Introduction to Making Graphs in Excel

For a beginner Excel user, the program can seem daunting with its many available menus to clean, analyze, and visualize data. But once you have the lay of the land, you'll find the tool isn't as scary as it is at first glance. I find the tool is well-organized and enables users to gain familiarity by using the same menus time and again to make changes to the spreadsheet and graph elements. That means we don't need to learn new menus, views, or buttons for every task. And in my opinion, the newer versions of Excel—through the Office 365 subscription model—are a big improvement over the older versions.

Don't get me wrong: there are some things about Excel that drive me crazy, and many of the graphs in the basic drop-down menu are drop-down dreadful. But the tool is built in such a way that we can make more advanced visualizations simply by understanding what it can and cannot do.

In this chapter, I introduce some basic parts of Excel for users who may have never opened the tool or made a graph. I focus primarily on the sections of the tool that are most relevant to creating effective graphs. There are many terrific blogs, resources, and books available if you want to learn other aspects of Excel like Pivot Tables, PowerQuery, macros and the Visual Basic for Applications (VBA) language, and more advanced formulas (see Appendix 3 for a curated list).

This chapter also provides some valuable information for the more experienced user. Not every Excel user knows about the *Plot Area* and the *Chart Area* or how to modify the properties of Excel charts. I've inserted this icon (🦉)

25

before the sections that I think everyone can benefit from learning, not just the first-time users. If you are experienced with Excel, you can jump to those icons to see if there are some nuggets of wisdom here to help you become a better Excel user.

The Excel View

Before we begin to explore using Excel, we must understand how the program is organized. The main area at the top of the Excel window is called the *ribbon* and each section of the ribbon (e.g., *Home*, *Insert*, etc.) are called *Tabs*. Within each Tab are a variety of options, menus, and buttons. Additional Tabs will appear when we are working within a specific element of Excel, like a graph or Pivot table. Below the ribbon is the *Formula Bar* where we can begin to type in a formula for our spreadsheet and Excel will prompt us with formula options (see next chapter) (Figure 4.1).
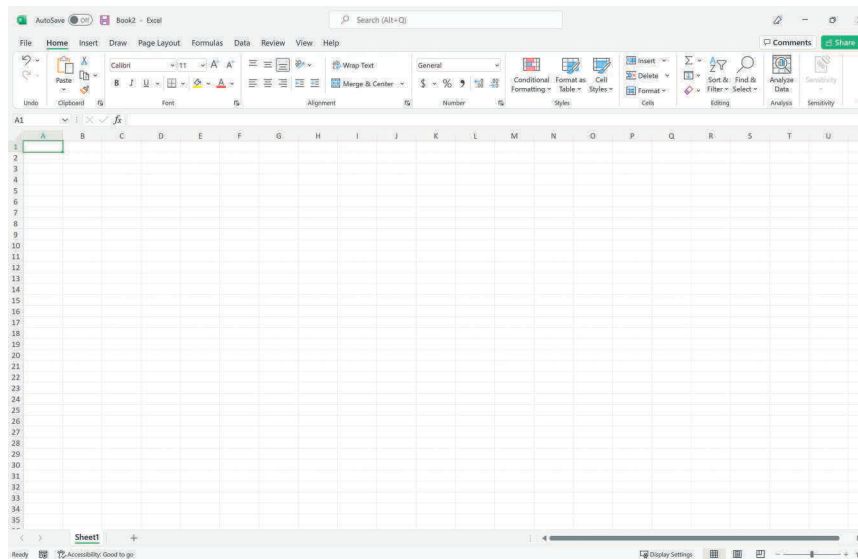

FIGURE 4.1

🐵 The narrow area above the ribbon is called the *Quick Access Toolbar* or *QAT*. When opening Excel for the first time, there will be two or three icons, likely for *Save* (🖫), *Undo* (↺), and *Redo* (↻). We can customize the QAT by selecting the down arrow on the far right side and choosing *Customize the Ribbon*. Once the dialog box opens up, we can add commands to the QAT or even create a new Tab. Adding a variety of commands to the QAT can help us be more productive by reducing the number of clicks we have to make.

