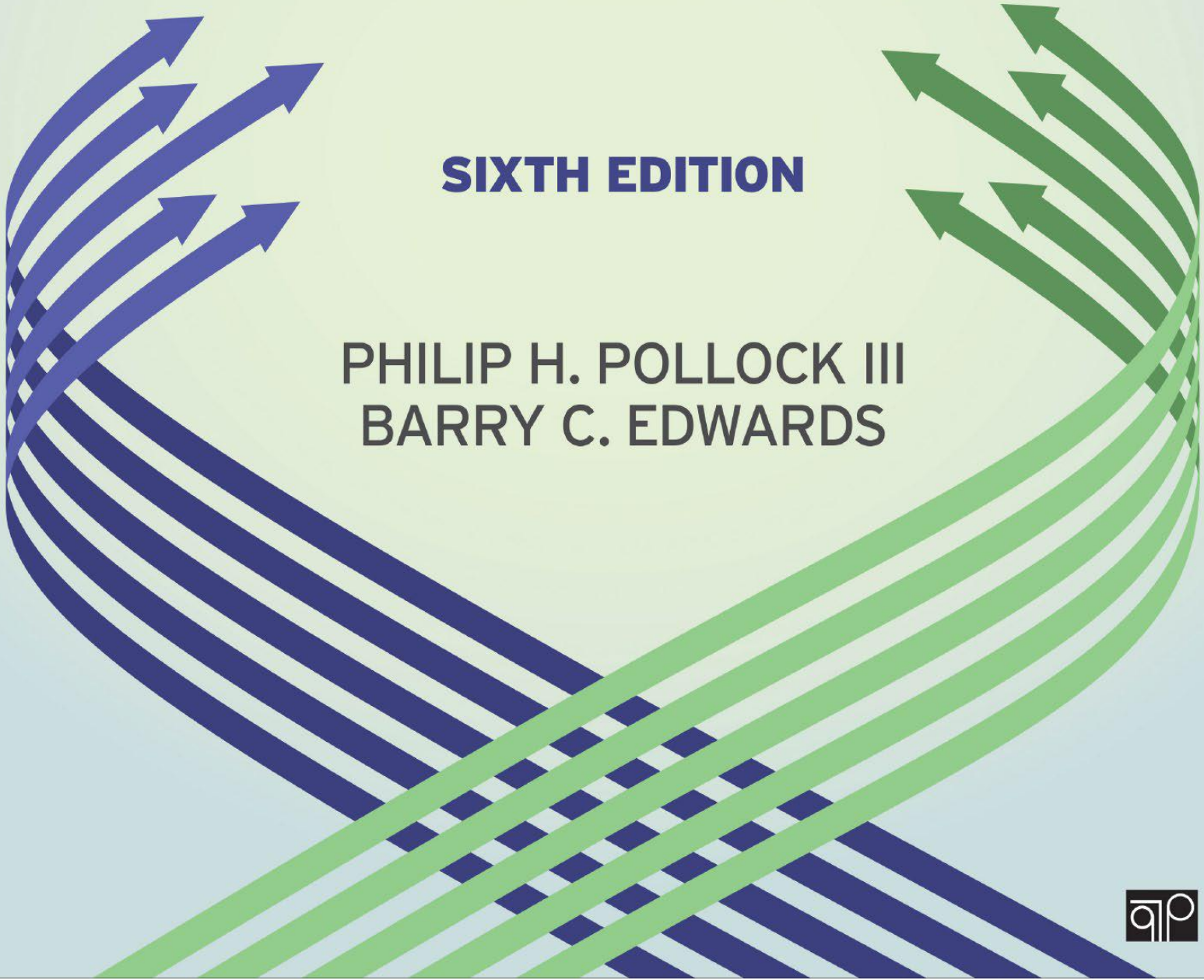
AN IBM SPSS® COMPANION TO POLITICAL ANALYSIS

SIXTH EDITION

PHILIP H. POLLOCK III
BARRY C. EDWARDS



An IBM® SPSS® Companion to Political Analysis

Sixth Edition

An IBM® SPSS® Companion to Political Analysis

Sixth Edition

Phillip H. Pollock III
University of Central Florida

Barry C. Edwards
University of Central Florida

SAGE | CQPRESS

FOR INFORMATION:

CQ Press
An Imprint of SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte Ltd
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Printed in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Pollock, Philip H., III, author. | Pollock, Philip H., III. Essentials of political analysis.

Title: An IBM SPSS companion to political analysis / Phillip H. Pollock III, University of Central Florida, Barry C. Edwards, University of Central Florida.

Other titles: International Business Machines statistical package for the social sciences companion to political analysis

Description: Sixth Edition, Revised and updated edition. | Washington, D.C. : CQ Press, A DIVISION OF SAGE, [2019] | System requirements: System requirements: Microsoft Windows-based computer; CD-ROM drive. | Companion text to: Essentials of political analysis / Philip H. Pollock. | Previous edition: 2016. | Includes bibliographical references and index.

Identifiers: LCCN 2019006730 | ISBN 9781506379654 (Paperback : alk. paper)

Subjects: LCSH: Political science—Research—Handbooks, manuals, etc. | Political statistics—Computer programs—Handbooks, manuals, etc. | Analysis of variance—Computer programs—Handbooks, manuals, etc. | SPSS (Computer file)—Handbooks, manuals, etc.

Classification: LCC JA86 .P64 2019 | DDC 320.0285/555—dc23
LC record available at https://lccn.loc.gov/2019006730

Acquisitions Editor: Scott Greenan
Content Development Editor: Scott Harris
Editorial Assistant: Lauren Younker
Production Editor: Bennie Clark Allen
Copy Editor: Christina West
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Jen Grubba
Cover Designer: Dally Verghese
Marketing Manager: Erica DeLuca

19 20 21 22 23 10 9 8 7 6 5 4 3 2 1

CONTENTS

FIGURES

PREFACE

Since Philip Pollock first published *The Essentials of Political Analysis* nearly twenty years ago, his textbooks have helped countless students learn how to interpret and conduct political analysis. As its name suggests, this book, *An IBM® SPSS® Companion to the Essentials of Political Analysis, Sixth Edition*, serves as a companion to the core textbook, helping students apply general concepts of political analysis using SPSS software.

The longevity of Pollock's political analysis series is a testament to the vital union of theory and practice in teaching political science research methods. Students who appreciate the practical side of research are better prepared to contribute to class discussions of methodological concepts and problems than those who study only abstract principles. Moreover, students often develop a usable skill they can hone as they continue their academic careers or pursue employment opportunities. Whatever their professional goals, students need a solid foundation in basic political analysis techniques. They should learn to manipulate variables, explore patterns, and graph relationships. They also need a working knowledge of powerful yet easy-to-learn software, such as IBM® SPSS® Statistics.* This book instructs students in using SPSS to construct meaningful descriptions of variables and to perform substantive analysis of political relationships. The chapters cover all major topics in data analysis, from descriptive statistics to logistic regression. A final chapter describes several doable research projects, shows how to find relevant data, and lays out a framework for a well-organized research paper. After completing this book, students will have become competent SPSS users, and they will have learned a fair amount about substantive political science, too.

In its essential features—multiple datasets, guided examples, screenshots, graphics instruction, and end-of-chapter exercises—this book continues in the tradition of previous editions. As before, we continue to assume that students using this workbook have never heard of SPSS and have never used a computer to analyze data. However, previous adopters will find some changes and improvements.

DATASETS AND CHAPTER EXERCISES

The SPSS datasets that accompany this book have been thoroughly revised and updated. There are four downloadable datasets: selected variables from the 2016 General Social Survey (the GSS dataset) and the 2016 American National Election Study (the NES dataset), as well as datasets on the 50 states (the States dataset) and 169 countries in the world (the World dataset). You can find names and descriptions for all of the variables in these datasets in the Appendix.

This book has always taken a two-step approach to skill-set learning: (1) Perform the guided examples. (2) Work the exercises. Following this approach, Chapters 1 through 10 include end-of-chapter exercises for students. The exercises are designed to give students opportunities to apply their new skills and to engage students in discovering the meaning of their findings and learning to interpret them. The exercises test a full range of competencies, and most chapters include at least one more-challenging exercise. Most exercises have multiple subparts to give students the opportunity to practice the steps of a research process, including stating hypotheses, analyzing data, and discussing results.

We have included exercises that reflect current scholarly debates in American politics, international relations, comparative politics, and the judicial system. Instructors should feel free to pick and choose exercises that best match the interests and needs of their classes. The pages of this book are perforated, three-hole punched, and 8.5 × 11 inches in size so instructors can use the end-of-chapter exercises as homework assignments that students complete and submit. The formatting of this edition allows

* SPSS is a registered trademark of International Business Machines Corporation.

students to tear out and submit exercises without losing substantive chapter content. After completing this book, students not only are confident SPSS users, they will have also learned a fair amount about substantive political science, too.

Instructors can obtain a complete solutions manual with instructor access to the book's website (edge.sagepub.com/pollock).

CHAPTER ORGANIZATION

The chapter organization for this edition follows that of the previous edition. The "Getting Started" section, an informal preface for students, briefly explains why learning to analyze politics with SPSS is a vitally important skill, provides a roadmap of subsequent chapters, and tells students where they can download the Companion datasets featured throughout the book.

Chapter 1 introduces the SPSS Data Editor, discusses the output Viewer, and illustrates the print procedure. For the sixth edition, we've added some new material to Chapter 1. We've added a brief discussion of using keyboard shortcuts as an alternative to SPSS's graphical user interface (GUI). We've also added a section on formatting SPSS tables. SPSS now features some nice table templates, including an "Academic" style template that generates publication-style tables. Although we emphasize SPSS's GUI, we thought it appropriate to add a section on saving commands in syntax files to enable others to replicate our analysis, an increasingly important dimension of political science research.

Chapter 2 addresses how to describe the central tendency and dispersion of a variable. This chapter also shows how graphics, like bar charts and histograms, in conjunction with tables can enrich one's description of a variable. SPSS's Chart Editor makes it relatively easy to create compelling graphics. Chapter 2 also includes coverage of Case Summaries, which can be quite useful for providing insights into small datasets such as States and World. We've added a brief discussion of sample weights in Chapter 2. SPSS makes weighting observations easy and we use sample weights when analyzing GSS and NES survey data throughout the book.

Chapter 3 describes the main SPSS data transformation procedures, Recode and Compute. Students can use these procedures to create indicator variables, simplify variables, standardize variables, and create additive indexes. We demonstrate each of these operations and show how to modify variable and value labels (a new section in this edition). This chapter also discusses Visual Binning, which is a powerful and efficient alternative to Recode, especially for collapsing interval variables into ordinal categories of roughly equal size.

In Chapter 4, which covers cross-tabulations and mean comparisons, students learn bivariate analysis. This chapter shows how to use graphics, like line charts and bar charts, and box plots to supplement numeric results. Chapter 5 takes the methods introduced an important step further by showing students how to make controlled comparisons. Students use the Crosstabs and Compare Means procedures to obtain and interpret controlled comparisons. This chapter also discusses graphic support for controlled comparisons. Chapters 4 and 5 feature fresh examples and updated screenshots but cover the same procedures as prior editions.

Chapter 6 uses the One-Sample T Test and Independent-Samples T Test procedures to show students how to make inferences about the means of interval-level variables. In this edition, we show students how to visualize mean comparisons using Error Bar graphs that show point estimates and confidence intervals—it's a nice way to reinforce the core concepts of inferential statistics. We've also added a section on making inferences about proportions using an example that cuts across several subfields: public support for the government spying on U.S. citizens. The standard SPSS toolkit leaves a lot to be desired for testing inferences about proportions, but we are careful to note where SPSS provides only approximate answers and how to calculate *z*-scores. Like the prior edition, this edition avoids confusing students on one-tailed versus two-tailed tests of significance. Instead, Chapter 6 focuses exclusively on two-tailed tests and on the 95 percent confidence interval.

Chapter 7 covers chi-square and measures of association for nominal and ordinal variables. We cover the same methods in this chapter as covered in prior editions, but we now show students how they can depict the analyses using multiple line charts (for ordinal-level variables) and clustered bar charts (for nominal-level variables). Additionally, we've updated the featured example in this chapter to examine public support for use of force to solve international problems, a topic that's relevant to multiple subfields.

In Chapter 8, students work through guided examples to learn to use the Correlate (Bivariate) procedure and the Regression (Linear) procedure. Chapter 8 also discusses advanced scatterplot editing using the Chart Editor. The featured example in this edition looks at the correlation among different interval-level measures of democracy around the world and the effect of democratic development on global conflict levels. This chapter now concludes with a brief discussion of formatting regression results tables for presentation or publication.

Chapter 9 shows students how to use multiple regression analysis to conduct different types of analysis. Students learn how to perform regression analysis with multiple dummy variables and model interaction in multiple regression. This chapter gives students the opportunity to apply some of the variable transformation methods they learned in Chapter 3. We've added a section on graphing an interaction relationship on a scatterplot; seeing the analysis should help students properly interpret an interaction term in multiple regression results.

