USING EXCEL FOR DATA ANALYSIS

LEARNING OBJECTIVES

In this chapter you will learn to:

- Navigate Excel workbooks
- Extend Excel's data analysis capabilities with add-ins
- Generate results with formulas, functions, and tools
- Define and use names in Excel
- Distinguish data from analysis and keep them separate
- Print and save worksheets
- Identify academic-style tables and figures
- Get additional help with Excel functions and tools

EXCEL USAGE

- *Excel Companion* files: Debate Experiment Workbook, Presidential Elections Workbook
- File ▶ Open, Print, Save
- File ▶ Options ▶ Adds-ins
- Data Analysis ToolPak: Descriptive Statistics (as example)
- Named functions: AVERAGEIF, AVERAGEIFS (as examples), FORMULATEXT
- Formulas ▶ Show Formulas, Define Name, Name Manager
- Help ▶ Help
- Briefly mentioned: Formulas ▶ Insert Function; Insert ▶ Text ▶ Table, Text Box, AVEDEV, DEVSQ, COUNTIF

In this chapter, we take a quick tour of the Excel program. Since most students have used Excel before, a lengthy introduction is unnecessary. We assume that you have some familiarity with Windows-based programs—that is, what it means to copy and paste, right-click, move the cursor, drag an object onscreen, and open, close, and save files.[1] We focus on using Excel for data analysis, a topic rich enough to fill this book.

For maximum benefit, practice the steps and procedures we discuss here on your own computer. We've based most of this book on Excel 365 on a Windows machine. This should be the Excel version most widely available to students. That said, we've tried to avoid tailoring the book too closely to one version of Excel because some aspects of the program are revised and redesigned with new versions. It is quite possible things may look a little different on your computer.

You can run Excel on a Macintosh operating system. There aren't major differences in the program on a Mac computer, but there are some differences in how Mac users operate programs generally. We hope that when we refer to the "Ctrl" key or right-clicking, you can translate those Windows-based instructions into their Mac OS equivalents.

READING IN *ESSENTIALS*

We cover the definition and measurement of political science concepts in Chapter 1 of the sixth edition of *The Essentials of Political Analysis*, pp. 1–33.

1.1 WORKBOOKS AND WORKSHEETS

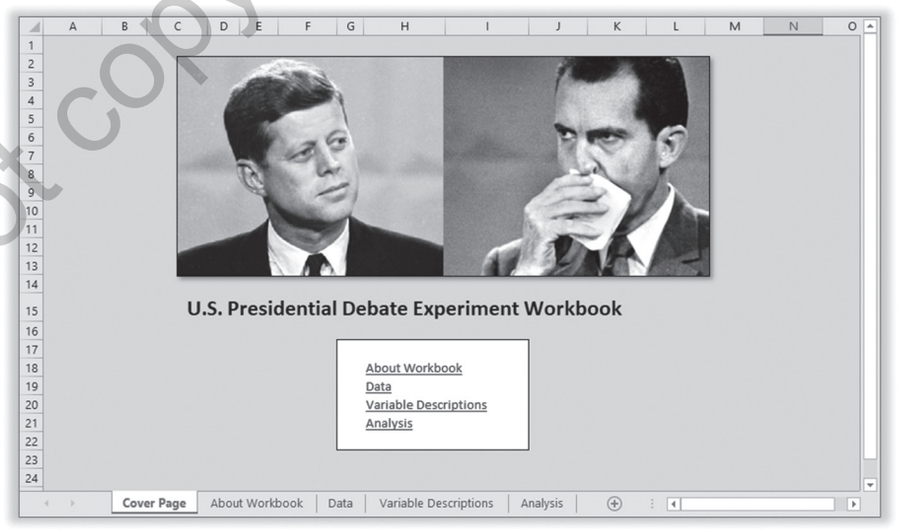
Workbook: the file that contains one or more Excel worksheets; usually ends in .xls or .xlsx.

Worksheet: a spreadsheet organized into rows and columns; may reference functions and other worksheets.

Workbooks are Excel's way of organizing and storing information. An Excel **workbook** contains one or more **worksheets**, which most people call spreadsheets. For this *Excel Companion to Political Analysis*, we've created several Excel workbooks that contain multiple worksheets (see "Getting Started" and the Appendix for details).

Figure 1-1 shows the cover page of our Debate Experiment Workbook. To open this workbook, locate the "Debate Experiment Workbook.xlsx" file in the folder where you downloaded it and double-click it. If you already have Excel running, select File ▸ Open and find the file on your computer.

FIGURE 1-1 ⬢ Excel Companion's Debate Experiment Workbook

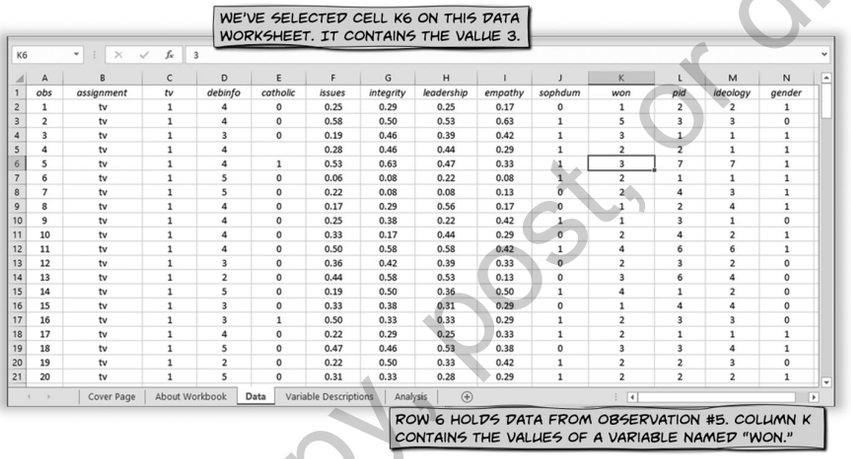


[1] If you are totally new to Excel, there are many resources available to introduce you to the program. Check out the computer reference section of a local library or your campus bookstore.

You can navigate this workbook's worksheets by clicking text links on the cover page or the tabs listed at the bottom of the window. As you can see, the Debate Experiment Workbook has four worksheets following the cover page: a page about the workbook, the data, variable descriptions, and some analysis of the data (which we discuss in Sections 6.2 and 9.5). Workbooks don't need cover pages, basic graphic design, or hyperlinks, but Excel allows you to organize and store information in an attractive and user-friendly manner. When you're using Excel for data analysis, we encourage you to make your work clear and user-friendly, especially since you'll be the one using your workbooks!

Figure 1-2 shows the workbook's "Data" worksheet. When researchers talk about datasets, this is what they have in mind. By convention, each row represents an observation and each column provides a different piece of information about the observations. These data were collected in a political science experiment conducted with college students. Each row contains information related to a student who participated in the experiment. Some of the columns contain demographic information about the student participants; other columns show students' impressions of the 1960 Presidential Debate.

FIGURE 1-2 ⬢ Data Worksheet of the Debate Experiment Workbook



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	*obs*	*assignment*	*tv*	*debinfo*	*catholic*	*issues*	*integrity*	*leadership*	*empathy*	*sophdum*	*won*	*pid*	*ideology*	*gender*
2	1	tv	1	4	0	0.25	0.29	0.25	0.17	0	1	2	2	1
3	2	tv	1	4	0	0.58	0.50	0.53	0.63	1	5	3	3	0
4	3	tv	1	3	0	0.19	0.46	0.39	0.42	1	3	1	1	1
5	4	tv	1	4		0.28	0.46	0.44	0.29	1	2	2	1	1
6	5	tv	1	4	1	0.53	0.63	0.47	0.33	1	3	7	7	1
7	6	tv	1	5	0	0.06	0.08	0.22	0.08	1	2	1	1	1
8	7	tv	1	5	0	0.22	0.08	0.08	0.13	0	2	4	3	1
9	8	tv	1	4	0	0.17	0.29	0.56	0.17	0	1	2	4	1
10	9	tv	1	4	0	0.25	0.38	0.22	0.42	1	1	3	1	0
11	10	tv	1	4	0	0.33	0.17	0.44	0.29	0	2	4	2	1
12	11	tv	1	4	0	0.50	0.58	0.58	0.42	1	4	6	6	1
13	12	tv	1	3	0	0.36	0.42	0.39	0.33	0	2	3	2	0
14	13	tv	1	2	0	0.44	0.58	0.53	0.13	0	3	6	4	0
15	14	tv	1	5	0	0.19	0.50	0.36	0.50	1	4	1	2	0
16	15	tv	1	3	0	0.33	0.38	0.31	0.29	0	1	4	4	0
17	16	tv	1	3	1	0.50	0.33	0.33	0.29	1	2	3	3	0
18	17	tv	1	4	0	0.22	0.29	0.25	0.33	1	2	1	1	1
19	18	tv	1	5	0	0.47	0.46	0.53	0.38	0	3	3	4	1
20	19	tv	1	2	0	0.22	0.50	0.33	0.42	1	2	2	3	0
21	20	tv	1	5	0	0.31	0.33	0.28	0.29	1	2	2	2	1

The variable names, brief yet descriptive, appear in row 1 (see Figure 1-2). Numbers in the *obs* column record each subject's unique identification number. Since row 1 holds the variable names, information related to subject number 1 is in row 2.