🦉Alternatively, instead of customizing the QAT, you can familiarize yourself with keyboard shortcuts. You're most likely familiar with common shortcuts like CTRL+C for *Copy*, CTRL+V for *Paste*, and CTRL+S for *Save*. There are hundreds of other shortcuts that my memory isn't strong enough to remember all of the time (and they differ between PCs and Macs). But one other incredibly useful shortcut is CTRL+1 (CMD+1 on Macs), which will bring you to the *Format > Current Selection* menu (the drop-down menu in the far left part of the *Format* tab) of whatever you have selected. Want to format the horizontal (x-) axis in your chart? Select it and hit CTRL+1. Want to format the cells in the spreadsheet? Select them and hit CTRL+1. For me, CTRL+1 is as important as all of the other standard shortcuts we have come to know and love.

The spreadsheet itself—called "worksheets" in Excel parlance—takes up the main area of the view and consists of rows (horizontal) and columns (vertical), with the intersections called cells. Rows are numbered in Excel, and their default height depends on the computer's monitor and resolution—rows on my Windows Surface Pro have a default height of 38 pixels, while rows on my MacBook Pro have a default height of 16 pixels. Columns are labeled with letters, and their size will also depend on the screen—columns on my Surface Pro have a default width of 136 pixels and columns on my MacBook Pro have a default width of 65 pixels. Both dimensions can be adjusted by selecting the column/row and right-clicking or by dragging the column/row edge. The spreadsheet gridlines can also be modified or turned off by using the *Show* menu in the *View* tab.

You may notice that Excel shows two numbers for row heights and column widths. A column width unit in Excel is equal to the width of one character, with the default on my Surface Pro equal to 47 characters or 136 pixels. Row height is measured in pixels *and* points, which can make it a little annoying to compare the dimensions of rows and columns. Fortunately, if we stick to pixels and don't worry about converting to characters or points, it's not such a big deal. Excel determines the width of columns by the character width of the font, which can vary between monitors, screen resolutions, operating systems (e.g., Windows and Mac), and font families. As a result, some of the column widths or row heights in the tutorials might not look perfect on every machine, but we can adjust them to fit our preferences.

At the bottom of the Excel window are some basic controls of the worksheet. Each worksheet has a default name and number (e.g., Sheet1, Sheet2), but the name and color of the tabs can be controlled by right- or double-clicking. We can toggle between sheets by clicking on the desired sheet or by using the left/right triangles/arrows. We can also control the view (*Normal*, *Page Layout*, *Page Break*) with the buttons on the right side and the zoom level with the slider next to that.

A few other keyboard combinations can come in handy when selecting data and navigating through the various menus.

1. CTRL+Arrow keys (CMD+Arrow on Macs) will bring us to the last used cell in the current data region. If there is data in cells A1 to A100, place the cursor in cell A1 and press CTRL+Down Arrow. The cursor will go all the way down to cell A100.

2. SHIFT+CTRL+Arrow (SHIFT+CMD+Arrow on Macs) will let us *select* all of the data in the cells going in the direction of the arrow. In the previous example, if we pressed SHIFT+CTRL+Down Arrow, we would select all of the data in cells A1 to A100.

3. ALT+. Many of the buttons in the ribbon have a dedicated keyboard shortcut assigned to them. With any cell selected, if we press the ALT key plus a letter, number, or function key, we can shortcut right to that tab or command. Personally, I'm a mouse-first user, so I don't rely on the ALT+ commands (plus, they don't work on Macs), but ALT+ commands can help you work faster. Table 4.1 contains a curated list of the most important shortcuts.