Chapter 10 covers binary logistic regression. Students learn how to conduct logistic regression analysis, interpret results, and visualize relationships in terms of probabilities: marginal effects at the means (MEMs) and marginal effects at representative values (MERs). In this edition, we analyze a contemporary civil rights issue: whether businesses that provide wedding services should be required to serve gay couples.

Chapter 11 guides students on conducting their own political analysis. We've revised our discussion of "doable ideas" to identify seven political science topics students can research using the four Companion datasets. Like prior editions, this chapter shows students how they can input other datasets into SPSS to conduct their own research.

These chapters are organized in the way that we typically teach our research methods courses. We prefer to cover the logic of description and hypothesis testing before introducing inferential statistics and statistical significance. We assign our students exercises and problems sets, often working with them in lab sessions. However, with a little rearranging of the chapters, this book will prove useful for instructors who do things differently. For example, if your course culminates in a significant research paper, you may prefer to assign Chapter 11 (Doing Your Own Political Analysis) at the beginning of the course, perhaps after the Introduction to SPSS (Chapter 1). As noted above, we know that political science research methods courses are often taught by instructors with different subfield interests. While focusing on the essential methods used by all subfields, we have attempted to engage varied substantive interests with our examples and exercises. Of course, one size cannot fit all, so we encourage instructors to add and subtract material to help students learn to interpret and conduct political analysis.

LEARNING TO ANALYZE POLITICS WITH SPSS

As before, each chapter is written as a step-by-step tutorial, taking students through a series of guided examples and providing many annotated screenshots. Because of the revised and updated datasets, all of the examples and screenshots have been updated. To augment the first step, we have produced a set of screencasts. These are short, tutorial videos that demonstrate essential methods using SPSS and the Companion datasets. Screencasts cover all of the guided examples—plus some other topics of interest, such as creating bubble plots and producing nicely formatted SPSS tables in Microsoft Word.

Screencast
Introduction to SPSS

To make the screencasts as accessible as possible, we've embedded QR codes ("quick response" codes) to related screencasts in the text. With a QR reading app, students can simply point their

smartphone cameras at a screencast link, which looks like a bar code stamp, and that tutorial video will play on their phone. These smartphone apps can be downloaded for free and they are built into some web browsers. We encourage students to watch screencast videos on their phones as they practice new skills on their computers.

Students can find and view all of our screencast tutorials on the book's website (edge.sagepub.com/pollock), where they can also download all of the datasets featured in this book.

The text guides students through essential methods of analyzing politics with SPSS step by step. We've prepared hundreds of screenshots to show students exactly how to use the program's GUI. To help students use the book as a reference, this edition continues to list procedures covered in each chapter. In this edition, we also highlight the procedures we discuss. For example, in Chapter 1, students learn they can select File ▶ Open ▶ Data to open one of the Companion datasets. (The underlined letters in the command show the keyboard shortcut, another convention introduced in this edition to help students get the most out of SPSS.)

Adopters can obtain a complete set of SPSS syntax files that can be run to replicate the guided examples and the exercises. Instructors may find these files useful for re-creating the graphics, for performing the exercises, or for troubleshooting students' difficulties with the examples or exercises.

SPSS 25 is featured here, but anyone running release 12 or later can profitably use this book. As far as we can tell, SPSS 25 uses the same interfaces for the core procedures discussed in this book as it did in prior versions. SPSS 25 tables and graphs, however, have a decidedly more modern design than those produced by earlier versions.

Any student who has access to SPSS, either the full version or the "Grad Pack" version, can use this book. The Grad Pack version of more recent editions of SPSS takes the place of "Student" versions of older editions of SPSS. Unlike the Student version, the Grad Pack versions of SPSS are not limited to analyzing a small number of variables and observations, so we no longer need to create pared-down versions of the companion datasets as we have for past editions of this book. To perform logistic regression analysis in Chapter 10, however, students will need access to the full version of SPSS, as this feature is not available in the Grad Pack version.

There are many commonalities across post-12 releases, including the graphic dialogs and the Chart Editor. There are currently three ways to obtain unedited charts: Chart Builder, Graphboard Template Chooser, and Legacy Dialogs. Although we have attempted to migrate to the more recent tools (for how many more releases will SPSS support routines labeled "legacy"?), the Legacy Dialogs still offer superior intuitiveness and flexibility. (One notable exception is bubble plots, which can only be created in the Graphboard Template Chooser. Bubble plots are not covered in detail in this book, but a screencast, mentioned in Chapter 8, demonstrates how to create this graphic form.) In any event, this edition carries forward the emphasis on elegant graphic displays to complement and clarify empirical results. We have sought to instruct students in using the Chart Editor to emulate the techniques advocated by Edward R. Tufte and other experts on the visual display of data.

ACCOMPANYING CORE TEXT

Instructors will find that this book makes an effective supplement to any of a variety of methods textbooks. However, it is a particularly suitable companion to our own core text, *The Essentials of Political Analysis*. The textbook's substantive chapters cover basic and intermediate methodological issues and ideas: measurement, explanations and hypotheses, univariate statistics and bivariate analysis, controlled relationships, sampling and inference, statistical significance, correlation and linear regression, and logistic regression.

As noted above, each chapter also includes end-of-chapter exercises. Students can read the textbook chapters, do the exercises, and then work through the guided examples and exercises in *An IBM® SPSS® Companion to Political Analysis*. The idea is to get students in front of the computer, experiencing political research firsthand, early in the academic term. An instructor's solutions

manual, available for download online at edge.sagepub.com/pollock and free to adopters, provides solutions for all of the textbook and workbook exercises.

ACKNOWLEDGMENTS

We received more than a few friendly e-mails suggesting ways to improve this book, and we are grateful for this advice. We also thank current and past reviewers for pointing us in the right direction: Holly Brasher, University of Alabama at Birmingham; Matthew Davis, University of Delaware; Jason Kehrberg, University of Kentucky; Thad Kousser, University of California, San Diego; Nancy Martorano, University of Dayton; Matthew Streb, Northern Illinois University; Brian Vargus, Indiana University–Purdue University Indianapolis; Julian Westerhout, Illinois State University; Lindsay Benstead, Portland State University; William Field, Rutgers, The State University of New Jersey; Rob Mellen, Mississippi State University; Brian Frederick, Bridgewater State College; Krista Jenkins, Fairleigh Dickinson University; Renato Corbetta, University of Alabama at Birmingham; Changkuk Jung, SUNY College at Geneseo; and Wesley Hussey, California State University, Sacramento.

We thank our University of Central Florida colleagues Bruce Wilson and Kerstin Hamann for helping us with ideas for exercises on comparative politics. We also give special thanks to Bill Claggett of Florida State University, who attended graduate school at the University of Minnesota with Philip Pollock, for sharing his SPSS know-how with us.[1] Bill passed away in 2017. He will be missed by his students, colleagues, family, and many friends. Many encouraging people have helped us make this a better book, including all the students who have taken the Scope and Methods of Political Science course with us over the years. Any remaining errors, however, are ours.

We gratefully acknowledge the encouragement and professionalism of everyone associated with the College Division of CQ Press and SAGE Publications: Scott Greenan, acquisitions editor; Lauren Younker, editorial assistant; Scott Harris, senior content development editor; Bennie Clark Allen, production editor; Christina West, copy editor; and Erica DeLuca, marketing manager. It is a real joy to work with such a talented team. We owe an enduring debt to Charisse Kiino, vice president of editorial—without her commitment, this entire project would still be an incoherent jumble of datasets and syntax files lying fallow on an old computer.

[1] It was Claggett who alerted us to an ancient flaw in the Compute Variable procedure: Multiply 0 times missing, and SPSS interprets the product as 0, not as missing. This flaw—a serious defect, in our opinion—is discussed in Chapter 9.

Getting Started

To get started with this book you will need

- Access to an SPSS-compatible computer with an Internet connection

As you have learned about political research and explored techniques of political analysis, you have studied many examples of other people's work. You may have read textbook chapters that present frequency distributions, or you may have pondered research articles that use cross-tabulation, correlation, or regression analysis to investigate interesting relationships between variables. As valuable as these learning experiences are, they can be enhanced greatly by performing political analysis firsthand—handling and modifying social science datasets, learning to use data analysis software, obtaining your own descriptive statistics for variables, setting up the appropriate analysis for interesting relationships, and running the analysis and interpreting your results.

This book is designed to guide you as you learn these valuable practical skills. In this volume you will gain a working knowledge of SPSS, a data analysis package used widely in academic institutions and business environments. SPSS has been in use for many years (it first appeared in 1968), and it contains a great variety of statistical analysis routines—from basic descriptive statistics to sophisticated predictive modeling. It is extraordinarily user friendly. You can execute most of the data analysis procedures discussed in this book using SPSS's graphical user interface (GUI). Although this book assumes that you have practical knowledge of your computer's operating system and know how to perform elemental file-handling tasks, it also assumes that you have never heard of SPSS and that you have never used a computer to analyze data of any kind. By the time you complete the guided examples and the exercises in this book, you will be well on your way to becoming an SPSS aficionado. The skills you learn will be durable, and they will serve you well as you continue your educational career or enter the business world.

This book's chapters are written in tutorial fashion. Each chapter contains several guided examples, and each includes exercises at the end. You will read each chapter while sitting in front of a computer, doing the analyses described in the guided examples, and analyzing the datasets that accompany this text. Each data analysis procedure is described in step-by-step fashion, and the book has many figures that show you what your computer screen should look like as you perform the procedures. Thus, the guided examples allow you to develop your skills and to become comfortable with SPSS. The end-of-chapter exercises allow you to apply your new skills to different substantive problems.

This book will provide you with a solid foundation in data analysis. You will learn to obtain and interpret descriptive statistics (Chapter 2), to collapse and combine variables (Chapter 3), to perform cross-tabulation and mean comparison analysis (Chapter 4), and to control for other factors

that might be affecting your results (Chapter 5). Techniques of statistical inference (Chapters 6 and 7) are covered, too. On the more advanced side, this book introduces correlation and linear regression (Chapter 8), and it teaches you how to use dummy variables and how to model interaction effects in regression analysis (Chapter 9). If you are running the Full Version of SPSS, Chapter 10 provides an introduction to logistic regression, an analytic technique that has gained wide currency in recent years. Chapter 11 shows you how to read data into SPSS, and it provides guidance on writing up your results.

DOWNLOADING THE DATASETS

To access the datasets that you will analyze in this book, navigate to this site:

edge.sagepub.com/pollock

In the left navigation pane of the page, hover over the title of this book, *An IBM® SPSS® Companion to Political Analysis, 6th edition*, and click the Datasets link (Figure I-1).

Download the datasets to a USB drive or other portable storage (or download the datasets to the default location, and then copy them to a USB drive or cloud storage you can access from anywhere you plan on working with the datasets). There are four datasets.

1. **GSS** (file name: gss.sav). This dataset has selected variables from the 2016 General Social Survey (GSS), a random sample of 2,867 adults aged 18 years or older, conducted by the National Opinion Research Center and made available through the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan.[1] Some of the scales in the GSS dataset were constructed by the authors. All other variables retain their original names. The names and basic descriptions of variables in the GSS dataset appear in Appendix Table A-1.[2]
2. **NES** (file name: nes.sav). This dataset includes selected variables from the 2016 American National Election Study (NES), a random sample of 4,271 citizens of voting age, conducted by the University of Michigan's Institute for Social Research (ISR) and made available through the ICPSR.[3] With the exception of scales constructed by the authors, all variables in the NES dataset retain the variable names assigned to them by the ISR. The names and basic descriptions of variables in the NES dataset appear in Appendix Table A-2.[4]
3. **States Dataset** (file name: states.sav). This dataset includes variables on each of the fifty states. These variables were compiled by the authors from various sources. The names and basic descriptions of variables in the states dataset appear in Appendix Table A-3.