The worksheet's row numbers are listed along the left-side margin. The worksheet's columns are identified by letters A, B, C, D, and so forth that appear along the top margin, just above the variable names.

The column letters and row numbers together create **cell references**, unique locations for every cell in a worksheet. In Figure 1-2, we've clicked on cell K6. The active cell is identified in the upper-left corner of the window (the name box) and its content (the number 3) is displayed in the formula bar as well as in the cell itself. Cell K6 contains subject number 5's impression of who won the debate. All of subject 5's information is in the cell range A6:N6. All of the subjects' responses on who won the debate are in the cell range K1:K172. A range of cells, sometimes called a **field**, is identified by its upper-left and bottom-right cells.

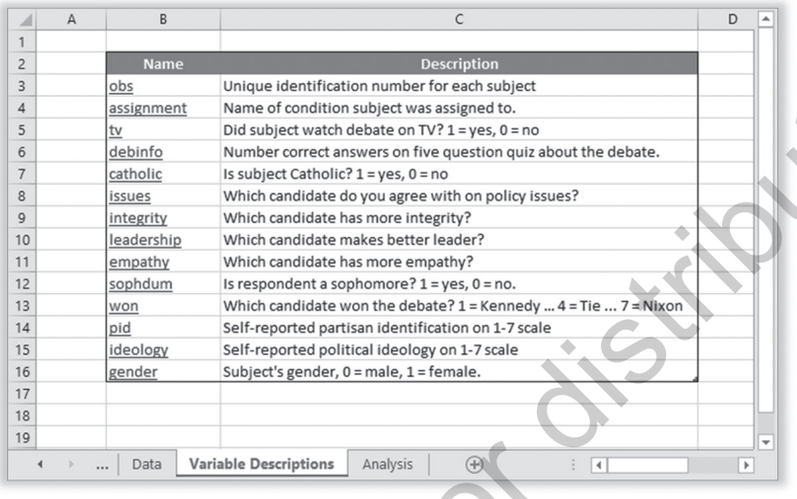
Cell references: combinations of letters and numbers that identify cell(s) on a worksheet; can be absolute or relative.

Field: a set of adjacent cells; identified by upper-left and bottom-right cell references.

Datasets aren't always useful without additional information to put the data in context. Storing information as numbers is efficient, but it's not immediately clear what these numbers mean. To better understand these data, you need to see how the variables are coded. To see how the Debate Experiment variables are coded, click the Variable Descriptions tab (Figure 1-3). This worksheet

clarifies the meaning and measurement of each variable in the dataset. This bare-bones description of variables in the dataset lists their names and numeric coding. The information on this worksheet is commonly called a dataset's codebook.

FIGURE 1-3 ⬢ Variable Descriptions in the Debate Experiment Workbook



	Name	Description
3	obs	Unique identification number for each subject
4	assignment	Name of condition subject was assigned to.
5	tv	Did subject watch debate on TV? 1 = yes, 0 = no
6	debinfo	Number correct answers on five question quiz about the debate.
7	catholic	Is subject Catholic? 1 = yes, 0 = no
8	issues	Which candidate do you agree with on policy issues?
9	integrity	Which candidate has more integrity?
10	leadership	Which candidate makes better leader?
11	empathy	Which candidate has more empathy?
12	sophdum	Is respondent a sophomore? 1 = yes, 0 = no.
13	won	Which candidate won the debate? 1 = Kennedy ... 4 = Tie ... 7 = Nixon
14	pid	Self-reported partisan identification on 1-7 scale
15	ideology	Self-reported political ideology on 1-7 scale
16	gender	Subject's gender, 0 = male, 1 = female.

Variable names, which run across the top row of the Data worksheet, serve as labels when Excel does analysis. You want variable names to be concise and descriptive. Researchers often write variable names without spaces between words by using underscores to connect words.

Finally, a word about missing values. 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 a survey, some questions do not apply to everyone, or the response may not be clear. In coding the data, researchers may give special numeric codes to missing values, enter "N/A," or simply leave those cells blank. It's important to be aware of how any missing values are encoded. When you use an existing variable to create a new variable, Excel may not automatically transfer missing values on the existing variable to missing values on the new variable. Some data analysis tools cannot handle missing values. Later in this volume, we discuss how to handle missing values.

1.2 ACTIVATING DATA ANALYSIS ADD-INS

Add-ins: add specialized features to Excel beyond the basic functions; similar to apps you download to add features to a device.

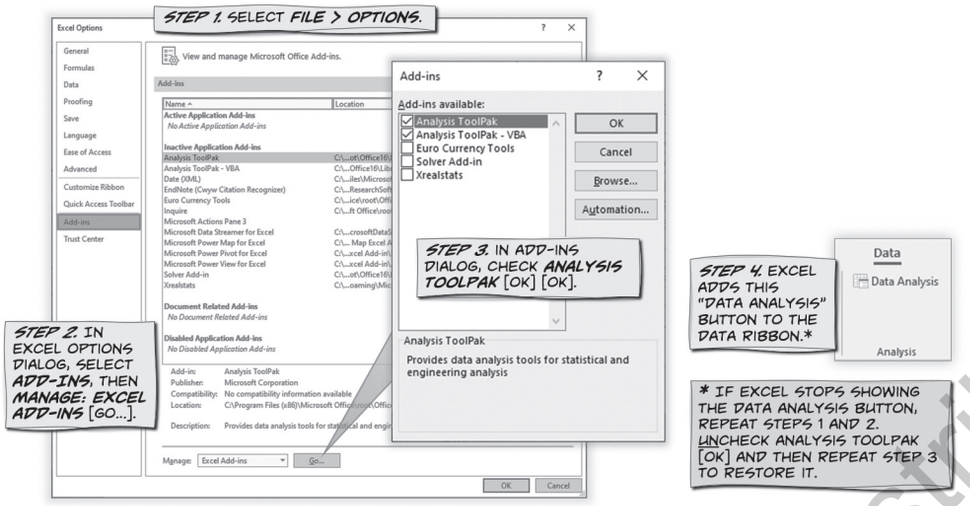
To use Excel for political analysis, you need to activate some **add-ins** that expand the program's functionality. To active Excel's Data Analysis ToolPak, which we use extensively in this book, select File ▶ Options ▶ Add-ins (see Figure 1-4). Below the list of active and inactive add-ins, select manage "Excel Add-ins" and click "Go . . ." In the Add-ins dialog, check Analysis ToolPak and Analysis ToolPak – VBA, and then click OK. After Excel activates these add-ins, you'll find a new button for Data Analysis in the ribbon under Data.

To further enlarge Excel's capacity for data analysis, we recommend activating Excel's Solver add-in and downloading the Real Statistics add-in. We use these Excel add-ins to conduct logistic regression analysis in Chapter 14. To activate the Solver add-in, which implements a variety of optimization algorithms, follow the steps illustrated in Figure 1-4 for activating the Data Analysis ToolPak, except also check the Solver add-in. You'll then find a new button for Solver in Excel's **ribbon** in the Data section.

Ribbon: the row of tabs and icon buttons that spans the top of Excel's graphical user interface.

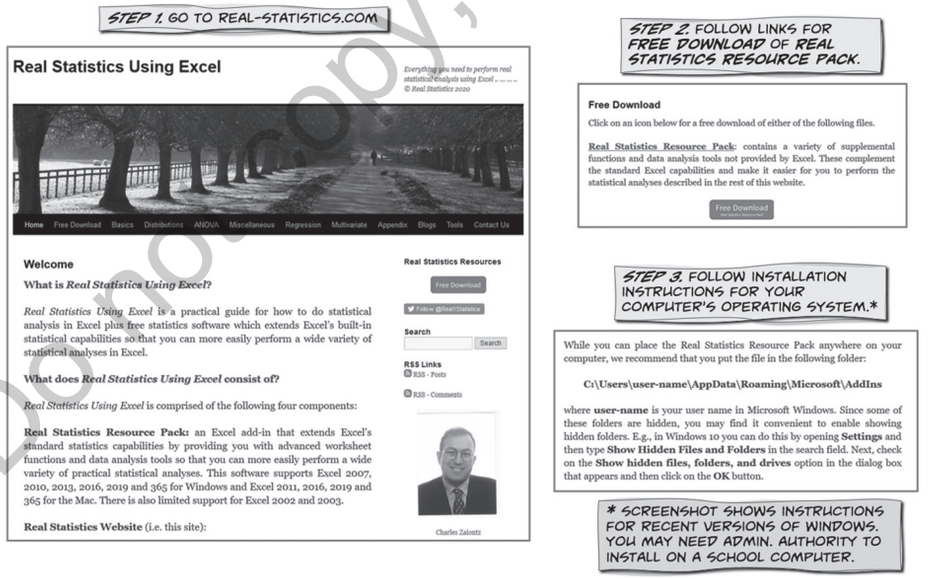
You can obtain a free copy of the Real Statistics add-in from the Real Statistics Using Excel website (http://www.real-statistics.com). Go to the website, click the "Free Download" link, then click "Real Statistics Resource Pack," and follow the link to download the Resource Park for the version of Excel

FIGURE 1-4 ⬢ Activating Excel Add-Ins for Data Analysis



you are running (see Figure 1-5). The Resource Pack is available for Mac OS and older versions of Excel. Follow the installation instructions.[2] You need to activate the Solver add-in before installing the Real Statistics add-in. Move the downloaded file to the appropriate location on your computer. After you have the Real Statistics add-in moved to the correct location on your machine, return to the Excel program. Follow the steps illustrated in Figure 1-4, except now add a check next to "Xrealstats" and click OK. To access this add-in, you will select Add-ins ▸ Real Statistics ▸ Data Analysis Tools. You may need to restart Excel to see these new features.[3]

FIGURE 1-5 ⬢ Downloading and Installing the Real Statistics Add-In



[2]The Real Statistics Resource Pack is available at http://www.real-statistics.com/free-download/real-statistics-resource-pack/.