TABLE 4.1 Selected Useful Excel Keyboard Shortcuts

Shortcut	PC	Mac
Copy	CTRL+C	CMD+C
Cut	CTRL+X	CMD+X
Paste	CTRL+V	CMD+V
Undo last action	CTRL+Z	CMD+Z
Redo last action	CTRL+Y	CMD+Y
Find	CTRL+F	CMD+F
Save	CTRL+S	CMD+S
Save as	CTRL+SHIFT+S	CMD+SHIFT+S
Add hyperlink	CTRL+K	CMD+K
Format (almost) anything	CTRL+1	CMD+1
Go to next worksheet	CTRL+PgDn	Option+Right Arrow
Go to previous worksheet	CTRL+PgUp	Option+Left Arrow
Select entire row	SHIFT+SPACE	SHIFT+SPACE
Select entire column	CTRL+SPACE	CTRL+SPACE
Select entire worksheet	CTRL+A	CMD+A
Move active cell/menu right in a selection	TAB	TAB
Move active cell/menu left in a selection	SHIFT+TAB	SHIFT+TAB

Setting

Making

Moving

Excel Graph Types

To insert a chart in Excel, we must navigate to the *Insert* tab. There are about eight main chart types listed in the default *Charts* tab of the ribbon (again, the layout looks a little different between PCs and Macs) (Figure 4.2).
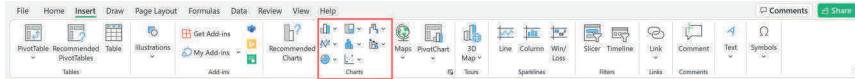

FIGURE 4.2

Inserting a chart is easy—select the data in the worksheet and select one of the charts from the *Insert* menu. Excel will place the graph closest to your cursor, but you can always move it to wherever you want it in the worksheet. Once these steps are done, a few things will happen.

1. A new *Chart Design* tab appears in the ribbon. This tab enables us to add different elements to the chart, select data, and change the chart type (Figure 4.3).

   The *Chart Design* menu contains some especially important parts:

   a. On the far-left area of the menu is the *Add Chart Element* button. Clicking on this button will reveal a number of ways we can add,
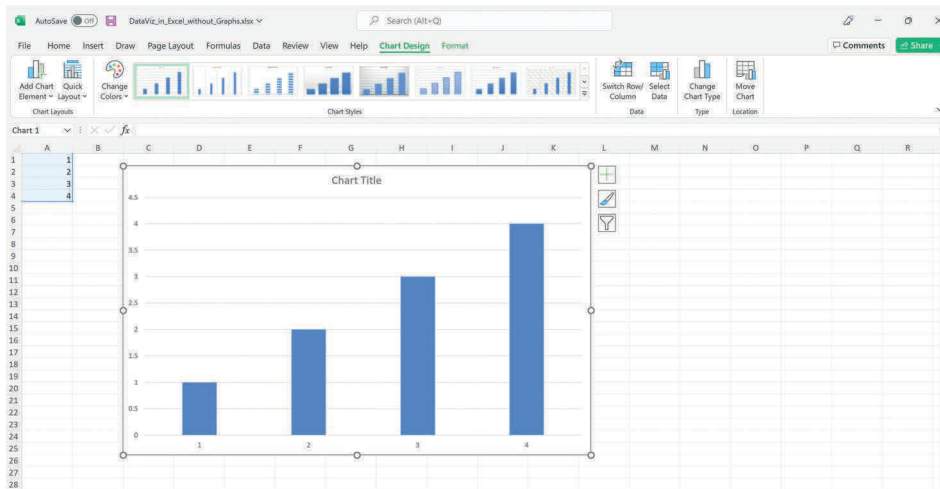