---

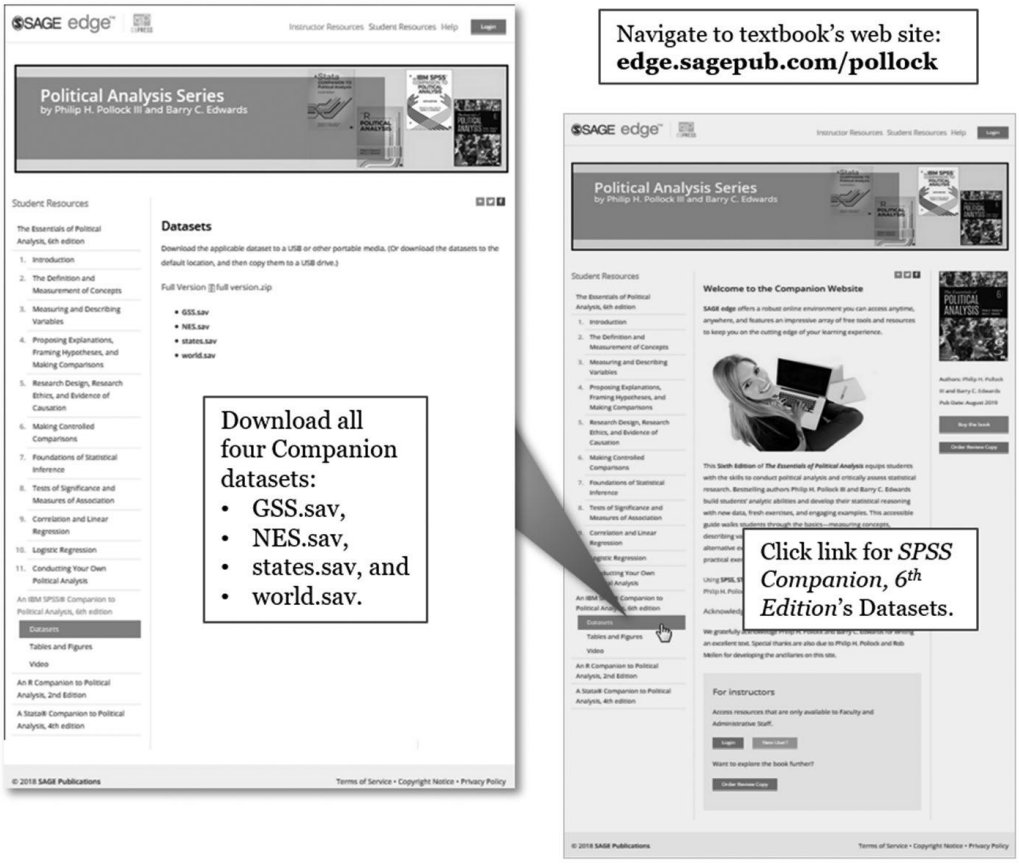
[1] The GSS Dataset (GSS) was created from the General Social Survey 1972–2016 Cumulative Data File. Smith, Tom W., Michael Hout, and Peter V. Marsden, General Social Survey, 1972–2016 [Cumulative File], ICPSR36797-v1 (Chicago, IL: National Opinion Research Center/Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributors], 2017-11-14), https://doi.org/10.3886/ICPSR36797.v1. We encourage students to explore the full version of the 2016 GSS dataset online as well as other years of the survey for additional insights.

[2] To find information on coding and question wording, visit the following link at the University of California–Berkeley's Social Data Archive and search the alphabetical variable list: http://sda.berkeley.edu/D3/GSS16/Doc/hcbk.htm.

[3] American National Election Studies, University of Michigan, and Stanford University, ANES 2016 Time Series Study, ICPSR36824-v2 (Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2017-09-19), https://doi.org/10.3886/ICPSR36824.v2.

[4] For specific coding and question wording, go to the following link and search codebooks: http://www.electionstudies.org/studypages/anes_timeseries_2016/anes_timeseries_2016.htm.

FIGURE I-1 Downloading Companion Datasets from the SAGE Edge Website



4. **World Dataset** (file name: world.sav). This dataset includes variables on 169 countries. Many of these variables are based on data compiled by Pippa Norris (John F. Kennedy School of Government, Harvard University) and made available to the scholarly community through her Internet site.[5] Other variables were compiled by the authors from various sources. The names and basic descriptions of variables in the world dataset appear in Appendix Table A-4.

As you work your way through this book, you will modify these datasets—recoding some variables, computing new variables, and otherwise tailoring the datasets to suit your purposes. You will need to make personal copies of the datasets and store them on a removable drive, such as a USB flash drive, or cloud storage that you can access from home, school, or wherever else you may want to work with the datasets.

When you begin each chapter's guided examples, or when you do the exercises, you will want to insert your personal media into the appropriate computer drive. SPSS will read the data from the drive. (Chapter 1 covers this operation in detail.) If you make any changes to a dataset, you can save the newly modified dataset directly to your drive. Alternatively, your computer lab's administrator may permit you to work on datasets that have been copied to the lab computer's desktop or to a folder designated for such a purpose. In any case, if you have modified a dataset during a data analysis session, it is important that you copy the dataset to your personal drive and take the datasets with you.

[5] http://www.pippanorris.com.

SPSS FULL AND GRAD PACK VERSIONS: WHAT IS THE DIFFERENCE?

Campus computer labs run Full Version SPSS. Your institution may offer Full Version SPSS as an app, or you may have rented it. Alternatively, along with this workbook, you may have purchased a "Grad Pack" Version of SPSS, which is currently available in base, standard, and premium editions. After you install the Grad Pack Version of SPSS, you can run it from your own desktop or laptop computer.

In terms of the guided examples and exercises in this book, how does the Grad Pack Version compare with Full Version? Since this book focuses on core statistical and graphing methods, the SPSS Grad Pack, in any edition, should suffice. The base edition of the SPSS Grad Pack, however, is not equipped to perform logistic regression analysis, which is the subject of Chapter 10. If your course covers Chapter 10, or you plan on doing logistic regression analysis, you'll want to purchase the standard edition of the Grad Pack or plan to use a Full Version for that analysis.

WATCH SCREENCASTS FROM SAGE EDGE

Screencast

Introduction to SPSS

To augment the step-by-step instructions in this text, we have produced a set of screencasts. These are short, tutorial videos that show you how to use SPSS and analyze the companion datasets. Screencasts cover all of the guided examples—plus some other topics of interest.

To make screencasts as accessible as possible, we've embedded QR codes ("quick response" codes) in the text. With a QR reading app, you can simply point your smartphone camera at a screencast link and that tutorial video will play automatically. These smartphone apps can be downloaded for free and they are built in on some web browsers. We encourage you to watch screencast videos as you practice new skills on your computer. If you want to watch a screencast on your computer, you can find and view all of our screencast tutorials on the book's website: **edge.sagepub.com/pollock**.

CHAPTER

1

Introduction to SPSS

Watch screencasts of the guided examples in this chapter. **edge.sagepub.com/pollock**

Procedures Covered

File ▶ Open ▶ Data
Edit ▶ Options
Utilities ▶ Variables
Analyze ▶ Descriptive Statistics ▶ Frequencies
File ▶ Print
File ▶ Open ▶ Output
Format ▶ TableLooks (in Table Editing window)
File ▶ New ▶ Syntax

In this chapter, we take readers on a quick 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.

Screencast
Introduction to SPSS

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.

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 (Figure 1-2).

[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.

FIGURE 1-1 SPSS "Welcome Screen"

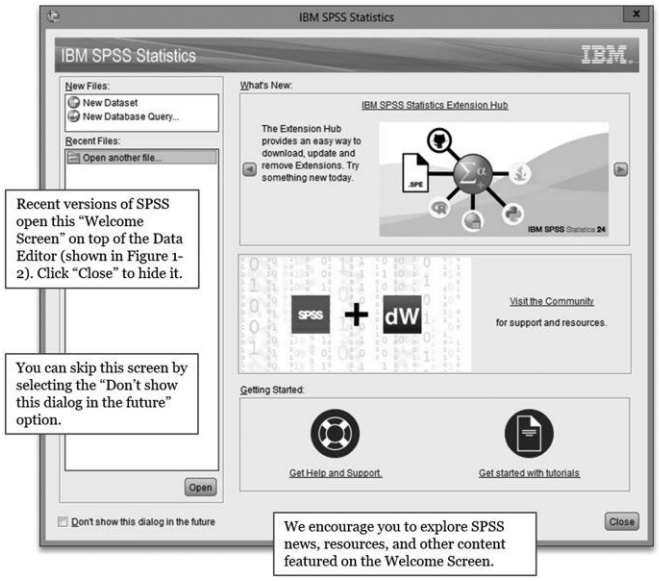
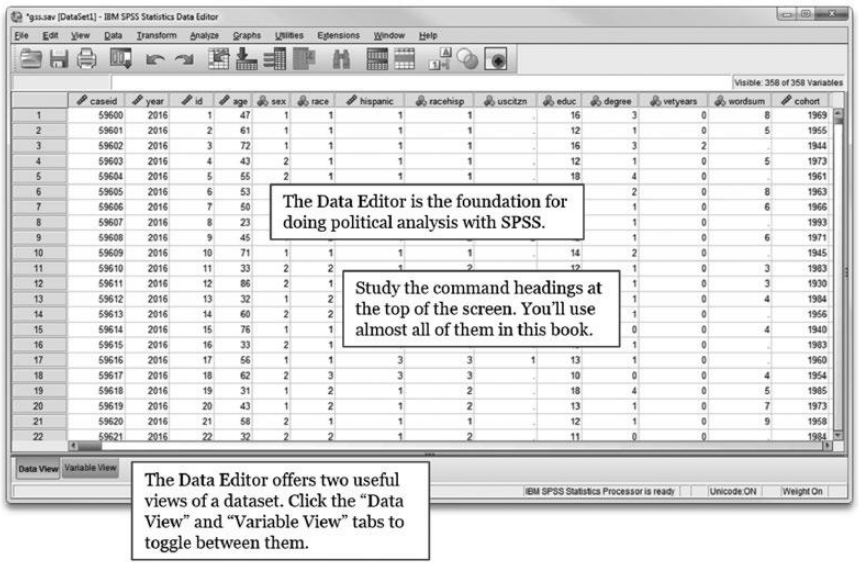


FIGURE 1-2 SPSS Data Editor in Data View

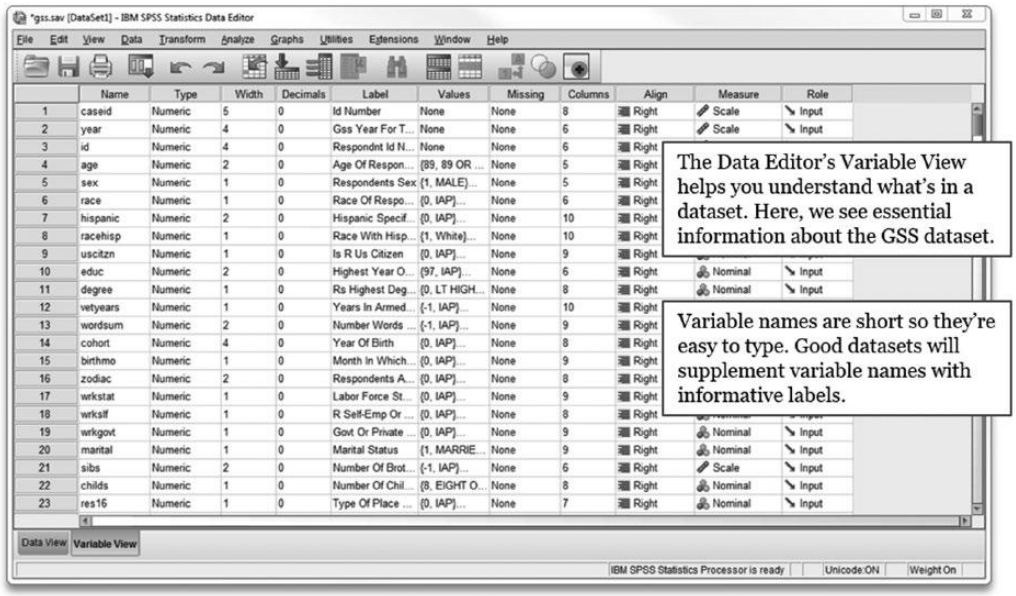


Notice the two tabs at the bottom of the Data Editor window: Data View and Variable View. The SPSS Data Editor offers two "views" of the dataset. Both views are useful. Select Data View or Variable View by clicking one of the tabs at the bottom of the Data Editor.

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. 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.

Scroll right to see information recorded from the first respondent. You can tell that the first respondent in the dataset is 47 years old (see the first value in the "age" column). You can also

FIGURE 1-3 SPSS Data Editor in Variable View



see that this respondent has a 1 in the "sex" column and a 1 in the "race" column. Storing information as numbers is efficient, but it's not immediately clear what these numbers really mean. To paint a more complete word portrait of this respondent, you need to see how all the variables are coded.

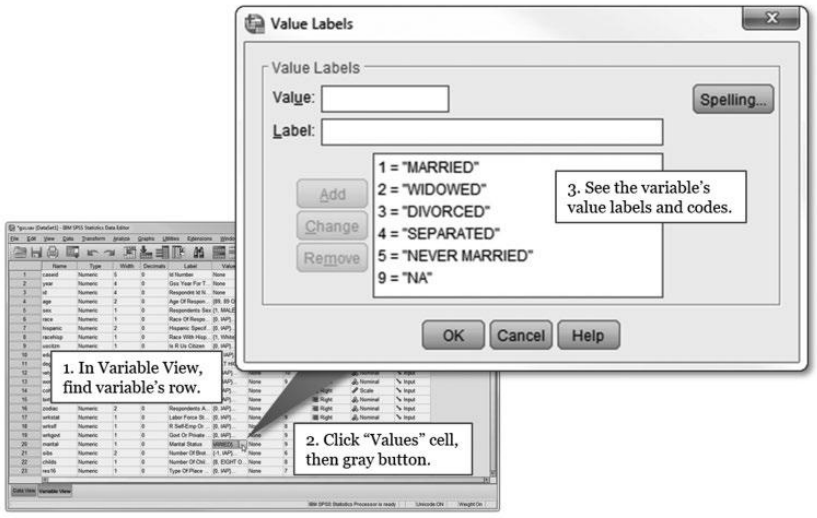
To see how the GSS variables are coded, click the Variable View tab (Figure 1-3). The Variable View, among other useful features, shows the word labels that the researcher has assigned to the numeric codes. This view shows complete information on the meaning and measurement of each variable in the dataset. (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 Name, Label, Values, and Missing. 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 labels, long descriptors (up to 256 characters are allowed), to provide more detailed information about variables. For example, when SPSS analyzes the GSS variable mobile16, 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 mobile16 shows up as "Geographic Mobility Since Age 16"—a bit more descriptive than "mobile16."