[3]If you encounter problems installing and using the Real Statistics add-in, see the troubleshooting tips on its website, including what to do if your machine prompts you for a password or it reports a compilation error when you use the add-in.

We've noticed these add-ins sometimes disappear from Excel's ribbon even after we've activated them. If this happens to you while you're working through examples and exercises, select File ▶ Options ▶ Add-ins, deselect the add-ins, click OK, and then repeat the steps in Figure 1-4 and/or Figure 1-5 to restore the add-ins to the ribbon.

1.3 FORMULAS, FUNCTIONS, AND TOOLS

Once data are saved and organized in an Excel workbook, you can analyze data with Excel formulas, functions, and data analysis tools. Before getting into the technical details of specific procedures, it helps to understand how formulas, functions, and tools work generally.

Formulas: worksheet statements that start with = and output values to worksheets, usually using mathematical operators.

You can use **formulas** to calculate cell values. To calculate a value using a mathematical formula, type an equals sign (=) in a cell and then type the expression as you would enter it into a calculator. See Table 1-1 for list of mathematical operators in Excel. You can use real numbers in formulas, such as =10+2, as well as cell references, like =10+B2. Excel will calculate values according to the standard order of operations; use parentheses to control the order of operations and make your expressions clearer.

TABLE 1-1 ⬢ Mathematical Operators in Excel

Operator	Purpose
+	Addition
–	Subtraction
/	Division
*	Multiplication
^	Exponents

Functions: routines called by name to perform computations and generate output.

Excel **functions** generate cell values based on a pre-programmed routines. Every function has a unique name that's written in all capital letters. A function's name usually tells you what it does, but it's not always clear. When you're working with a new function, the main things to know are its name, what it's used for, its arguments, and its output. A function's arguments are the inputs it needs to do its job. If a function is going to count something, you need to tell it which cells to count (required argument) and you may want to tell it to only count numbers, ignore duplicates, and so on (optional arguments).

General syntax: a generic statement of an Excel function that lists all its arguments.

The **general syntax** of an Excel function looks something like this:

=FUNCTION(arg1,arg2,arg3)

FUNCTION is a generic placeholder for an Excel function's name. Arguments are enclosed in parentheses after the function's name, separated by commas. A function can have any number of arguments. You must complete the required arguments in order and may add the function's optional arguments, separated by commas, enclosed in parentheses.

There are several ways to insert functions into worksheets:

- Click on a cell and type an equals sign (=) followed by the function name and its arguments. If you start typing a function call, Excel has some nice autocomplete and live tip features to help you use the function correctly.
- Click the f_x button and select a function using the Insert Function dialog.
- Select Formulas ▶ Insert Function and then choose a function to insert using the Insert Function dialog.
- Select Formulas and then an option from the Function Library.

These four ways to insert functions are simply different ways of doing the same thing. You can try them all to see what works best for you. If you're using a new function, Excel's graphical user interface (GUI) can help break it down for you; once you're familiar with it, you might find the keyboard route is faster.

When you insert an Excel function, the worksheet stores the function and its arguments in the cell and displays the function's output. You see the function's result and not the function itself, unless the cell is active. If you use insert functions by completing function **dialogs**, the equivalent function call will be inputted into a worksheet cell for you, which can help you learn how to use new functions.

Dialogs: Excel's graphical user interfaces for functions and tools.

Excel's **data analysis tools** are implemented through the add-ins we discussed in the previous section. These tools execute multiple functions and statistical methods from one dialog, which can save you a lot of time. Where possible, we'll show how to do analysis using data analysis tools, but you will need to learn how to use both tools and functions to do political analysis with Excel.

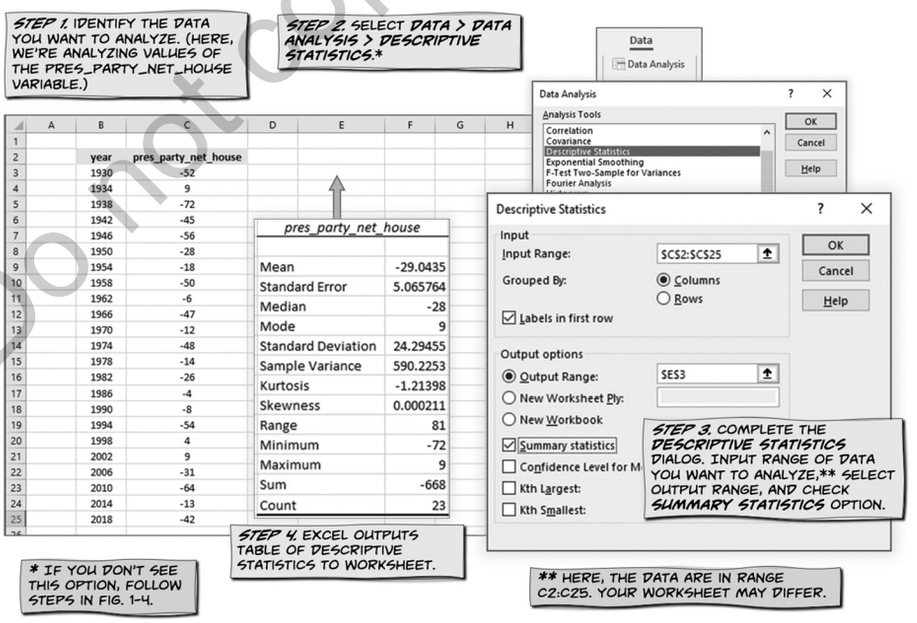
Data Analysis tools: a special set of tools for data analysis that are part of the Data Analysis ToolPak.

Functions and tools serve similar purposes, but there are important differences. You can't execute statistical procedures using the Data Analysis ToolPak, Solver, or Real Statistics by typing function calls into cells. Instead, you use specialized dialog windows. Like functions, tools output results to worksheets, but they typically output "just the numbers" and not the formulas that produced those numbers. Moreover, data analysis tools output tables of results that occupy multiple columns and rows rather than the single cell of output generated by most Excel functions.

If you update the data a function operates on, the results will update automatically. In contrast, if you update the data a tool operates on, the results will not update automatically; you'll need to run the tool again. When you execute a procedure using the GUI, Excel temporarily stores the information you inputted so you can return to the same window and adjust your selections. This is particularly useful when you're executing complex methods for the first time, making graphics, or performing the same operation repeatedly on different variables. If you want to repeat the analysis in another session, you need to fill out the tool dialogs again, a minor inconvenience since Excel tools are easy to use once your data are in order and you know how to use the tool dialogs.