FIGURE 4.3

subtract, or edit elements in the chart, including axis titles, gridlines, data labels, and more (Figure 4.4). We will use this menu extensively in the tutorials. As you gain more experience using the various menus and creating visualizations, I will assume you can find the menu on your own and won't describe it each time.
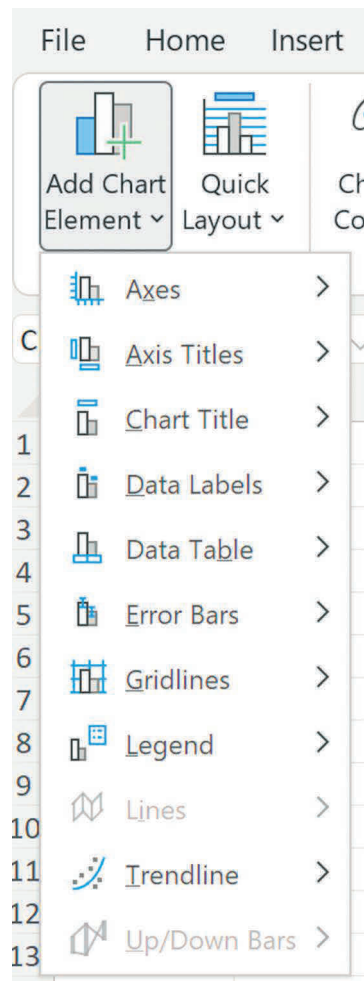
Setting

Making

Moving

FIGURE 4.4

  b. Toward the right side of the tab is the *Switch Row/Column* button. Excel orients the selected data based on the program's default behavior, so if our data stretch across columns, Excel might plot it along the rows and vice versa. We can easily switch the orientation of our data by selecting this button rather than changing the layout of the data (Figure 4.5).

  c. Next to that button is the *Select Data* button, which we can use to add, subtract, and edit the data in the chart. We can also access this button by right-clicking in the chart and choosing the *Select Data* option in the menu. We will add and edit data in our charts throughout the book, so it's important to know how to use this menu as we will use it over and over again (Figures 4.6–4.8).
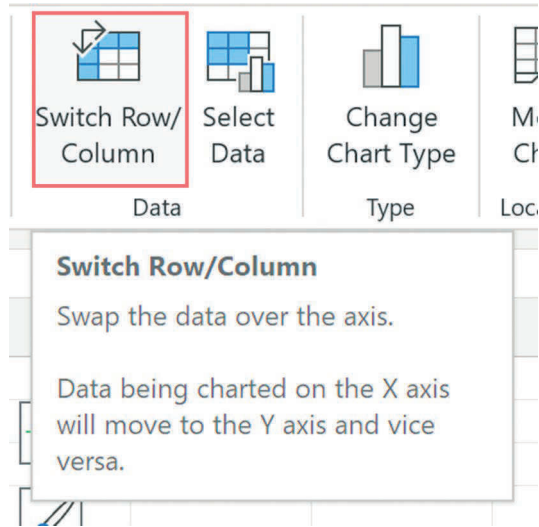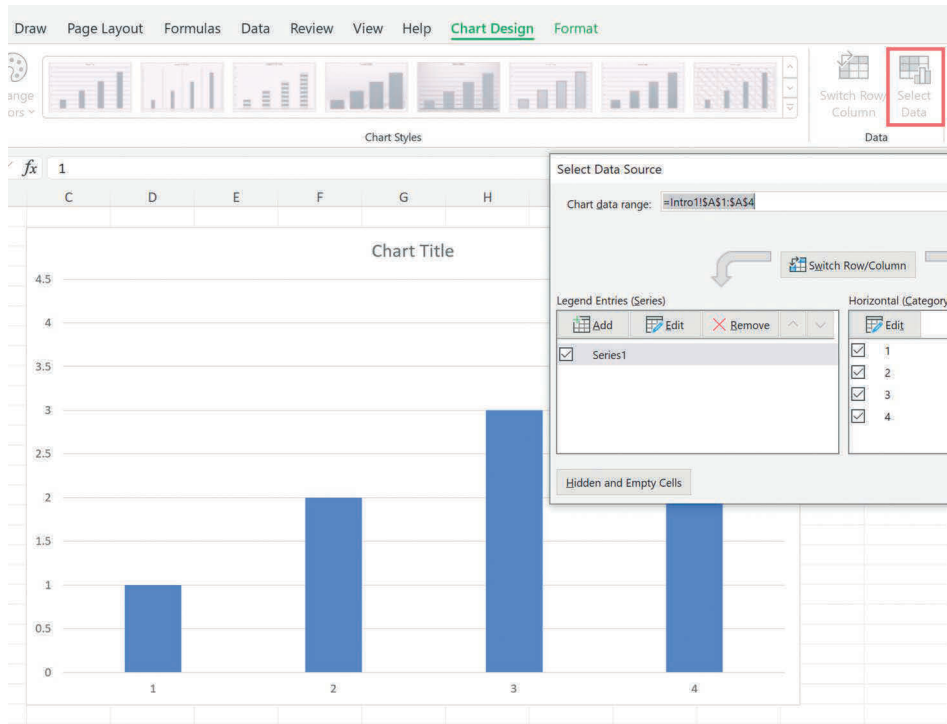
FIGURE 4.5