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 sex and race variables to find out what it means when they're coded 1. To find out the value labels for the sex 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 sex. Unless you instruct it to do otherwise, SPSS will apply these labels to its analysis of the sex 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 1-4). 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 1-4 Value Labels of a Variable in the Data Editor



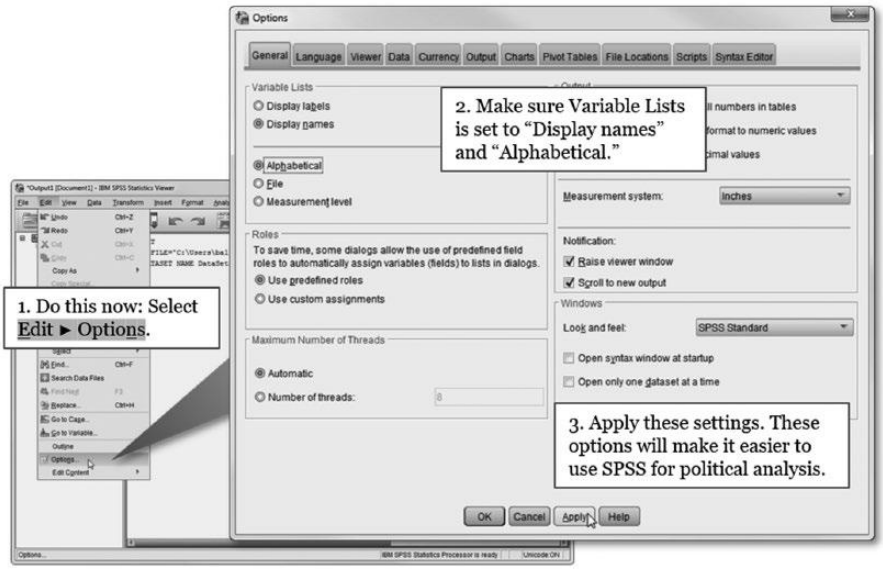
Finally, a word about the Missing column. Sometimes a dataset does not have complete information for all variables and observations. This happens for a variety of reasons; researchers may add or remove questions from the survey, some questions may not apply to everyone, or the response may not be clear. 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 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.

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 opened the Options menu, you are set to go and can click the Cancel button.)

FIGURE 1-5 Setting Options for Variable Lists

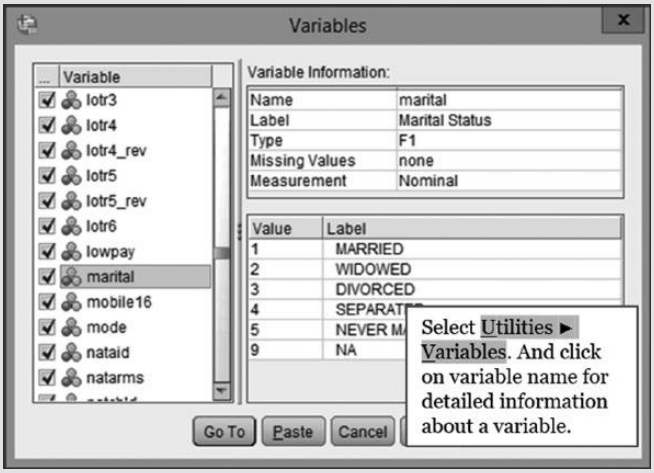


A CLOSER LOOK: VARIABLES UTILITY

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 (Figure 1-6).

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.

FIGURE 1-6 Retrieving Coding Information



THE VIEWER

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 methods of 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. For this example, select Analyze ▶ Descriptive Statistics ▶ Frequencies. The Frequencies window appears (Figure 1-7).

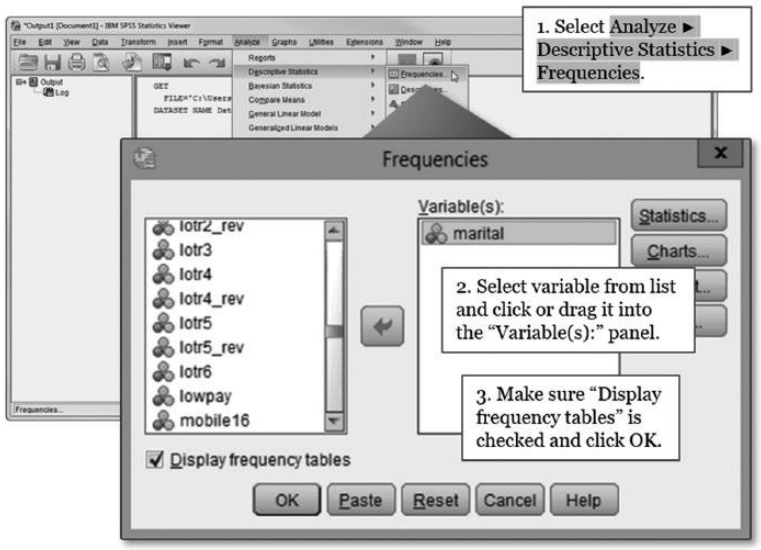
You'll notice that the Frequencies window has two panels. On the right is the (currently empty) Variable(s) panel. This is the panel where you enter the variables you want to analyze. On the left you see the names of all the variables in GSS in alphabetical order, just as you specified in the Options 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 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 1-8).

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 1-8.

When you execute a procedure using the GUI, SPSS temporarily stores the information you inputted so you can return to the same window and adjust your selections. This is particularly useful

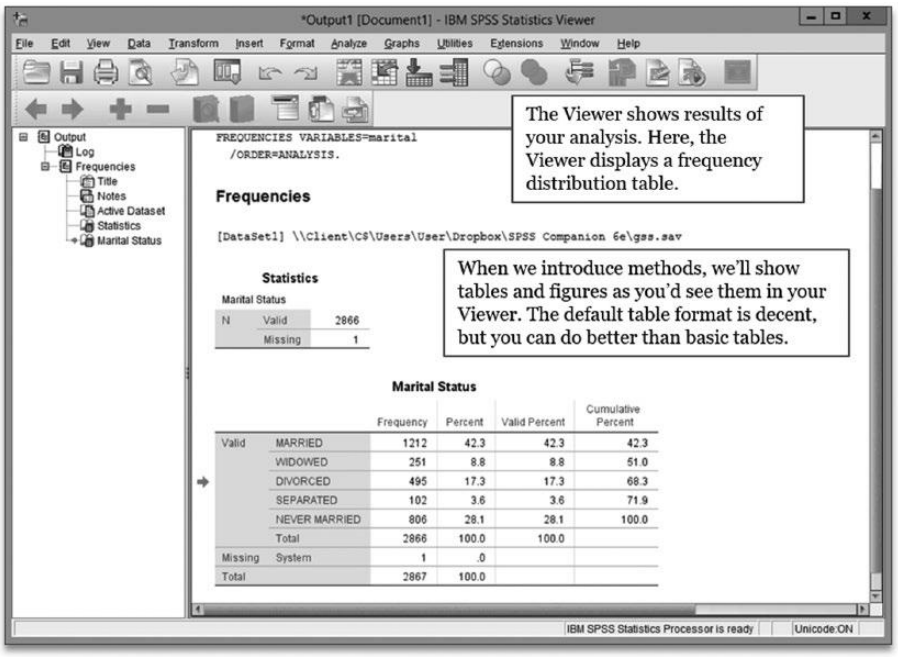
FIGURE 1-7 Analyzing Frequencies



[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.)

FIGURE 1-8 Sample Table Output in the SPSS Viewer



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 1-7 and 1-8, 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.

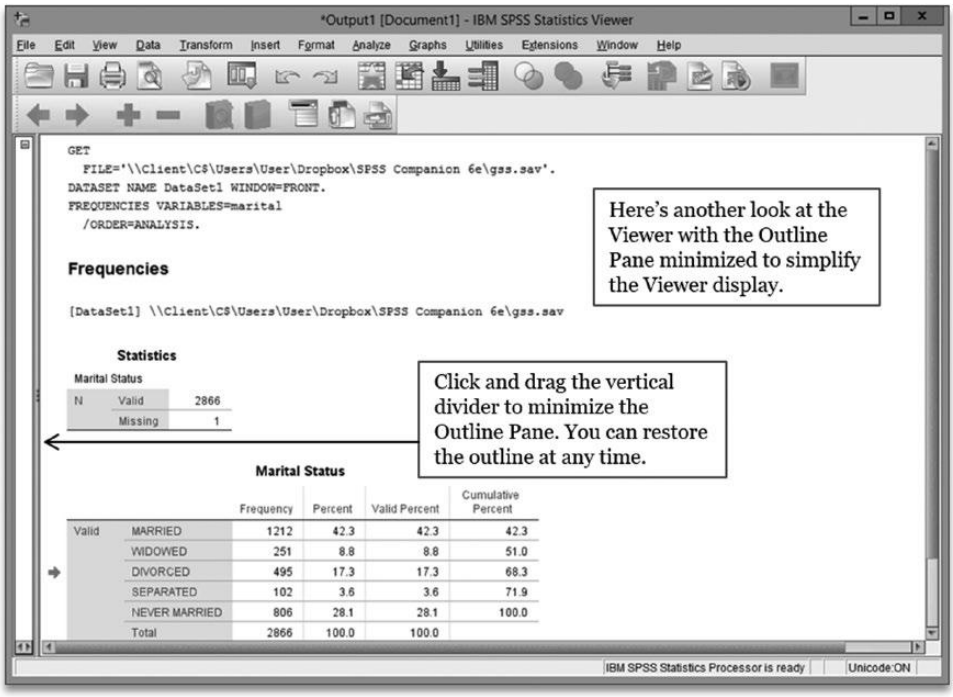
A CLOSER LOOK: KEYBOARD SHORTCUTS

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", 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 notations above, 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 1-9. 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.

FIGURE 1-9 SPSS Viewer with the Outline Pane Minimized



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 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.
- 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.

SELECTING, PRINTING, AND SAVING OUTPUT

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

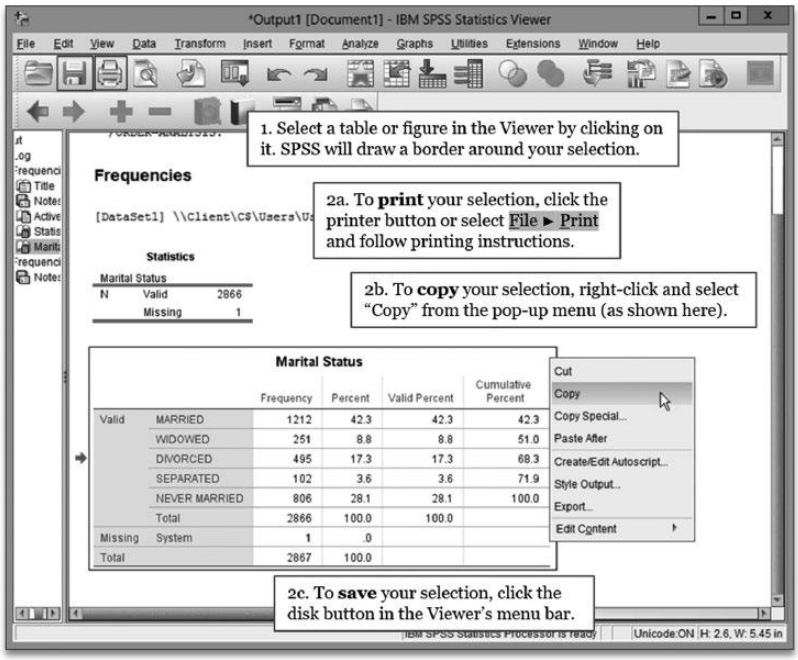
Printing desired results requires, first, that you select the output or portion of output you want to print. A quick and easy way to select a single table or chart is to place the cursor anywhere on the desired object and click once. For example, if you want to print the marital frequency distribution 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 1-10). Click the Printer icon on the Viewer menu bar or select File ▶ Print. The Print window opens. In the window's Print Range panel, the radio button next to "Selected output" should already be clicked. Clicking OK would send 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 wanted to print the frequency table and the statistics table, first click 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 1-10). Recent versions of SPSS have greatly improved table formatting over prior versions and the table copied from the Viewer now looks decent in a document.

To save your Viewer output, simply click the familiar Save icon on the Viewer menu bar (refer to Figure 1-10 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.

FIGURE 1-10 Selecting, Printing, and Saving Output



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 (see Figure 1-11). 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.