An example should help drive this general discussion of Excel functions and tools home. We'll examine descriptive statistics in detail in Chapter 2, so for now focus on the general mechanics of Excel functions and tools. Figure 1-6 shows a worksheet that lists the number of House seats the U.S. president's party has gained or lost during midterm elections since 1930 (cells C3:C25). We can use the Descriptive Statistics tool in the Data Analysis ToolPak to analyze these values. Select

FIGURE 1-6 ⬢ Using a Tool in the Data Analysis ToolPak

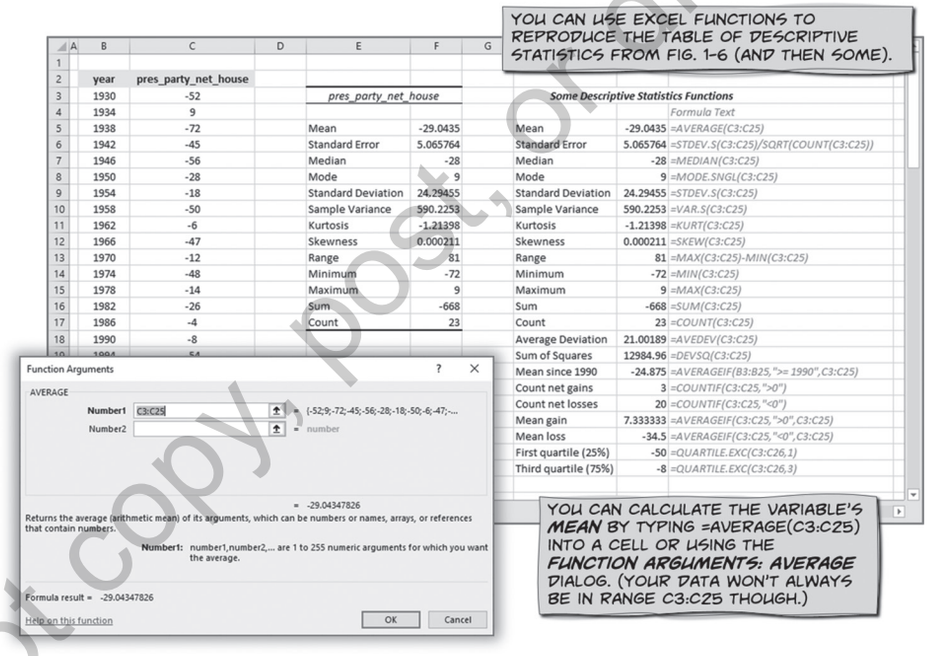


Data ▶ Data Analysis ▶ Descriptive Statistics, complete the Descriptive Statistics dialog, and click OK. The tool outputs a table of descriptive statistics to the worksheet. This is a handy tool to generate a set of common descriptive statistics.

We can use Excel's built-in functions to replicate all the statistics generated by Excel's data analysis tool, and then some. For example, Excel's AVERAGE function will calculate the mean of the numbers in cells C3:C25 for us. Column J in Figure 1-7 shows the function text used to generate the values seen in Column I.[4] We can describe data using Excel functions like AVEDEV and DEVSQ that aren't incorporated into the descriptive statistics tools. You can see that the only required argument for most of these functions is the cell range to evaluate, C3:C25. Some statistics, like standard error and range, are computed by using function results in formulas.

If Excel starts inserting or modifying cell references when you mean to move the cursor backward or forward, press the Esc key to escape from the formula. If you make a mistake, keep in mind that you can undo most mistakes by pressing Ctrl-Z.

FIGURE 1-7 ⬢ Using Excel Functions to Calculate Statistics

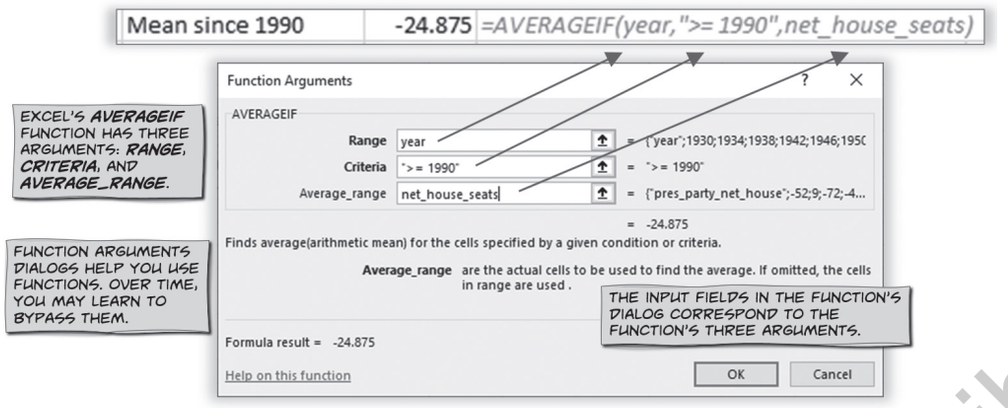


Arguments: values supplied to a function inside parentheses, separated by commas; they're settings for the function.

In Figure 1-7, you can see some Excel functions with multiple **arguments**. For example, Excel's AVERAGEIF function has three arguments. It follows the generic function syntax described earlier. In this example, we used AVERAGEIF to compute the mean seat change in midterms since 1990 with the following arguments: (1) the range of values to evaluate, which in this case are the election

[4]We use Excel's FORMULATEXT function to show the underlying functions on many of the worksheets discussed in this book. FORMULATEXT helps make worksheet functions transparent and easier to manage.

FIGURE 1-8 ⬢ Executing a Function with Multiple Arguments



years in B3:25; (2) the criteria for inclusion, in which the year value must be greater than or equal to 1990; and (3) the values to be averaged, which are in C3:C25. Notice that the three argument fields in the AVERAGEIF dialog (Figure 1-8) correspond to the three arguments in the function executed on the worksheet.

We can use the AVERAGEIF function again, with different argument values, to calculate the mean seat loss (−34.5 seats) and the mean seat gain (+7.33 seats). Data analysis tools will often suffice but Excel functions can help us drill down and extract more detailed information or want one specific statistic rather than an extensive table of statistics.

1.4 DEFINING AND USING NAMES

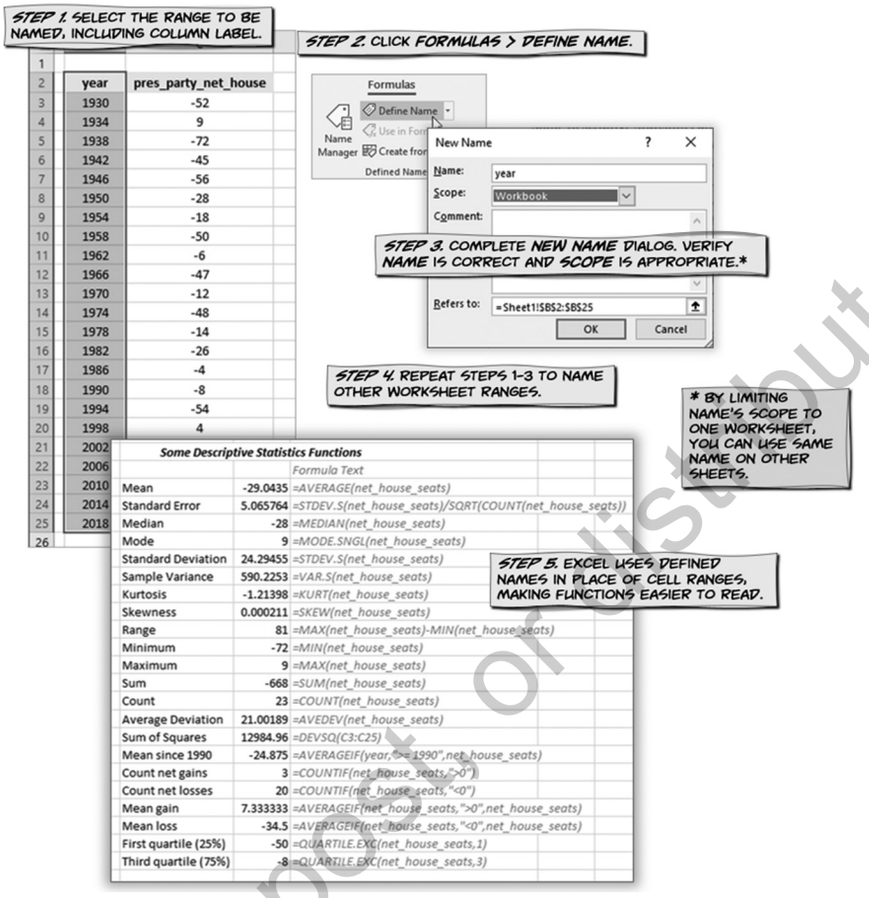
In the previous section we applied a battery of Excel functions to values stored in cells C3:C25 on a worksheet (see formula text in Figure 1-7). Referencing cells C3:C25 gets the job done, but it's hard to decipher a statement like =AVERAGEIF(B3:B25,">= 1990",C3:C25) without working backward to understand what's in the cell ranges. Fortunately, Excel allows us to name cell ranges and use these names in functions.

Names: you can name cells or cell ranges to make them easier to remember and read in formulas.

Defining and using **names** makes it easier to analyze data with Excel. It's a lot easier to remember a variable by name than by its column letter, especially when you move worksheet elements around. We've named variables in the workbooks that accompany this book and encourage you to define and use names in your own political analysis.

To illustrate, we'll define and use names in our analysis of the number of House seats the U.S. president's party has gained or lost during midterm elections. To name a range of cells, highlight the range, along with any labels you have made, and select Formulas ▶ Define Name. For this analysis, we'll name the years in cells B3:B25 and the net seat changes in C3:C25 (see Figure 1-9). We highlight cells B2:B25, select Formulas ▶ Define Name, check the automated entries in the New Name dialog, and click OK. We repeat the process for cells C2:C25. We don't have to use the column label as the defined name; here, we'll change it to net_house_seats. Defined names must start with a letter or underscore, and they can't include white spaces or conflict with any existing name in the workbook. Use descriptive names so their meaning is clear to you. Arbitrary names, however clever they may be, defeat the purpose of defining and using names, so avoid them.