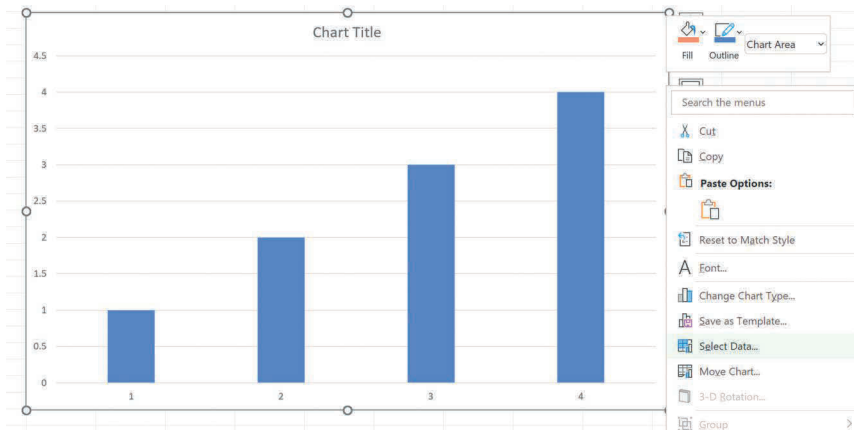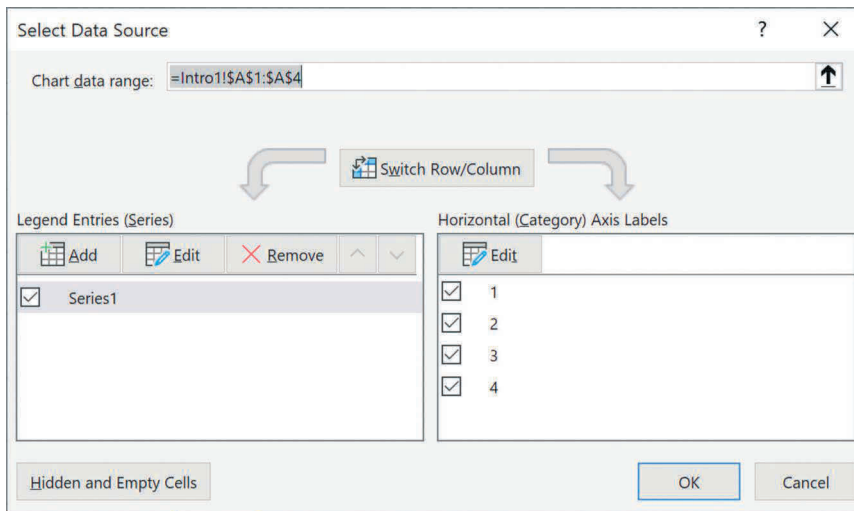**FIGURE 4.6**