HOW TO FORMAT AN SPSS TABLE

When you analyze political science data, you'll create 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 Exporting Viewer Output

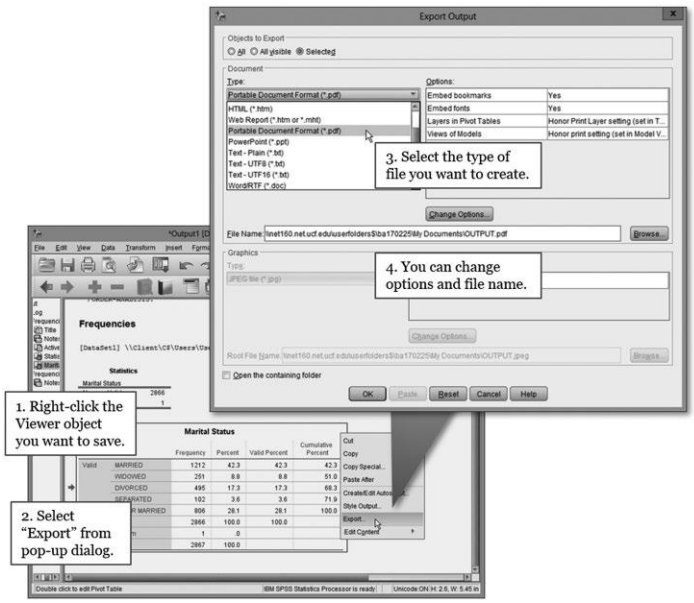
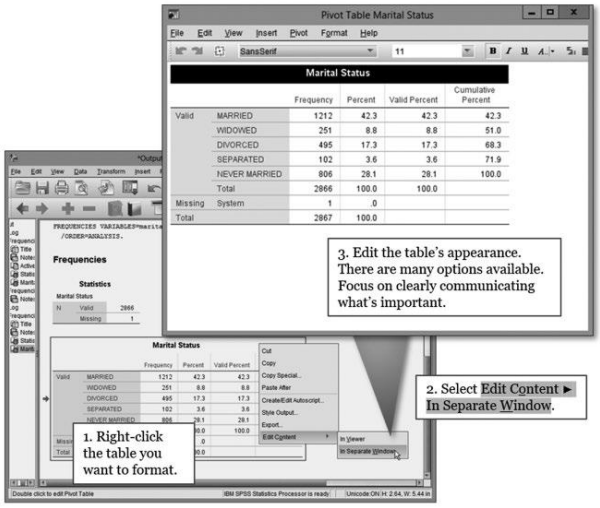


FIGURE 1-12 Formatting a Table



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 Content ▶ In Separate Window" option. You'll see your table in a new, separate window with formatting tools (see Figure 1-12). This table editing window 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.

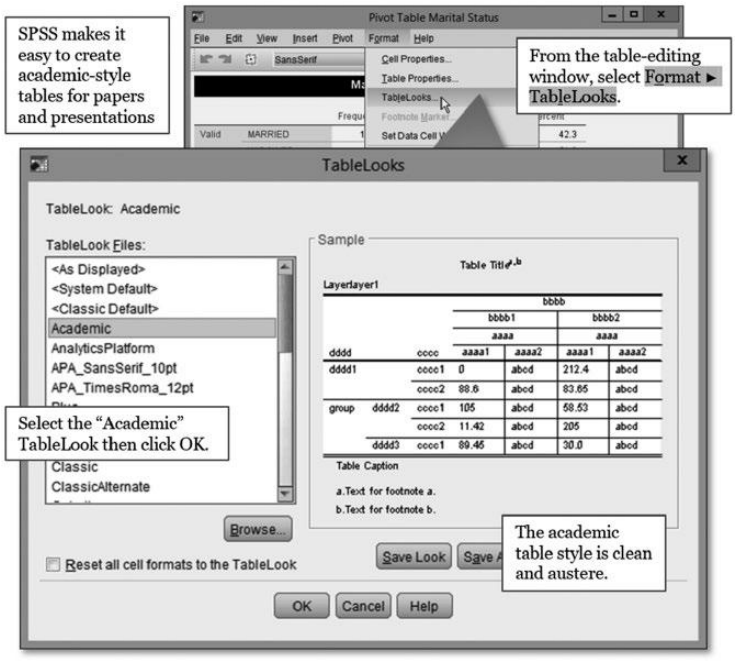
Here's a suggestion to help you quickly create professional-style tables. The Academic-style tables are particularly good. In the separate table editing window, select Format ▶ TableLooks (this procedure is only available in the table editing window). TableLooks are a pre-defined set of table styles. There are many styles to choose from, 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 close the separate table editing window by clicking the X button in the upper-right corner. Now 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 this:

Marital Status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MARRIED	1212	42.3	42.3	42.3
	WIDOWED	251	8.8	8.8	51.0
	DIVORCED	495	17.3	17.3	68.3
	SEPARATED	102	3.6	3.6	71.9
	NEVER MARRIED	806	28.1	28.1	100.0
	Total	2866	100.0	100.0	
Missing	System	1	.0		
Total		2867	100.0		

FIGURE 1-13 Creating Academic-style Tables



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, select Academic from the TableLook options, and then click OK. Tables in this pre-defined 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.

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 framework for 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 some data analysis. The procedures you execute using SPSS's GUI can also be stated as text commands. If you look closely at Figure 1-9, you'll see that SPSS displays the text equivalent of the frequency analysis of the marital variable in the Viewer:

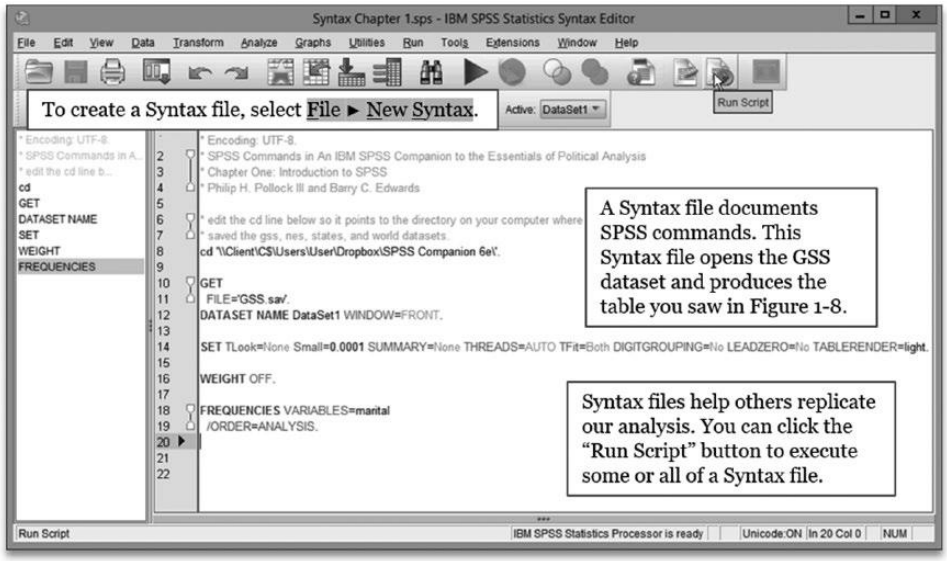
```
FREQUENCIES VARIABLES=marital
/ORDER=ANALYSIS.
```

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 can copy and paste these text-equivalent commands into a syntax file. To create a complete

FIGURE 1-14 The Syntax Editor



syntax file, 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).

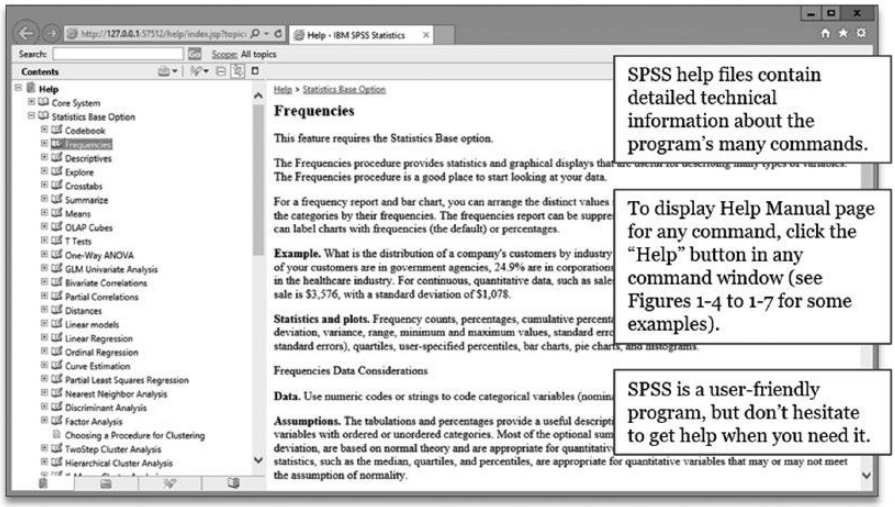
Another way to save your commands in a syntax file is the "Paste" button that appears in GUI windows that execute commands. Take another look at Figure 1-7, the procedure we used to generate a frequency distribution table for the marital variable. Next to the OK button, you'll see the Paste button. If 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 add the commands to open the dataset and some user-friendly comments, but it's a very convenient feature for generating syntax files that replicate your analysis.

GETTING HELP

To view the formal how-to manual 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 1-7). SPSS retrieves the technical manual information and displays it in a web browser (Figure 1-15).

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. You can find a complete list of screencasts on the SAGE Edge website for this book: **edge.sagepub.com/pollock**.

FIGURE 1-15 SPSS Help Manual



Name: _____ Date: _____

E-mail: _____ Section: _____

CHAPTER 1 EXERCISES

1. (Dataset: gss.sav. Variables: income16, attend.) In this chapter, we spent some time using the Data View and the Variable View to describe the first respondent in the GSS dataset. In this exercise, you will use your familiarity with the Data Editor to find out this respondent's income (income16) and how often this respondent attends religious services (attend).

   A. With the GSS dataset open, go to the Data View. What numeric code does the first respondent have on income16? A code of (fill in the blank) _____. Go to the Variable View. Just as you did earlier in this chapter, find income16 and click in the Values cell. What is this respondent's income? (circle one)

   $25,000 to $29,999 $60,000 to $74,999 $170,000 or over

   B. Return to the Data View. What is this respondent's code on the variable attend? A code of (fill in the blank) _____. Go to the Variable View. How often does this respondent attend religious services? (circle one)

   Never Once a year 2–3 times a month

2. Suppose that you have just opened the World, States, or NES 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? (check two)

   - Display labels
   - Display names
   - Alphabetical
   - File
   - Measurement level

3. The GSS dataset contains the variable happy, which asks respondents how happy they are. Find the variable happy in the Data Editor's Variable View and answer the following questions.

   A. What is the descriptive label for the variable happy? _____

   B. Respondents who say they are "very happy" have which numeric code? (circle one)

   1 2 3

4. Generate a frequency distribution table for the variable happy in the GSS dataset using the Analyze ▶ Descriptive Statistics ▶ Frequencies procedure.

   A. Print the table.

   B. Copy/paste the table into a blank document. Edit the table for appearance and readability, and then print it.

5. SPSS works with different types of files to analyze data and produce results. 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	?

6. A political scientist wants to analyze civic culture in the United States. Civic culture is an important concept but is difficult to measure empirically. The researcher could consider several different variables.

   A. Which variable in the States dataset records the number of years of social studies that states require students to take to graduate high school? _____

   B. Which variable in the States dataset records the percentage of the voting age population that turned out to vote in the most recent federal election (for which data are available)? _____

   C. Which variable in the States dataset records the percentage of state residents who frequently attend religious services? _____

D. Which variable in the States dataset records the percentage of state residents who do voluntary community service? _____

7. Chapter 1 showed you how to view detailed information about a variable named marital in the GSS dataset. This variable identifies the survey respondents' marital statuses with numeric codes and assigns a label to each numeric value. The NES dataset contains a similar variable, also named marital, that records the marital statuses of its respondents. The numeric coding for marital in the NES is a little different than its GSS counterpart, however. It's important to pay close attention to how variables are coded. Fill in the following table to identify the labels that correspond to numeric codes of the NES marital variable.

NES Variable "marital"	
Numeric Code	**Value Label**
1	?
2	?
3	?
4	?
5	?
6	?

8. A political scientist wants to study health outcomes in countries around the world. Good government may improve health outcomes; a healthy country may also be more politically stable. To study the health-politics relationship, the researcher could consider several variables in the World dataset that measure health in countries around the world. For each variable, briefly describe what it measures in your own words.

A. fertility _____

B. hiv_percent _____

C. infant_mortality _____

D. spendhealth _____

E. unnoncom _____

CHAPTER

2

Descriptive Statistics

Watch screencasts of the guided examples in this chapter. **edge.sagepub.com/pollock**

Procedures Covered

Analyze ▶ Descriptive Statistics ▶ Frequencies
Data ▶ Weight Cases
Graphs ▶ Legacy Dialogs ▶ Histogram
Analyze ▶ Reports ▶ Case Summaries

Analyzing descriptive statistics is the most basic—and sometimes the most informative—form of analysis you will do. Descriptive statistics reveal two attributes of a variable: its typical value (central tendency) and its spread (degree of dispersion or variation). The precision with which you can describe central tendency for any given variable depends on the variable's level of measurement. For nominal-level variables you can identify the *mode*, the most common value of the variable. For ordinal-level variables, those whose categories can be ranked, you 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 you can obtain the mode, median, and arithmetic *mean*, the sum of all values divided by the number of cases.