FIGURE 1-9 ⬢ Defining and Using Named Ranges in Excel Functions

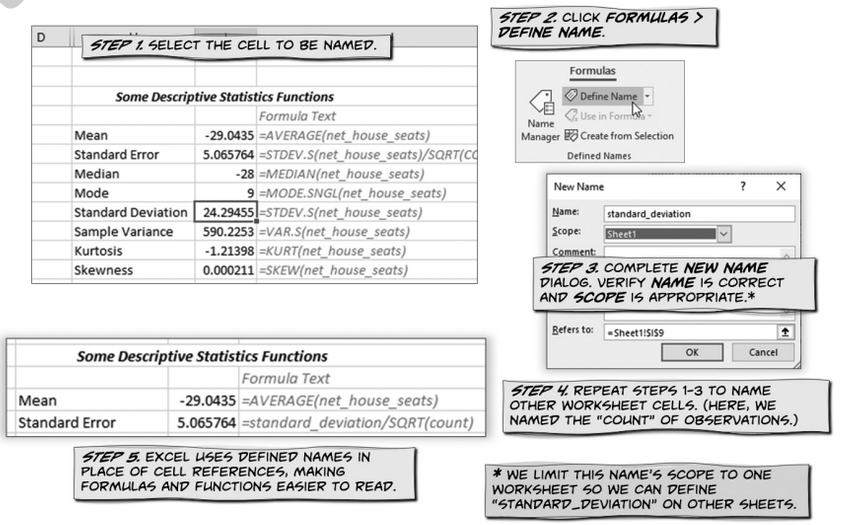


year	pres_party_net_house
1930	-52
1934	9
1938	-72
1942	-45
1946	-56
1950	-28
1954	-18
1958	-50
1962	-6
1966	-47
1970	-12
1974	-48
1978	-14
1982	-26
1986	-4
1990	-8
1994	-54
1998	4
2002	
2006	
2010	
2014	
2018	

Some Descriptive Statistics Functions		*Formula Text*
Mean	-29.0435	=AVERAGE(net_house_seats)
Standard Error	5.065764	=STDEV.S(net_house_seats)/SQRT(COUNT(net_house_seats))
Median	-28	=MEDIAN(net_house_seats)
Mode	9	=MODE.SNGL(net_house_seats)
Standard Deviation	24.29455	=STDEV.S(net_house_seats)
Sample Variance	590.2253	=VAR.S(net_house_seats)
Kurtosis	-1.21398	=KURT(net_house_seats)
Skewness	0.000211	=SKEW(net_house_seats)
Range	81	=MAX(net_house_seats)-MIN(net_house_seats)
Minimum	-72	=MIN(net_house_seats)
Maximum	9	=MAX(net_house_seats)
Sum	-668	=SUM(net_house_seats)
Count	23	=COUNT(net_house_seats)
Average Deviation	21.00189	=AVEDEV(net_house_seats)
Sum of Squares	12984.96	=DEVSQ(C3:C25)
Mean since 1990	-24.875	=AVERAGEIF(year,">= 1990",net_house_seats)
Count net gains	3	=COUNTIF(net_house_seats,">0")
Count net losses	20	=COUNTIF(net_house_seats,"<0")
Mean gain	7.333333	=AVERAGEIF(net_house_seats,">0",net_house_seats)
Mean loss	-34.5	=AVERAGEIF(net_house_seats,"<0",net_house_seats)
First quartile (25%)	-50	=QUARTILE.EXC(net_house_seats,1)
Third quartile (75%)	-8	=QUARTILE.EXC(net_house_seats,3)

Having defined what's meant by "year" and "net_house_seats" in this workbook, we can use these defined names in place of ordinary cell references. Excel will even recognize them as we type them into functions, allowing us to autocomplete function calls and avoid typos. A statement like =AVERAGEIF(year, ">= 1990", net_house_seats), while not plain English, is as clear and user-friendly as we can make it (Figure 1-10).

FIGURE 1-10 ⬢ Naming Calculated Statistics



Some Descriptive Statistics Functions		*Formula Text*
Mean	-29.0435	=AVERAGE(net_house_seats)
Standard Error	5.065764	=STDEV.S(net_house_seats)/SQRT(CO
Median	-28	=MEDIAN(net_house_seats)
Mode	9	=MODE.SNGL(net_house_seats)
Standard Deviation	24.29455	=STDEV.S(net_house_seats)
Sample Variance	590.2253	=VAR.S(net_house_seats)
Kurtosis	-1.21398	=KURT(net_house_seats)
Skewness	0.000211	=SKEW(net_house_seats)

Some Descriptive Statistics Functions		*Formula Text*
Mean	-29.0435	=AVERAGE(net_house_seats)
Standard Error	5.065764	=standard_deviation/SQRT(count)

If you're using the result of one function elsewhere on a worksheet, you can name that result and use the name on the worksheet. For example, the standard error of the sample mean (which we cover in Section 8.3.3) is based on two other values calculated on this worksheet: standard deviation and count. To make the standard error calculation clearer, we select cell I9, the cell where we calculate standard deviation, and then click Formulas ▸ Define Name. One change we'll make in the New Name dialog is to change the scope of the name from the entire workbook to only the worksheet where this cell is found. We might calculate a different standard deviation or count on another worksheet.[5] We define standard_deviation, count, minimum, and maximum on this worksheet and use these defined names to further clarify the functions we use on the worksheet.[6] Defined names don't change the results of our data analysis, but they make it a lot easier to understand what you're doing.

A CLOSER LOOK: KEYBOARD SHORTCUTS

Some Excel users may find navigating the GUI cumbersome after a while. The good news for those who prefer the keyboard to the mouse is you can navigate the GUI using keyboard shortcuts. The bad news is that these keyboard shortcuts defy simple explanation and are basically random keystroke combinations. If you do the same task repeatedly, it might be worth finding the "cheat codes" on one of the extensive lists of Excel keyboard shortcuts online.[7]

1.5 SEPARATING ANALYSIS FROM DATA

We've tried to emphasize the difference between an Excel workbook and the worksheets within a workbook. If you're accustomed to single-page spreadsheets, the distinction between a workbook and a worksheet may seem semantic to you, but it's not. To use Excel successfully for data analysis, we encourage you to keep your data and analysis on separate worksheets.

If you've done data analysis with statistics software like SPSS, Stata, SAS, or R, you know that these programs allow you to save the commands you execute and the data you analyze in separate files. With these programs, you can see your data, a script of commands, and the results of your analysis in separate windows. This structure speeds computations and helps you keep your work organized. If you're accustomed to executing code in statistics software, it's strange to see analysis executed in a dataset. The strict structure of statistical programs is not entirely feasible with Excel, but keeping analysis and data on separate worksheets is generally a good idea.

If you expect to analyze some data more than once, we suggest saving the data on one worksheet, saving variable descriptions on another, and conducting your data analysis on separate pages. The data worksheet should not contain formulas or functions; even if you use some formulas or functions to generate new variables, the values should be fixed on the data worksheet.

When you use Excel functions to conduct data analysis, make your use of functions transparent and implement the analysis in a way that's easy to understand (if another person examines it or you use your own work later). You can display formula/function text on a worksheet using the

[5]If you later realize that you want a name defined for the entire workbook or want to limit the scope of a name to a single worksheet, select Formulas ▸ Name Manager to edit, add, or delete defined names.

[6]Unfortunately, you can't refer to cell ranges by defined name in Excel tool dialog, so it's important to understand cell ranges and keep worksheets organized with clear labels.

[7]For example, see Exceljet, "222 Excel Shortcuts for Windows and Mac," https://exceljet.net/keyboard-shortcuts.

FORMULATEXT function. If you're not sure whether a worksheet contains formulas, click Formulas ▶ Show Formulas to show all the formulas on a worksheet. If you're using a data analysis tool, it can be helpful to include some commentary on your analysis worksheet to explain what you've done. If you want to distinguish comments from the analysis contained in worksheet cells, try Insert ▶ Text ▶ Text Box. In other environments, researchers incorporate comments into computer code to explain some processes in plain English. Excel worksheets save function statements and their results in the same space, which can obscure how you conduct the analysis.[8]