Setting

Making

Moving

FIGURE 4.7

FIGURE 4.8

Sometimes when we add or edit a new series, Excel will insert a placeholder in the selection box. That text will be a "={1}" character, and we will need to overwrite or delete it before inserting our data series. If we don't overwrite it, we will end up with a data entry like "={1}=Sheet1!$A$1:$A$10" and Excel will show an error message. To fix our data, we will need to correct the formula or re-insert the reference.

d. Finally, the *Change Chart Type* button (found by right-clicking on the chart or by using the button in the *Chart Design* tab) enables us to change the chart type directly without having to delete what

we have (Figure 4.9). We can also use this menu to combine chart types, something we will do a lot in this book. On PCs, there is a *Combo* chart option at the bottom of this menu, which allows us to select a chart type for one data series and a different chart type for another data series. Be careful to note that if we select an entire chart and use the *Combo* chart option, Excel will change every series to the *Clustered Column* chart (which is apparently the default chart option). 🙊 Although it's not specified, every time you use the *Combo* chart option, it's best practice to select the chart series you want to change first before navigating to the *Combo* chart option (Figure 4.10).

On Macs, select the series in the chart you want to change, then select the new chart type in the *Change Chart Type* menu (Figure 4.11). There is no *Combo* chart option on Macs, so you need to have the series selected to change it.

Not all charts can be combined, and with few exceptions, we can only combine two chart types at a time.
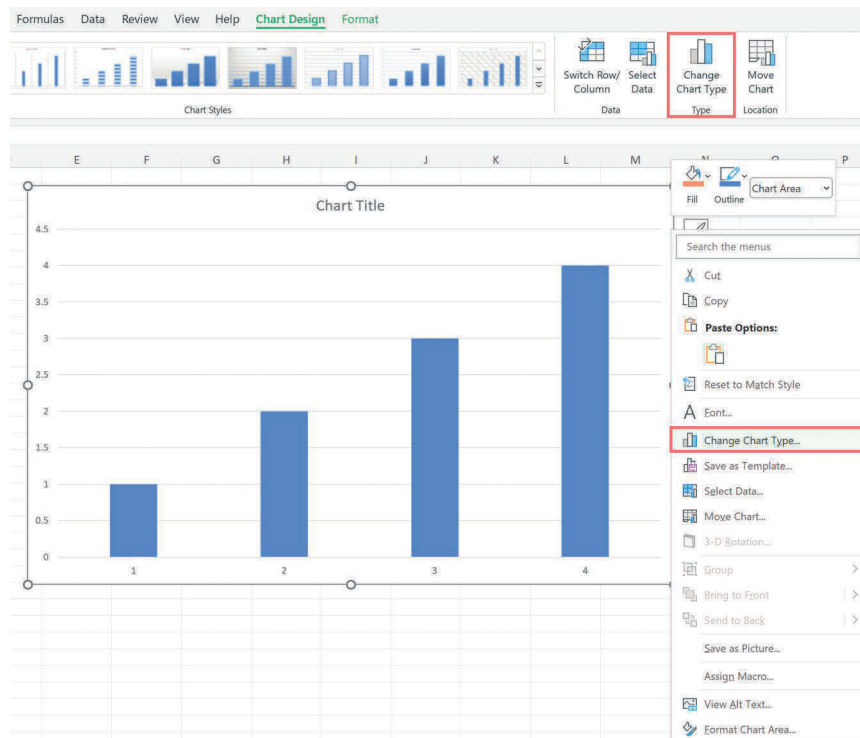

