents a binned range of values.
- **Interquartile range:** difference between first and third quartile values (the 25 percent and 75 percent values); the "middle half" of a distribution.
- **Interval level:** quantitative measurements, like length of time in hours or amount of money in dollars.
- **Kurtosis:** statistical measure of a distribution's peakedness. If greater than 3, it's more peaked than a bell curve; if less than 3, it's less peaked than a bell curve.
- **Level of measurement:** precision with which a variable is measured; some values are qualitative (nominal or ordinal level), while others are quantitative (interval level).
- **Median:** the 50 percent value of a variable; used to describe ordinal and interval variables.
- **Nominal level:** categories with no intrinsic order, like denominations of a religion, regions of the world, and nationalities of people.
- **Ordinal level:** categories that can be ranked or ordered, like low, medium, and high.
- **Percent:** in SPSS frequency tables, percent of all cases, including those with missing values.
- **Properties:** attributes like size, color, shape, and position that define the appearance and structure of graph components.
- **Quartiles:** values that divide ordered variable values into four equal parts; the 25, 50, and 75 percent values.
- **Range:** difference between a variable's minimum and maximum values.
- **Skewness:** statistical measure of a distribution's symmetry. If positive, the mean is greater than the median and the distribution has a longer right tail; if negative, the mean is less than the median and the left tail is longer.
- **Standard deviation:** widely used measure of a variable's dispersion; equal to the square root of variance.
- **Valid percent:** percentage in category excluding any missing or invalid cases.
- **Variance:** measure of a variable's dispersion; the sum of squared deviations from the mean divided by $n - 1$ using the following formula: $\sum(\boldsymbol{x_i} - \bar{\boldsymbol{x}}) / (\boldsymbol{n} - 1)$.
- **Variation:** the degree to which variable values differ from each other and the mean; synonymous with dispersion.
- **Weight variable:** weights used to correct for under- or overrepresentation of cases in a sample relative to a population; how much a case should count in analysis.

CHAPTER TWO EXERCISES

1. How you analyze a variable depends on its level of measurement. To apply the right methods, you must be able to determine if a variable is measured at the nominal, ordinal, or interval level.[10]
   - **A.** The States Dataset includes a variable named min_wage, which reports the minimum wage in each state in dollars and cents. What's the level of measurement of the min_wage variable?
   - **B.** The World Dataset includes a variable named frac_eth3, which records the level of ethnic fractionalization in countries as low, medium, or high.
2. Practice identifying the level of measurement of variables by completing the following table.[11]

Dataset	Variable	Level of Measurement (select one)	How Do You Know?
States	voter_id_law	☐ Nominal ☐ Ordinal ☐ Interval	
States	opioid_rx_rate	☐ Nominal ☐ Ordinal ☐ Interval	
World	gender_equal3	☐ Nominal ☐ Ordinal ☐ Interval	
World	gender_inequality	☐ Nominal ☐ Ordinal ☐ Interval	
World	typerel	☐ Nominal ☐ Ordinal ☐ Interval	

3. You can create frequency distribution tables for variables measured at the nominal level as well as variables measured at the ordinal level. The tables look similar, but there are differences. You can add a column of cumulative percentages when you're working with an ordinal-level variable. You can't add a column of cumulative percentages to the frequency distribution table of a nominal-level variable.[12] Why is that?

4. Both bar charts and histograms are used to visually display the dispersion of a variable's values. Bar charts and histograms sometimes look very similar, but there are important differences between them.[13] How are histograms different from bar charts? Why would you use a histogram to display the dispersion of an interval-level variable instead of a bar chart?

5. According to the Inter-Parliamentary Union, an international organization of parliaments, 23.7 percent of members of the U.S. House of Representatives are women.[14] How does the United States compare to other democratic countries? Is 23.7 percent comparatively low, comparatively high, or typical for a national legislature? The World Dataset contains a variable named womenleg, which records the percentage of women in the lower house of the legislature in each of 168 democracies.

A. Use SPPS to obtain descriptive statistics for womenleg.[15] Use your results to fill in the blanks: The mean of womenleg is equal to _____. The median of womenleg is equal to _____. The minimum is equal to _____ and the maximum is equal to _____.

B. Analysts generally prefer to use the mean to summarize a variable's central tendency, except in cases where the mean gives a misleading indication of the true center of the distribution. Make a considered judgment. For womenleg, can the mean be used or should the median be used instead? Explain your answer.

C. Recall that 23.7 percent of U.S. House members are women. Suppose a women's advocacy organization vows to support female congressional candidates so that the U.S. House might someday "be ranked among the top one-fourth of democracies in the percentage of female members." According to your analysis, women would need to constitute _____ percent of the House to meet this goal.