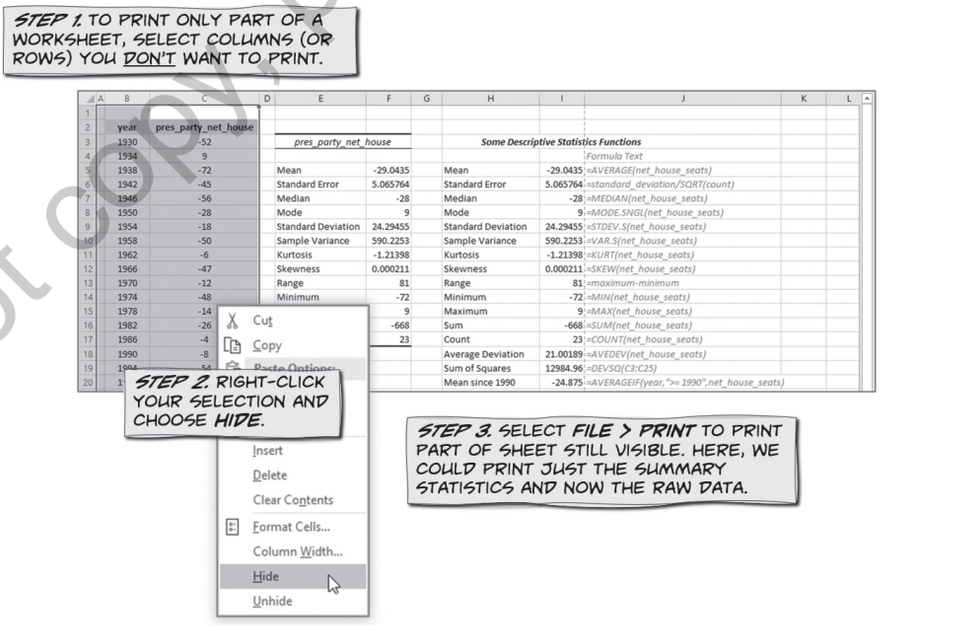
Don't insert the results of your analysis or charts you make on the data worksheet. If you need to reorganize data to use a data analysis tool, copy the variables you're working with to a new worksheet and leave the original data intact on the data worksheet. If an analysis worksheet gets too large and/or complicated, try separating your work onto several worksheets. Using separate worksheets for your data and analysis may seem a bit clumsy at first if you're used to working with simple, single-page spreadsheets, but it will help keep your work organized and make you more efficient in the long run.

1.6 PRINTING AND SAVING YOUR WORK

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

Printing desired results requires, first, that you decide what you want to print. To print, select File ▶ Print. In the print settings, you can choose, among other things, to print an entire worksheet, one worksheet, or just preselected parts of a worksheet.

FIGURE 1-11 ⬢ Printing a Selection of the Worksheet

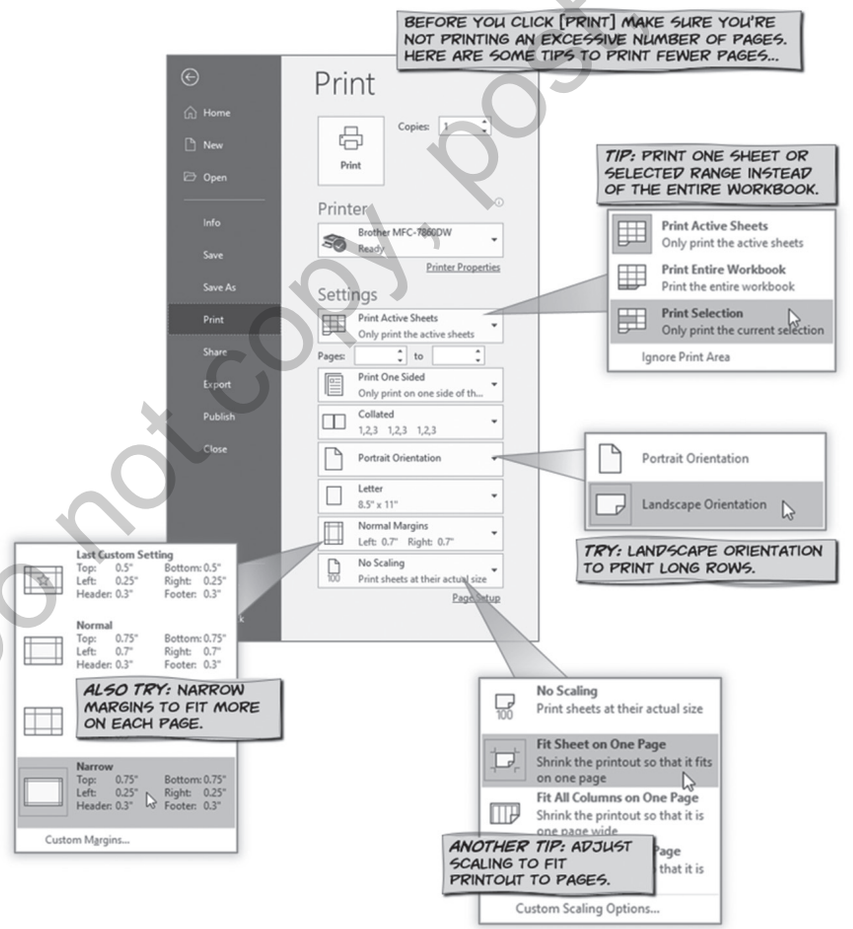


[8]You can audit a workbook in Excel to check how work is done and verify that results reported in a paper match results in a workbook, but you can't run a script file to replicate an analysis the way you can with other programs.

Before you click the Print button, always check the print preview and how many pages it says you're about to print. Worksheets can be odd-sized and/or very large. If you're not careful, you'll waste a lot of printer paper on pages with just one column. Here are some things you can do to be efficient with your printing:

- Hide rows and columns you don't need to print. To hide a row or column, click the row number or column letter to select the entire row/column, right-click, and select the Hide option. If, for example, we didn't want to print the columns of data, we'd select columns A:C, right-click, and hide them temporarily for printing purposes (see Figure 1-11).
- Adjust column widths to reduce white space on the page. It's nice to have some space between tables, but it can push one column to another page. To resize columns (or rows), position the cursor between column letters in the margin so it becomes a double-arrowed cross and then drag the column separators to narrow the columns.
- Highlight the worksheet range(s) you want to print before you click File ▶ Print and change the settings to "Print Selection." Check the print preview before you print.
- Change the print orientation to landscape if you are printing a wide area (see Figure 1-12).
- Use narrow margins, so long as your printer can handle it.
- Adjust print settings to scale the worksheet to print on one page, fit the columns to one page (if it's too long to fit on one page), or fit the rows to one page (if it's too wide for one page). This option usually works best in conjunction with our other suggestions for printing worksheets efficiently (see Figure 1-12).

FIGURE 1-12 ⬢ Printing Options in Excel



To save the work you've done in an Excel workbook, simply click File ▸ Save. If you're going to do data analysis with Excel at home, school, and elsewhere, we suggest you save your workbooks in a place you can access from multiple locations, like cloud storage. Then you can pick up where you left off.

To save an Excel graphic as an image file so you can insert it into another document, right-click the graphic, select Save as Picture . . . , and then save it in an appropriate location with a descriptive name. Excel supports a variety of image file formats such as jpg, gif, and png. You can also copy a chart from a worksheet and paste it into other documents, but this isn't always an option. For example, if you want to embed an image into a fillable pdf form or upload it to a website, you'll need to save it as a stand-alone image file.

1.7 FORMATTING TABLES AND FIGURES

Tables: in an academic context, tables report numbers in rows and columns; within Excel, a cell's ranges can be defined as a table.

Charts: visual representations of data and analysis; also called plots or graphs.

When you analyze political science data, you'll create **tables** and **charts** to summarize your results. Excel offers some easy-to-use options for formatting tables and charts. You can choose from an assortment of fonts, colors, borders, and styles; there are limitless possibilities. With this in mind, it's important to remember three things when you're formatting tables and graphs. First, always make tables and figures clear and easy to read. Second, when it comes to matters of style, consider the norms and expectations of your audience. Third, learn what formatting should be done in Excel and what you can do more effectively in other programs.

We create tables and figures to help us communicate the results of our analysis. Good tables and figures are concise and easy to navigate. No one likes trying to decipher results from a mess of numbers. The reader should be able to read everything; nothing should be cut off or too small to read. If you change colors, make sure the text clearly contrasts with the background. Computers may restrict variable names, but we can use plain English on tables and figures. Don't include unnecessary information just because the computer generates it. By default, Excel will calculate values to many decimal places, which is good, but all those numbers look overwhelming on a table or figure. There's rarely any reason to report values to more than two decimal places.[9]

Excel gives you many options for customizing the look and feel of tables and figures. You can change the colors, borders, backgrounds, fonts, alignments, and more.[10] When it comes to matters of style, keep your audience in mind. If you're doing analysis for an academic paper or assignment you're submitting to a professor, use academic-style tables and figures. If you're making a table or chart for a presentation, use bigger fonts and more colors and show only what's essential. If you're making tables and charts for exploratory analysis or to practice new methods, basic tables and charts are fine since you're the only one who needs to read them.

When you're formatting a table or figure that's destined for an assignment, paper, or presentation, keep in mind that you can continue editing something you made in Excel in another program. If you copy a table from Excel into a Word document, you can make additional adjustments, like changing fonts and merging cells, and this can be more efficient than editing the table in Excel. If the table or chart title already appears in your document, you can omit the title in Excel because it would be redundant in the final document. On the other hand, use Excel to round numbers to two or three decimal places because doing this in other programs is tedious and error prone.