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 having hundreds of anonymous cases, such as the American National Election Study (NES) 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. Case Summaries lets you see firsthand how specific cases are distributed across a variable that you find especially interesting.

HOW SPSS STORES INFORMATION ABOUT VARIABLES

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,

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 politically 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. It is important, therefore, to provide SPSS with labels for each code so that the software's analytic work makes sense to the user.

INTERPRETING MEASURES OF CENTRAL TENDENCY AND VARIATION

Finding a variable's central tendency is ordinarily a straightforward exercise. Simply read the computer output and report the numbers. Describing a variable's degree of dispersion or variation, however, often requires informed judgment.[1] Here is a general rule that applies to any variable at any level of measurement: 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. In other words, the number of cases in one category equals the number of cases in every other category.

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, dictatorship, or anarchy. What would be the modal response, or the economic system preferred by most people? Democracy. Would there be a great deal of dispersion, with large numbers of people choosing the alternatives, dictatorship or anarchy? Probably not.

[1] In this chapter we use the terms *dispersion*, *variation*, and *spread* interchangeably.

In other instances, however, 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 spread out more evenly 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, many people will likely be found at the extremes of "very interested" and "not very interested." In this instance, the amount of dispersion in a variable—its degree of spread—is essential to understanding and describing it.

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. The lower the level of measure, the more limited our toolkit for describing central tendency and dispersion. 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).

DESCRIBING NOMINAL VARIABLES

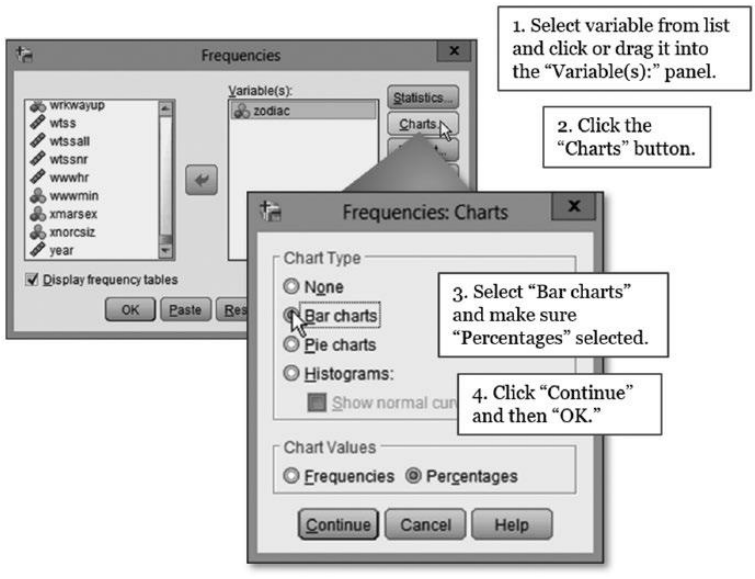
Screencast

Analyze a Variable with Dispersion

For this and the next few analyses, you will use the GSS dataset. Open the GSS dataset by double-clicking the GSS.sav file or, if you already have SPSS running, select File ▶ Open ▶ Data and locate GSS.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 for "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 GSS dataset, zodiac, which records respondents' astrological signs. Select Analyze ▶ Descriptive Statistics ▶ Frequencies. Scroll down to the bottom of the left-hand list until you find zodiac. Click zodiac into the Variable(s) panel. To the right of the Variable(s) panel, click the Charts button

FIGURE 2-1 Obtaining Frequencies and a Bar Chart (nominal variable)



(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 has produced two items of interest in the Viewer: a frequency distribution table of respondents' astrological signs 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.

Statistics

Respondents Astrological Sign

N	Valid	2777
	Missing	90

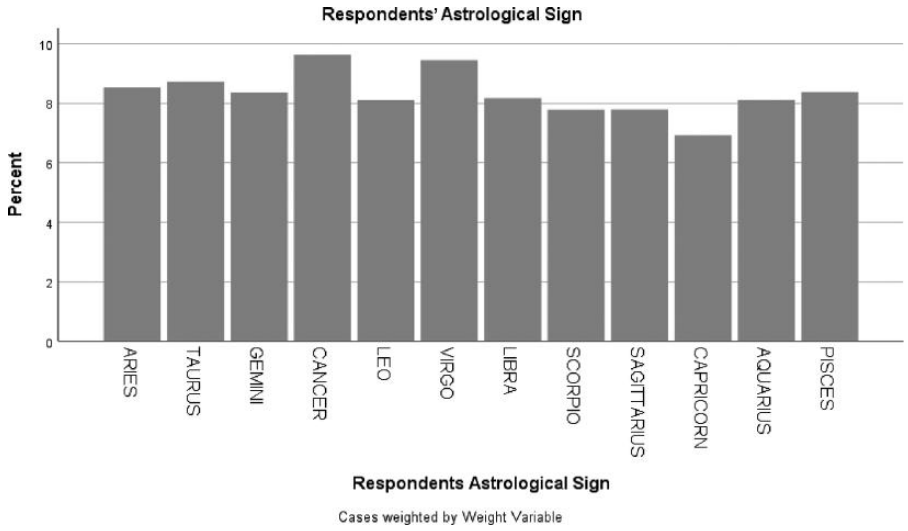
Respondents' Astrological Sign

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ARIES	237	8.3	8.5	8.5
	TAURUS	242	8.4	8.7	17.3
	GEMINI	232	8.1	8.4	25.6
	CANCER	267	9.3	9.6	35.3
	LEO	225	7.9	8.1	43.4
	VIRGO	262	9.2	9.4	52.8
	LIBRA	227	7.9	8.2	61.0
	SCORPIO	216	7.5	7.8	68.8
	SAGITTARIUS	216	7.5	7.8	76.6
	CAPRICORN	192	6.7	6.9	83.5
	AQUARIUS	225	7.9	8.1	91.6
	PISCES	233	8.1	8.4	100.0
	Total	2777	96.9	100.0	
Missing	System	90	3.1		
Total		2867	100.0		

The value labels for each astrological code appear in the leftmost column, with Aries occupying the top row of numbers and Pisces occupying the bottom row. There are four numeric columns: Frequency, Percent, Valid Percent, and Cumulative Percent. The Frequency column shows raw frequencies, the actual number of respondents having each zodiac sign. Percent is the percentage of *all* respondents, including missing cases, 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. Valid Percent tells us the percentage of non-missing responses in each value of zodiac. Finally, 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 zodiac, which cannot be ranked, the Cumulative Percent column provides no information of value.

Now consider the values in the Valid Percent column more closely. Scroll between the frequency distribution table and the bar chart, which depicts the zodiac variable in graphic form (Figure 2-2).

FIGURE 2-2 Bar Chart of a Nominal-level Variable



What is the mode, the most common astrological sign? For nominal variables, the answer to this question is (almost) always an easy call: Simply find the value with the highest percentage of responses. Virgo is the mode. When it comes to describing the central tendency of nominal-level variables like zodiac, our toolkit is limited to identifying the variable's mode.

Does the zodiac variable have little dispersion or a lot of dispersion? Again, study the Valid Percent column and the bar chart. Apply the following rule: *A variable has no dispersion if the cases are concentrated in one value of the variable; a variable has maximum dispersion if the cases are spread evenly across all values of the variable.* Are most of the cases concentrated in Virgo, or are there many cases in each sign of zodiac? Because respondents are spread out—all astrological signs are about equally represented—you would conclude that zodiac has a high level of dispersion. Looking at the bar chart of zodiac in Figure 2-3, it may be tempting to say the distribution is highest in the middle, but remember that the order of nominal-level values is essentially arbitrary; Virgo is not the middle zodiac sign, so the peak that appears to be in the middle of the bar chart is not a true feature of zodiac's dispersion.

A CLOSER LOOK: WEIGHTING THE GSS AND NES 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 (NES). Before proceeding, you need to learn about a feature of these datasets that will require special treatment throughout the book.

In raw form, the GSS and NES 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, raw survey data produce incorrect results.

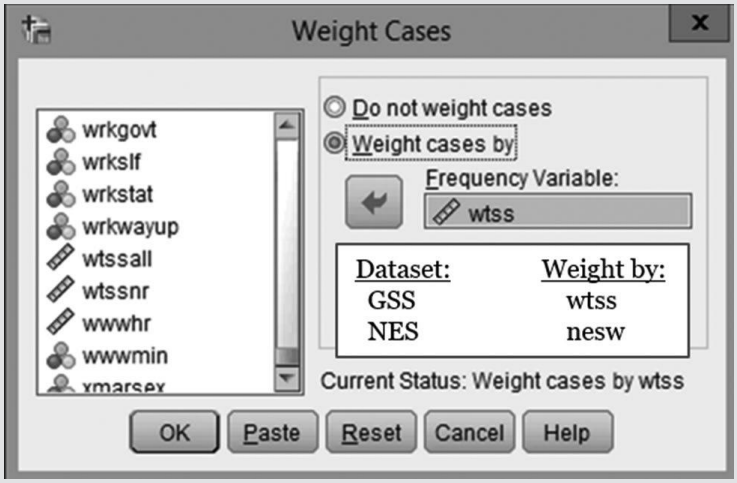
(Continued)

(Continued)

Fortunately, survey designers included the necessary corrective in the NES and GSS dataset: 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. 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 NES 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 2-3). For the NES dataset, the weight variable is *nesw*.

FIGURE 2-3 Weighting Observations in a Dataset



DESCRIBING ORDINAL VARIABLES

In this section, you will analyze and describe two ordinal-level variables in the GSS dataset, one of which has little variation and the other of which is more spread out. These ordinal-level variables examine public opinion on two social policy issues. Opinions on both questions are recorded on 5-point ordinal scales. We will use the same function we did to describe the nominal-level variable zodiac, so click the Analyze tab on the top menu bar of the Viewer and select Analyze ▶ Descriptive Statistics ▶ Frequencies.

SPSS remembers the preceding analysis, so the zodiac variable may still appear in the Variable(s) list. To begin a new descriptive analysis, click zodiac back into the left-hand list.

With your Frequencies dialog window cleared, scroll through the GSS dataset variable list until you find the variable "helppoor" and click on it so it is added to the Variable(s) list. The helppoor variable asks respondents to place themselves on a scale between 1 ("The government should take action to help poor people") and 5 ("People should help themselves"). SPSS should retain your earlier settings for Charts, so accompanying bar charts will appear in the Viewer.[2] Click OK.

[2] If you pressed the Reset button to clear zodiac 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.

SPSS produces descriptive statistics for the helppoor variable. To better understand how people responded to the helppoor question, we'll look at the variable's frequency distribution table and bar chart (Figure 2-4).

Statistics

Should Govt Improve Standard of Living?

N	Valid	1882
	Missing	985

Should Govt Improve Standard of Living?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	GOVT ACTION	339	11.8	18.0	18.0
	2	268	9.4	14.2	32.3
	AGREE WITH BOTH	807	28.2	42.9	75.1
	4	296	10.3	15.7	90.9
	PEOPLE HELP SELVES	172	6.0	9.1	100.0
	Total	1882	65.7	100.0	
Missing	System	985	34.3		
Total		2867	100.0		

Because helppoor is an ordinal variable, you can use both its mode and its median to describe central tendency. Its mode, clearly enough, is the response "Agree with both," which contains 42.9 percent of the cases. (If you get a different percentage in this category, make sure you've weighted observations properly.) What about the median? This is where the Cumulative Percent column of the frequency distribution comes into play. *The median for any ordinal (or interval) variable is the*

FIGURE 2-4 Bar Chart of an Ordinal Variable with Low Dispersion

category below which 50 percent of the cases lie. Is the first category, "Govt action," the median? No, this category contains fewer than half of the cases. How about the next higher category? No, again. The Cumulative Percent column still has not reached 50 percent. The median occurs in the "Agree with both" category (cumulative percentage, 75.1).

Now consider the question of whether helppoor has a high degree of dispersion or a low degree of dispersion. If helppoor had a high level of variation, then the percentages of respondents in each response category would be roughly equal, much like the zodiac variable that you analyzed earlier. So, roughly one-fifth of the cases would fall into each of the five response categories: 20 percent in "Gov't action," 20 percent in response category "2," 20 percent in "Agree with both," 20 percent in response category "4," and 20 percent in "People help selves." If helppoor had no dispersion, then all the cases would fall into one value. That is, one value would have 100 percent of the cases, and each of the other categories would have 0 percent. Which of these two scenarios comes closest to describing the actual distribution of respondents across the values of helppoor: the equal-percentages-in-each-category, high variation scenario, or the 100-percent-in-one-category, low variation scenario? It seems clear that helppoor is a variable with a relatively low degree of dispersion. "Agree with both," with 42.9 percent of the responses, contains nearly three times as many cases as its nearest rival ("Gov't action") and more than three times as many cases as any of the other response categories.