D. Run Analyze ▸ Reports ▸ Case Summaries. Click the country variable into the Variables box and womenleg into the Grouping Variable(s) box. Make sure to uncheck the box next to "Limit cases to first 100." Examine the output.[16]

Which five countries have the highest percentages of women in their legislatures?

1. _____

2. _____

3. _____

4. _____

5. _____

Which five countries have the lowest percentages of women in their legislatures?

1. _____

2. _____

3. _____

4. _____

5. _____

E. Create a nicely labeled histogram of the womenleg variable.[17] Give the horizontal axis (*x*-axis) the following label: "Percentage of Women in Legislature." Give the chart this main title: "Percentage Women Legislators in 168 Democracies." Submit the histogram.

Additional Exercises Available

Instructors can go to edge.sagepub.com/pollockspss7e for more exercises on fillable PDF forms.

ENDNOTES

1. In this chapter, we use the terms *dispersion*, *variation*, and *spread* interchangeably.
2. Some textbooks include a fourth level of measurement: ratio-level measurement. Ratio variables are a special subset of interval-level variables that have a true zero point. Income, measured in dollars, is a ratio-level variable because there is a true zero point: Zero dollars of income represents the complete absence of income. Temperature, measured in degrees Celsius or Fahrenheit, is an interval-level variable but not a ratio-level variable. Zero degrees Celsius or Fahrenheit does not represent the complete absence of temperature; these measures don't have a true zero point and, in fact, have different zero points.
3. If you pressed the Reset button to clear a different variable from the Variable(s) list, you'll need to click the Charts button again to have SPSS produce a bar chart with values specified as percentages.
4. A general guide: If the interval-level variable you are analyzing has 15 or more distinct values, go ahead and obtain the frequency distribution. If it has more than 15 distinct values, suppress the frequency distribution. Of course, you may not know how many distinct values a variable has until you generate a frequency distribution table to bar chat, so it may be necessary to do a trial run of the Frequencies command before deciding on appropriate options.
5. See U.S. Census Bureau, "America Is Getting Older," Press Release Number CB23-106, June 22, 2023, available at https://www.census.gov/newsroom/press-releases/2023/population-estimates-characteristics.html. Properly weighted, the GSS is representative of American adults but not the population generally because it excludes minors.
6. 62.52 is a curious 75th percentile value because the GSS Dataset's age variable is recorded in whole numbers. No one is observed to be 62.52 years old; how can this be a quartile value? Quartile values are determined by sorting observed values from smallest to largest and identifying the 25th, 50th, and 75th percentile values. When the quartile falls between two values, those values are averaged to obtain the quartile value. If, for example, a dataset has an even number of observations, the 50th percentile falls between two observations, and those observed values are averaged to determine the median; if the two midmost values are different integers, the median may have a 5 after the decimal point. Similarly, if the 25th or 75th percentiles fall between two observed values, the surrounding values are averaged to obtain quartile values. For GSS ages, the 75th percentile falls between 62 and 63 years old; SPSS averages 62 and 63 to find the 75th percentile value, but in doing so, it weights observations, resulting in the 62.52 quantile value.

7. Paul T. von Hippel, "Mean, Median, and Skew: Correcting a Textbook Rule," *Journal of Statistics Education* 13, no. 2 (2005). "Many textbooks teach a rule of thumb stating that the mean is right of the median under right skew, and left of the median under left skew. This rule fails with surprising frequency."
8. For demographic variables that are skewed, median values rather than means are often used to give a clearer picture of central tendency. One hears or reads reports, for example, of median family income or the median price of homes in an area.
9. The "limit cases to first ..." option does not do what you might think it does. If you limit cases to the first 10, SPSS will not display the 10 most/least dense countries in the world. Instead, it will analyze the first 10 rows of the dataset, which are the first 10 countries in alphabetic order, and report those 10 countries by ascending population density.
10. Section 2.1 tells you how to identify a variable's level of measurement.
11. Subsequent chapters build on your ability to identify levels of measurement, so you must master this skill.
12. To answer this question correctly, you need to understand the difference between nominal- and ordinal-level variables (covered in Section 2.1) and apply that understanding to table construction.
13. We discuss histograms as an alternative to bar charts in Section 2.6.
14. See the Inter-Parliamentary Union website at https://www.ipu.org.
15. See Section 2.5 for step-by-step instructions for generating descriptive statistics for interval-level variables with SPSS.
16. See Section 2.7 for guidance on obtaining case-level information.
17. Section 2.5.2 covers creating histograms for interval-level variables.