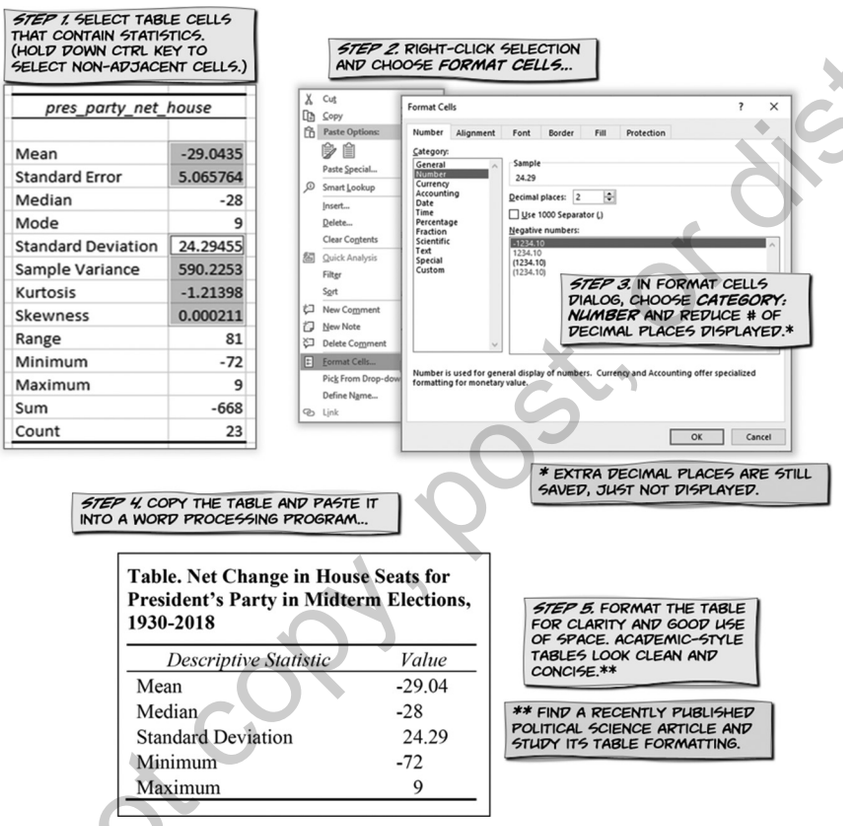
Let's apply these design and formatting principles to the table generated by the Descriptive Statistics tool seen in Figure 1-6. The tables produced by data analysis tools have a nice, clean academic look to them with simple borders and clear labels. To use a table like this in a paper or presentation, you

[9]This is a rule of thumb, not a law of political analysis. If you're reporting very small numbers, two decimal places may be insufficient; it's good to report at least one nonzero value.

[10]There are special table design tools in Excel that work on cell ranges specifically designated as "tables" in Excel. You can define a range of cells as a table by selecting the range and then clicking Insert ▸ Table. Once you've defined a table, you can use the table design tools on it.

only need to do some fine-tuning (see Figure 1-10). First, we'll reduce the number of decimal places on calculated values by selecting those cells (hold the Ctrl key down to select nonadjacent cells), right-clicking, and selecting the Format Cells . . . option. In the Format Cells dialog, we choose the Number category and set decimal places to 2. This rounds the skewness values to 0.00 so we can adjust that cell to show four decimal places; we see at least one nonzero digit there. We could make further edits in Excel, but we can also make edits in the word-processing document (e.g., Word) or presentation slide where we insert this table. To copy a range of worksheet cells to your computer's clipboard to paste into another document, simply select the range you want to copy, right-click, and select the "Copy" option. Once we put this table in context, we can delete rows of unnecessary descriptive statistics, give the table a title, and line up the decimal points. Our final product appears in Figure 1-13.[11]

FIGURE 1-13 ⬢ Formatting a Table of Descriptive Statistics



In this book, we'll show some tables and charts in the Excel default style to make it easier to follow our examples, but we suggest using academic-style tables and figures when you report your results.

1.8 GETTING HELP

To view the formal how-to manual for any Excel 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 COUNTIF function used earlier to generate descriptive statistics in addition to those

[11]Whether a row is necessary depends on how the table is used. Some of the descriptive statistics in the Excel table can be easily derived from other statistics. For example, if you report the minimum and maximum values, you don't need to report the range. The squared value of standard deviation is the sample variance.

reported by the descriptive statistics tool, you could click the Help button on that function's dialog (see Figure 1-14). Excel retrieves the technical manual information from Microsoft.com and displays it in a web browser. This help file includes a tutorial video, examples, solutions to common problems, and best practices (not all visible in the Figure 1-14 screenshot). To search all Excel help files, click Help ▶ Help and enter your search terms (see Figure 1-15).

FIGURE 1-14 ⬢ Getting Help with Excel Functions

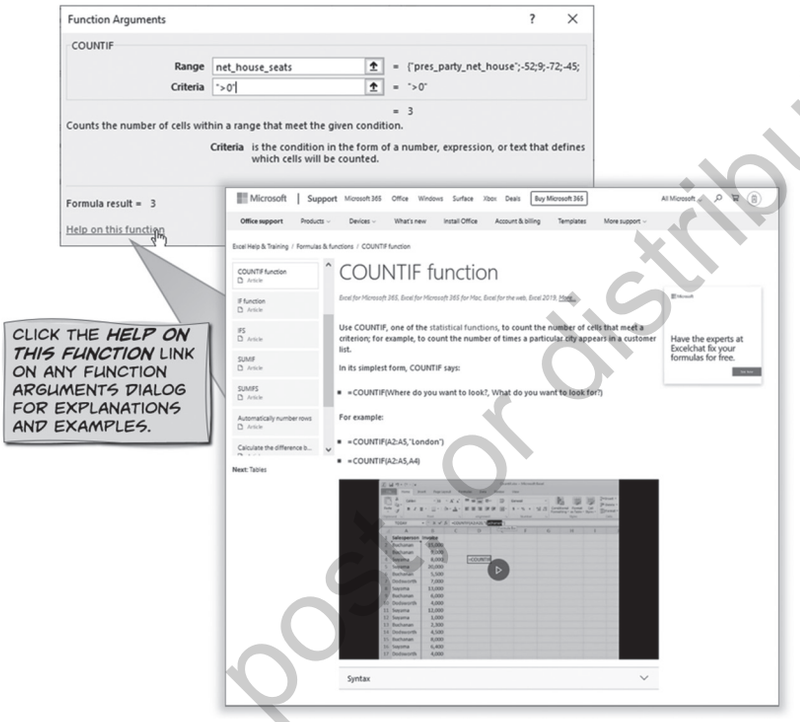
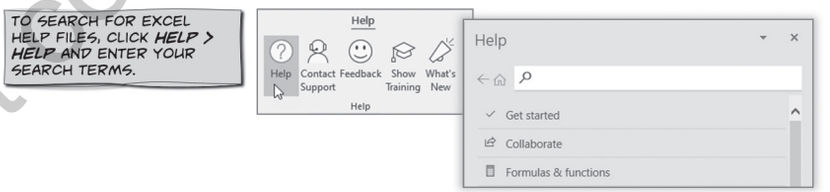
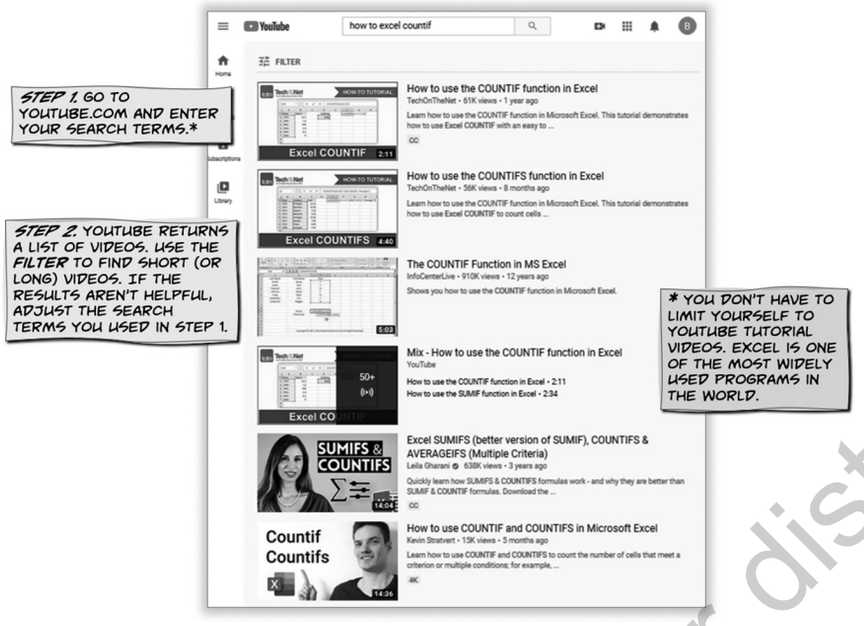


FIGURE 1-15 ⬢ Searching Excel Help Files



Microsoft Excel is one of the most widely used programs in the world. It should come as no surprise, then, that you can find help with Excel functions, shortcuts, data entry, and management tools on many different websites. We've compiled Excel tips and tutorial videos for every chapter in this book. Just point a smartphone camera at the chapter's QR code to access supporting online resources. If you learn how to use a function like COUNTIF, try searching "how to excel countif" on YouTube. There are hundreds of informative webpages and tutorial videos on just about every one of Excel's functions (see Figure 1-16).