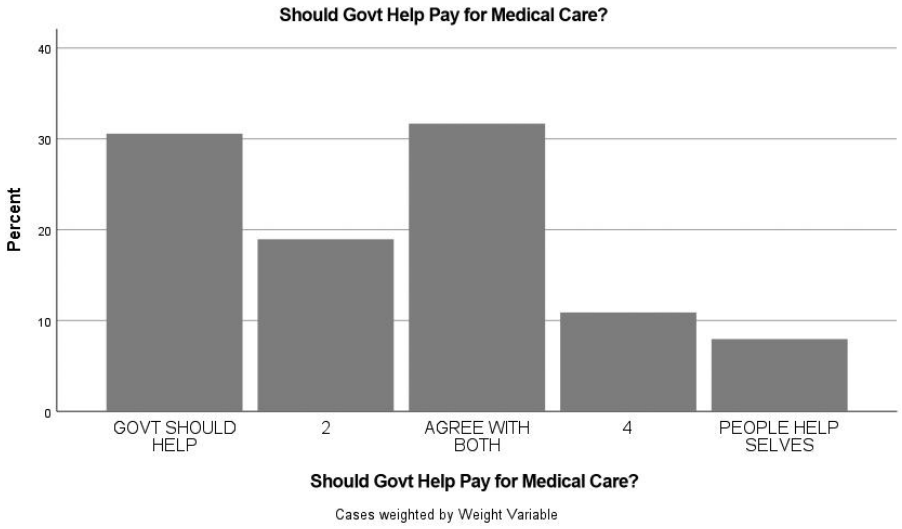
Now contrast helppoor's distribution with the distribution of the helpsick variable. The "helpsick" variable, using a similar 5-point scale, asks respondents about government responsibility or individual responsibility for medical care. You can produce a frequency distribution table and bar chart for helpsick (Figure 2-5) the same way we did to generate descriptive statistics for the helppoor variable. Review the preceding paragraphs as necessary to do this analysis.

Should Govt Help Pay for Medical Care?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	GOVT SHOULD HELP	579	20.2	30.6	30.6
	2	358	12.5	18.9	49.5
	AGREE WITH BOTH	599	20.9	31.7	81.2
	4	206	7.2	10.9	92.1
	PEOPLE HELP SELVES	150	5.2	7.9	100.0
	Total	1893	66.0	100.0	
Missing	System	974	34.0		
Total		2867	100.0		

Interestingly, helpsick has the same mode as helppoor ("Agree with both," with 31.7 percent of the cases), and the same median (again, "Agree with both," where the cumulative percentage exceeds 50.0). Yet with helppoor it seemed reasonable to say that "Agree with both" was the typical response. Would it be reasonable to say that "Agree with both" is helppoor's typical response? No, it would not. Notice that, unlike helppoor, respondents' values on helpsick are more spread out, with sizable numbers of responses falling in the first value ("Gov't action," with 30.6 percent), making it a close rival to "Agree with both" for the distinction of being the modal opinion on this issue. Clearly, the public is more divided—more widely dispersed—on the question of medical assistance than on the question of assistance to the poor.

FIGURE 2-5 Bar Chart of an Ordinal Variable with High Dispersion



A CLOSER LOOK: ANALYZING TWO VARIABLES AT ONCE

In the preceding section, we demonstrated how to describe two ordinal-level variables, helppoor and helpsick, using frequency distribution tables and bar charts. We analyzed one variable at a time to keep things simple as you learn a new skill and to draw your attention to the differences between these two variables. In the future, if you are describing the distributions of two or more variables, you may prefer to analyze multiple variables at once to work more efficiently. Watch the "Analyze Two Variables" screencast to learn how to analyze two (or more) variables at once.

Watch Screencast
Analyze Two Variables

USING THE CHART EDITOR TO MODIFY GRAPHICS

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

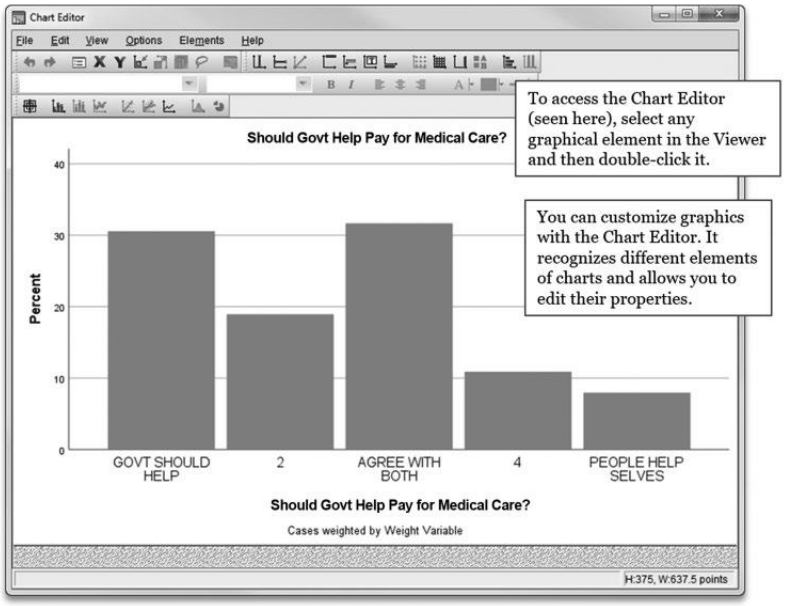
In this section, we show how you can use the Chart Editor to edit the bar chart you just created to show the distribution of public opinion on paying for medical care for sick people. We'll show how to change labels on the chart and also change the style and color of the bars. (The default style and color is rather uninspired, and it doesn't print well.)

In the Viewer, place the cursor anywhere on the bar chart and double-click. SPSS opens the Chart Editor (Figure 2-6).

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 elements as graphic, 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.

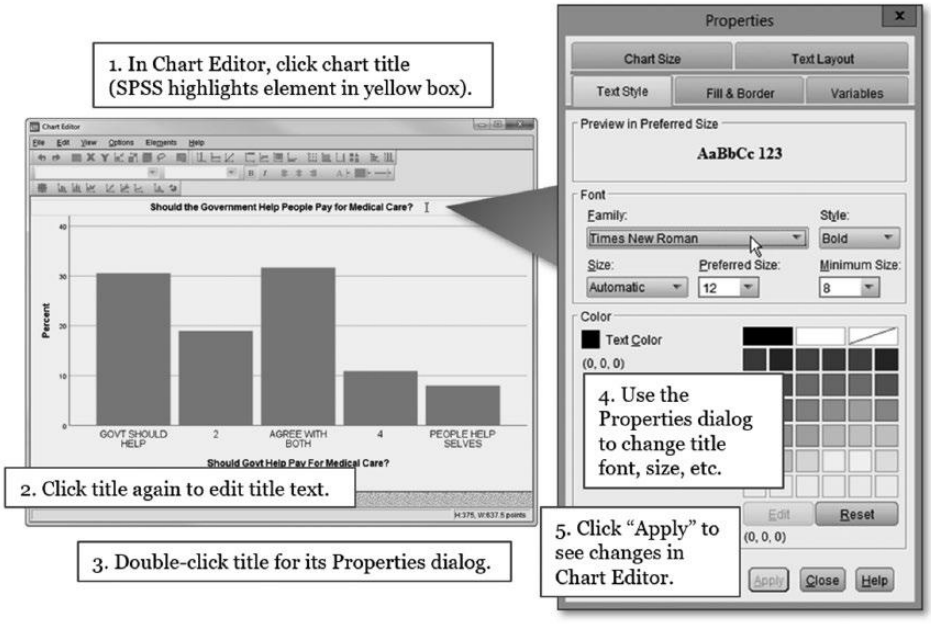
Place the cursor anywhere on the main title "Should Govt Help Pay For Medical Care?" and single-click it. SPSS selects the chart title. With the cursor still placed on the title, single-click again. SPSS moves the text into editing mode inside the chart (see the left side of Figure 2-7). The default

FIGURE 2-6 The Chart Editor



title is fine for understanding the distribution of this variable, but we could improve it to present the information to others. 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. While you're using the Chart Editor to edit chart text, you can delete the redundant *x*-axis label and provide more descriptive *x*-axis labels in place of the numbers "2" and "4."

FIGURE 2-7 Editing the Bar Chart Title



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 2-7) 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 text styles. 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 window. Alternatively, you can select the element and press the "Show Properties Window" button located near the upper-left corner of the Chart Editor window as we show in Figure 2-8. This opens the Properties window, the most powerful editing tool in the Chart Editor's arsenal. (*Special note:* 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 in a bar chart are plentiful. You can change their 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 2-8) 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 2-9).

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*. For example, age qualifies as an interval-level variable because its values impart each respondent's age in years.

Screencast

Frequency Analysis with an Interval Variable

FIGURE 2-8 Using the Properties Window to Edit Bars in a Bar Chart

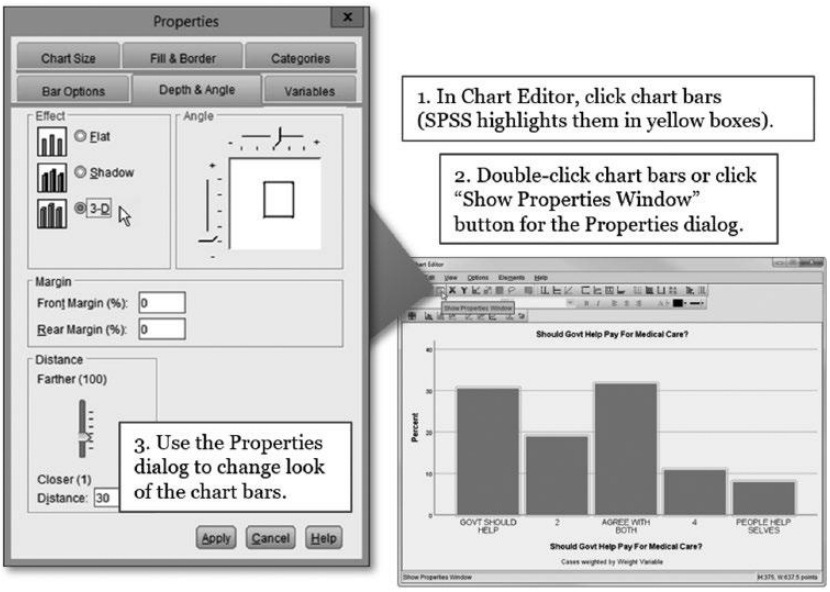
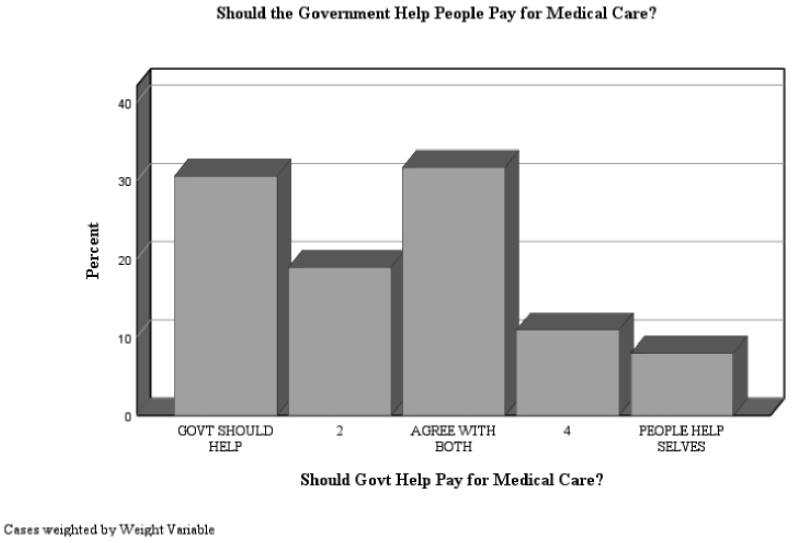


FIGURE 2-9 Edited Bar Chart in the Viewer



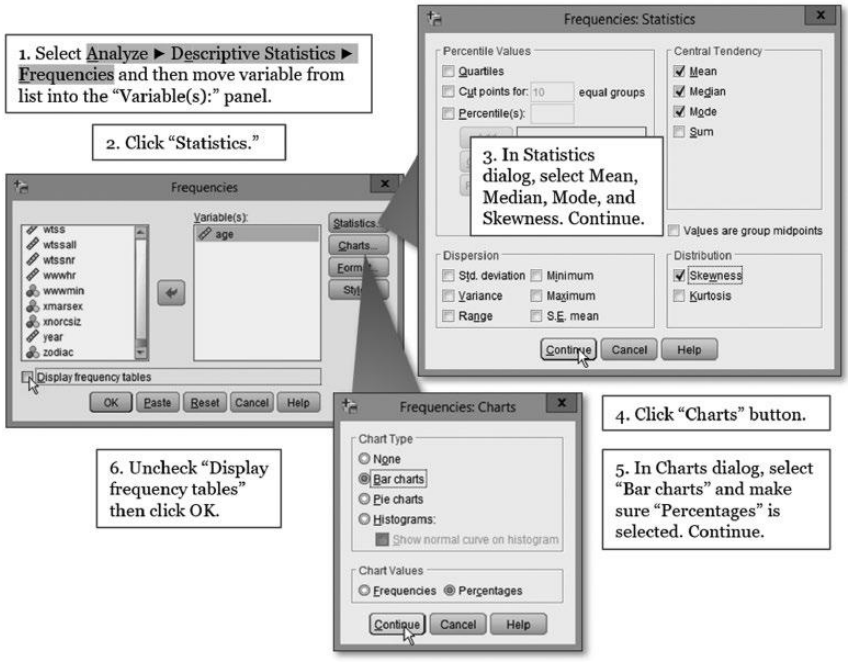
Because interval variables have the most precision, they can be described more completely than can nominal or ordinal variables; we have a relatively large toolkit available for describing variables measured at the interval level. For any interval-level variable, you can report its mode, median, and arithmetic average, or *mean*. In addition to these measures of central tendency, you can make more sophisticated judgments about variation. Specifically, you can determine if an interval-level distribution is *skewed*.