FIGURE 1-16 ⬢ Excel Function Tutorial Videos



Because there are excellent resources to help you learn how to use any Excel function, we encourage you to keep the big picture in mind and not sweat the details of every function. As you learn to use Excel for political analysis, organize your workbooks using worksheets, define and use names to make your work clearer, show functions on your worksheets, and communicate your results as clearly as possible. Once you know what formula, function, or tool to use, you can quickly learn how to use it. The real challenge is learning what methods to use in different situations and properly interpreting the results.

CHAPTER 1 EXERCISES

Name: _____ Date: _____

E-mail: _____ Section: _____

1. In Section 1.1, we discussed the structure of Excel workbooks and encouraged you to use worksheets to keep your analysis organized. In this exercise, you navigate the States Workbook to find information about an observation in the dataset.
   A. Open the States Workbook. Go to the Data worksheet. What numeric value does the first state (Alabama, sorted alphabetically) have on the gunlaw_rank variable? _____. What numeric value does the state have on the Gun_rank11 variable? _____.
   B. Based on the values you reported in part A, are gun control laws in Alabama more restrictive or less restrictive than the laws in most states? Consult the Variable Descriptions worksheet as necessary. Explain your answer.

      _____
      _____
      _____
      _____

2. You will use cell references to execute Excel functions and data analysis tools throughout this book. It's important that you understand how to use cell references.[12] The following figure shows some numeric values on an Excel worksheet. Replicate these entries on your own worksheet and use Excel functions to answer these questions.

	A	B	C	D	E
1	0	3	10	18	15
2	6	19	7	9	1
3	5	15	20	9	3
4	11	20	11	15	19
5	12	12	6	10	2

   A. What's the value of =LN(D3)? _____
   B. What's the value of =SUM(A1:A4)? _____
   C. What's the value of =MIN(C3:E5)? _____
   D. What's the value of =AVERAGE(A3:E3)? _____
   E. Write a function statement (like those you see in parts A–D of this question) that gives the mean of columns B and D.

      _____
   F. What is the mean of columns B and D? _____

3. What's the difference between Excel formulas, functions, and data analysis tools? In your answer, please provide a specific example of each.[13]

   _____
   _____
   _____
   _____
   _____

4. The following figure shows some numeric values on an Excel worksheet. Replicate these values on your own worksheet. For this exercise, you will apply and discuss an Excel function that may be unfamiliar to you. This exercise is meant to improve your understanding of the syntax of Excel functions. Use Excel to answer these questions.[14]

	A	B	C
1	-4	25	25
2	24	14	4
3	1	20	15
4	20	16	10
5	25	3	20

   A. What's the value of the following function statement?

      _____

      =AVERAGEIFS(C1:C5, A1:A5, ">0", B1:B5, "<15")
   B. What cell values are being averaged to produce the result you report in part A? Specify these cells using their row numbers and column letters. _____
   C. Let's switch up some function arguments. What's the value of the following function statement? _____

      = AVERAGEIFS(B1:B5, C1:C5, ">0", A1:A5, "<15")

[12]See Section 1.1 for a discussion of cell references.

[13]You can read about formulas, functions, and tools in Section 1.3.

[14]Refer to Section 1.3 for a discussion of function syntax and arguments. If you need assistance with the AVERAGEIFS function, Section 1.8 shows you how to help get with functions like this.

D. What cell values are being averaged to produce the result you report in part C? Specify these cells using their row numbers and column letters.
_____

E. The AVERAGEIFS function calls above have five arguments: average_range, criteria_range1, criteria1, criteria_range2, and criteria2. Complete the following table to get a better understanding of Excel function syntax.

Argument	Value in Part A	Value in Part C	Brief Description
average_range	C1:C5	B1:B5	
criteria_range1			
criteria1			
criteria_range2			
criteria2			

5. To do political analysis with Excel, you'll need to activate Excel's Data Analysis ToolPak. We also recommend that you activate the Solver add-in and download/install the Real Statistics add-in.[15]

   A. When you activate the Data Analysis ToolPak, the "Data Analysis" button appears in which section of the Excel ribbon? _____

   B. When you activate the Solver, the "Solver" button appears in which section of the Excel ribbon?
   _____

   C. When you download and install the Real Statistics add-in, the button for the Real Statistics add-in tools appears in which section of the Excel ribbon?
   _____

   D. Apply the correlation tool in the Data Analysis ToolPak to the table of values from Exercise 4. The input range should be cells $A$1:$C$5. The data are grouped by columns, they do not have labels in the first row, and the output range should be an open space on your worksheet.[16] Fill in the following table using the output from this tool (you can round the results to two decimal places). Don't worry about what these numbers mean. The purpose of this exercise is to practice using a data analysis tool.

	Column 1	Column 2	Column 3
Column 1	1.00		
Column 2	_____	1.00	
Column 3	_____	_____	1.00

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 analyze some variables in the States Workbook's dataset.

   A. Which variable records the number of years of social studies that states require students to take to graduate high school? _____
   (fill in the variable's name)

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

   C. Which variable records the percentage of state residents who frequently attend religious services? _____ (fill in the variable's name)

   D. Which variable records the percentage of state residents who do voluntary community service? _____ (fill in the variable's name)

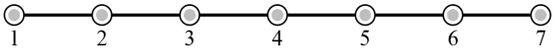
7. Researchers often use numbers to represent variable values instead of text. Using numbers simplifies datasets and speeds up computations. However, when you work with a dataset that uses numeric codes to represent variable values, you have to figure out what the codes mean to make sense

[15]See Section 1.2 for guidance on activating data analysis add-ins like the Data Analysis ToolPak, Solver, and Real Statistics.

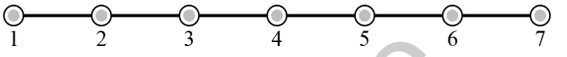
[16]See Section 1.3 for further guidance on using a data analysis tool.

of them. In this exercise, you'll figure out the meaning of three different variables in the Debate Experiment Workbook coded on a 1–7 scale.

A. What does the "pid" variable in this workbook's dataset measure? In addition to providing a written answer, please label the values on the 7-point scale provided.[17]



B. What does the "ideology" variable in this workbook's dataset measure? In addition to providing a written answer, please label the values on the 7-point scale provided.



8. Excel's data analysis tools do a decent job of formatting results in tables. If you're practicing new skills or doing some preliminary analysis, basic tables are fine since you're the only one who needs to read them. However, when you're sharing the results of some analysis with others, it's a good idea to format your results to make them clearer and more attractive.[18] To help you develop an eye for table formatting, carefully compare both of the following tables.

Basic Excel Table:

pres_party_net_house	
Mean	-29.0435
Standard Error	5.065764
Median	-28
Mode	9
Standard Deviation	24.29455
Sample Variance	590.2253
Kurtosis	-1.21398
Skewness	0.000211
Range	81
Minimum	-72
Maximum	9
Sum	-668
Count	23

Academic-Style Table:

Table. Net Change in House Seats for President's Party in Midterm Elections, 1930-2018

Descriptive Statistic	*Value*
Mean	-29.04
Median	-28
Standard Deviation	24.29
Minimum	-72
Maximum	9

Identify at least five differences between the basic table and the academic-style table.

1. _____
2. _____
3. _____
4. _____
5. _____

9. A political scientist wants to study health outcomes in countries around the world. To study the health–politics relationship, the researcher could analyze some variables in the World workbook's dataset that measure health in countries around the world. For each variable, briefly describe what it measures in your own words. To answer this question, use the Variable Descriptions worksheet and examine the variable's values in the dataset.

A. fertility

_____

_____

[17]If the variable descriptions in the Debate Experiment Workbook don't tell you enough to answer this question, search the Internet for additional information on seen-point party identification measures. This measure, along with the "ideology" measure covered in part B, is a classic political science variable.

[18]See the discussion of formatting academic-style tables in Section 1.7.

B. hiv_percent

_____

_____

C. infant_mortality

_____

_____

D. spendhealth

_____

_____

E. unnoncom

_____

_____

10. Excel is one of the world's most popular computer programs. This book does not cover everything you should know about Excel or all the things you can do with it. If you're looking for basic information about the program or specialized information about an Excel function not discussed in this book, there are many Excel resources on the Internet. For this exercise, conduct some independent research to identify good Excel resources. They can be websites, blogs, YouTube channels, message boards, and so on. Try to find resources that communicate information in a way that makes sense to you.

A. Identify and describe one good Excel resource that covers basic information, such as how to format cell values, change worksheet colors, and use the ribbon.

Name: _____

URL: _____

Describe what makes it a good resource:

_____

_____

B. Identify and describe one good Excel resource that covers specialized Excel functions, such as specialized functions for processing text strings and doing text analysis.

Name: _____

URL: _____

Describe what makes it a good resource:

_____

_____