Skewness refers to the symmetry of a distribution. If a distribution is not skewed, the cases tend to cluster symmetrically around the mean of the distribution, and they taper off evenly for values above and below the mean. If a distribution is skewed, by contrast, one tail of the distribution is longer and skinnier than the other tail. Distributions in which some cases occupy the higher values of an interval variable—distributions with a skinnier right-hand tail—have a *positive skew*. By the same token, if the distribution has some cases at the extreme lower end—the distribution has a skinnier left-hand tail—then the distribution has a *negative skew*. Skewness affects the mean of the distribution. 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.*

A step-by-step analysis of a GSS variable, age, will clarify these points. Select Analyze ▶ Descriptive Statistics ▶ Frequencies. If helppoor and helpsick are still in the Variable(s) list, click them back into the left-hand list. Click age into the Variable(s) list. You may notice that the icon next to the age variable looks different from the icon next to zodiac, helppoor, and helpsick; the icons signify the variable's level of measurement.

So far, this procedure is the same as in your analysis of zodiac, helppoor, and helpsick. When you are running a frequencies analysis of an interval-level variable, however, you need to adjust the settings for the Frequency analysis to get proper results. Here's a must-do: Click the Statistics button in the Frequencies window, as shown in Figure 2-10. The Frequencies: Statistics window appears. In the Central Tendency panel, click the boxes next to Mean, Median, and Mode. In the Distribution panel, click Skewness. Click Continue, returning to the main Frequencies window. Now click the Charts button. In the Charts dialog, make sure that "Bar charts" (under Chart Type) and "Percentages" (under Chart Values) are selected. Click Continue.

FIGURE 2-10 Requesting Statistics for an Interval Variable



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," appearing at the foot of the left-hand list.[3] For interval-level variables, like age, that have many categories, a frequency distribution table can run several output pages and is not very informative. Unchecking the "Display frequency tables" box suppresses the frequency distribution. Click OK.

SPSS analyzes the age variable and outputs the requested statistics and bar chart (Figure 2-11) 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. In addition, SPSS reports values for skewness and a statistic called standard error of skewness. When a distribution is perfectly symmetrical—no skew—it has skewness equal to 0. If the distribution has a skinnier right-hand tail—positive skew—then skewness will be a positive number. A skinnier left-hand tail, logically enough, returns a negative number for skewness.

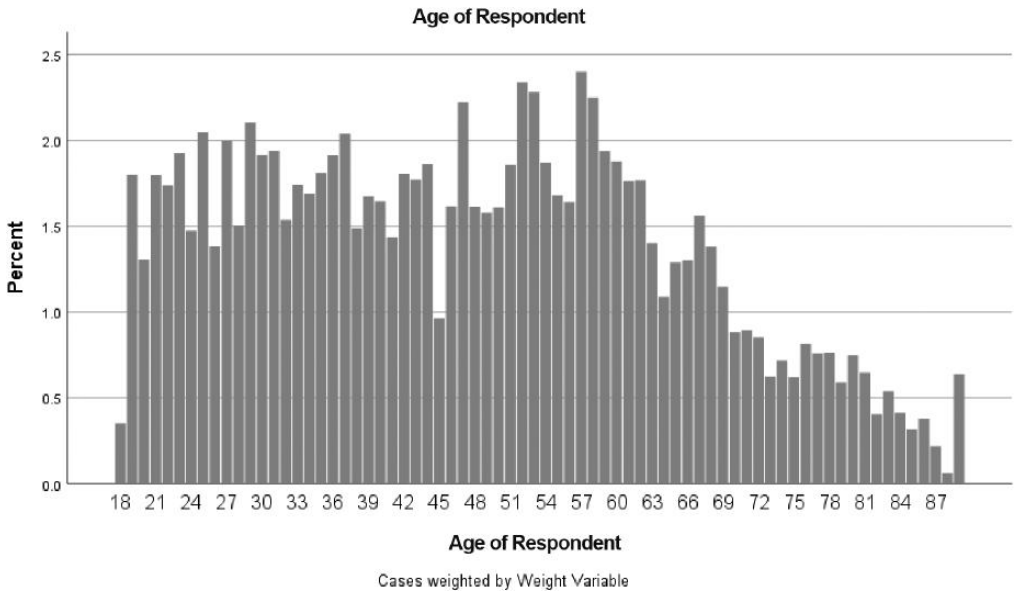
Statistics

Age of Respondent

N	Valid	2855
	Missing	12
Mean		47.56
Median		47.00
Mode		57
Skewness		.233
Std. Error of Skewness		.046

[3] A general guide: If the interval-level variable you are analyzing has 15 or fewer categories, go ahead and obtain the frequency distribution. If it has more than 15 categories, suppress the frequency distribution.

FIGURE 2-11 Bar Chart of the Interval-level Age Variable



For the age variable, the skewness statistic is positive (.233). This suggests that the distribution has a skinnier right-hand tail—a feature that is confirmed by the shape of the bar chart. Note also that the mean (47.56 years) is slightly higher than the median (47 years), a situation that often—although not always—indicates a positive skew.[4] Even so, the mean and median are less than 1 year apart. 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.[5]

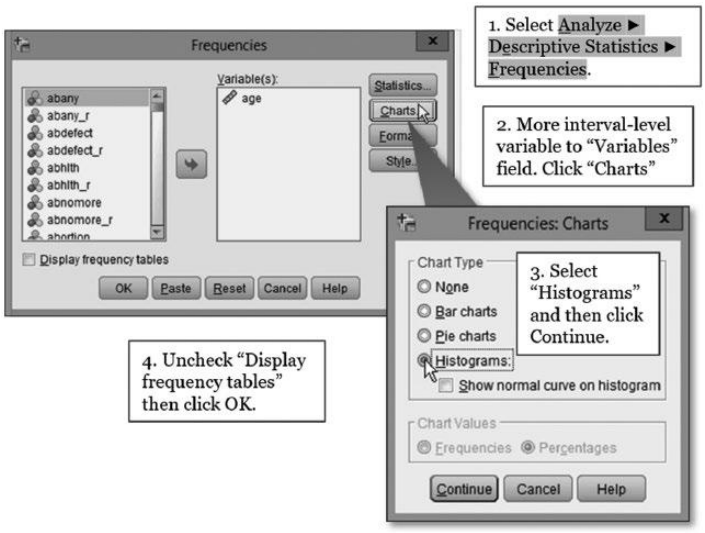
All the guided examples thus far have used bar charts for graphic support. For nominal and ordinal variables, a bar chart should always be your choice. For interval variables, however, you may want to ask SPSS to produce a histogram instead. What is the difference between a bar chart and a histogram? A bar chart displays each value of a variable and shows you the percentage (alternatively, the raw number) of cases that fall into each category. A histogram is similar to a bar chart, but instead of displaying each discrete value, it collapses categories into ranges (called bins), resulting in a compact display. Histograms are sometimes more readable and elegant than bar charts. Most of the time a histogram will work just as well as a bar chart in summarizing an interval-level variable. For interval variables with many unique values, a histogram is the graphic of choice. (Remember: For nominal or ordinal variables, you always want a bar chart.)

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, Mode, and Skewness. Click Continue. Click Charts, and then select the Histograms radio button in Chart Type. Click Continue. For this analysis, we do not

[4] 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." See http://www.amstat.org/publications/jse/v13n2/vonhippel.html.

[5] 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.

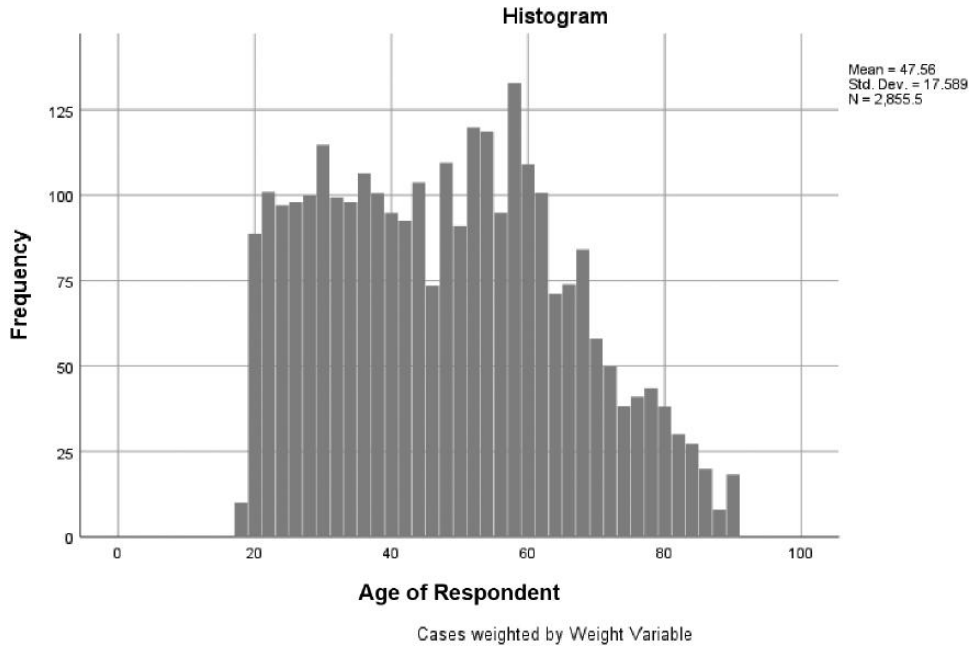
FIGURE 2-12 Creating a Histogram of an Interval Variable



need a frequency distribution table. In the Frequencies window, uncheck the "Display frequency tables" box. (Refer to Figure 2-12.) Click OK.[6]

This is a bare-bones run. SPSS reports its obligatory count of valid and missing cases, plus a histogram for age (Figure 2-13). 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

FIGURE 2-13 Histogram of the Interval-level Age Variable



[6] Alternatively, you can select Graphs ▶ Legacy Dialogs ▶ Histogram to produce a histogram. Move age into the Variable field and click OK. This procedure should produce the same output as the method described in the text.

years of age rather than 1 year of age. Now scroll up the Viewer to the bar chart of age, which you produced in the preceding analysis. Notice that the histogram has smoothed out the choppiness of the bar chart, though it still captures the essential qualities of the age variable.

Statistics

Age of Respondent

N	Valid	2855
	Missing	12

OBTAINING CASE-LEVEL INFORMATION WITH CASE SUMMARIES

Screencast
Analyze Case Summaries

When you analyze a large survey dataset, as you have just done, you generally are not interested in how respondent X or respondent Y answered a particular question; they're just some random people who happened to participate in the survey. Rather, you want to know how the entire sample of respondents distributed themselves across the response categories of a variable (for this purpose, their randomness is vitally important). Sometimes, however, you gather data on particular cases because the cases are themselves inherently important.

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 you are interested in identifying states that have the most/fewest laws restricting access to abortions. To begin this guided example, close the GSS dataset and open the States dataset. The States dataset contains a variable named abortlaw17. This variable records the number of legal restrictions on abortion access in each state in 2017 (out of 14 possible restrictions). Exactly which states impose the most restrictions? Which states impose the fewest restrictions? Where does your state fall on the list? Case Summaries can quickly answer questions like these. SPSS will sort states based on a "grouping variable" (in this example, abortlaw17) and then produce a report telling you which states are in each group.

With the States dataset open, click Analyze ▶ Reports ▶ Case Summaries.

To conduct the desired analysis, you need to do three things in the Summarize Cases window (see Figure 2-14):

1. Click the variable containing the cases' identities into the Variables window. In the States dataset, this variable is named state, which is simply the name of each state.
2. Click the variable you are interested in analyzing, abortlaw17, into the Grouping Variable(s) window.
3. Uncheck the "Limit cases to first. . ." option. This won't affect the analysis of state abortion laws because there are fewer than 100 states, but if this box is left checked when you analyze the World dataset, SPSS will limit the analysis to the first 100 countries and produce an incomplete analysis.

Click OK and consider the output. SPSS sorts the cases on the grouping variable, abortlaw17, and tells us which state is associated with each value of abortlaw17. For example, Vermont, with 0 legal restrictions on abortion access, is the state with the fewest restrictions on access to abortions. Which states impose the most restrictions? Scroll to the bottom of the tabular output. With 13 restrictions, Kansas and Oklahoma are tied for imposing the most restrictions.