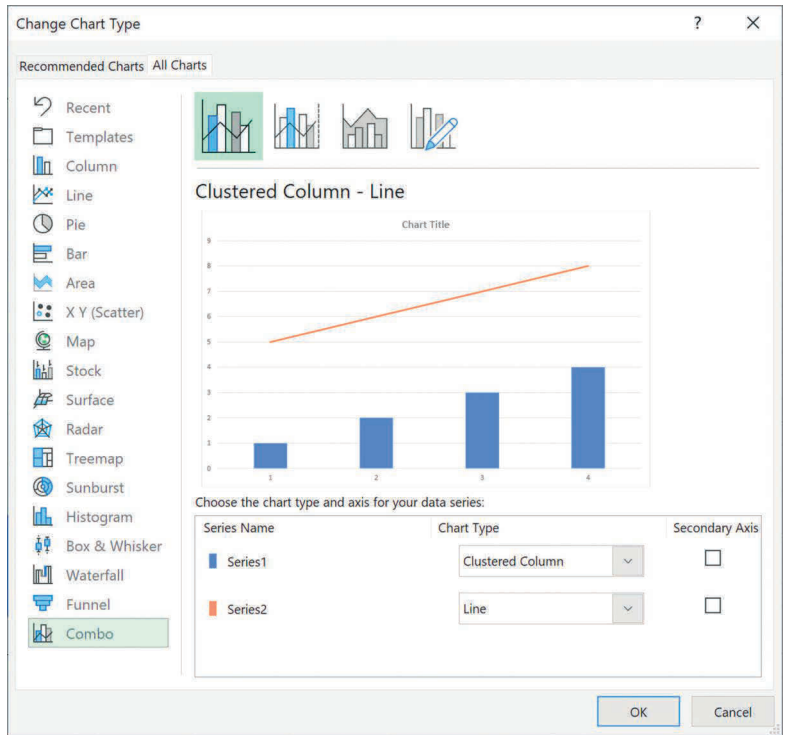**FIGURE 4.9** *Change chart type menu on PCs.*

Setting

Making

Moving

Setting

Making

Moving

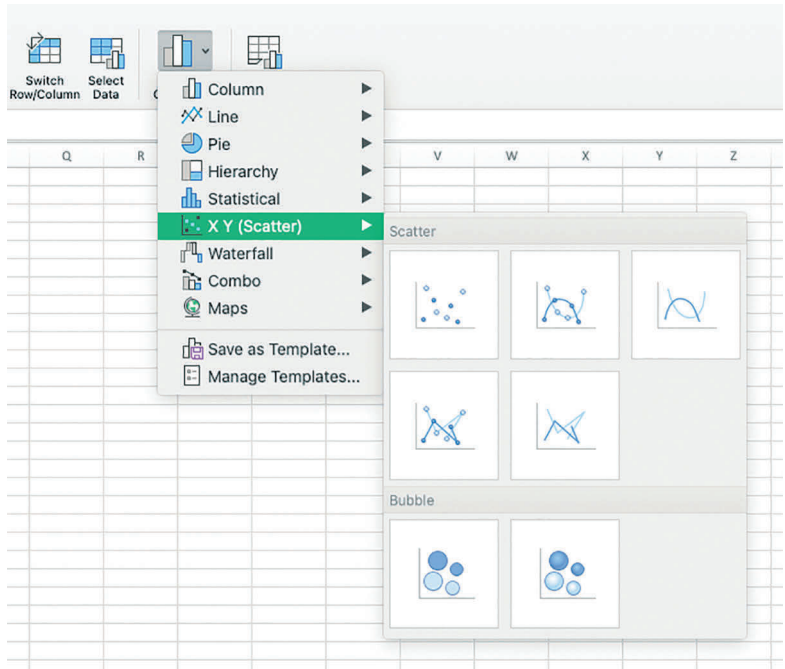**FIGURE 4.10** *Combo chart option menu on PCs.*

FIGURE 4.11 *Change chart type menu on Macs.*

2. When we select a chart, a new *Format* tab will appear. This tab houses various formatting options including colors, text options, and chart dimensions (Figure 4.12).
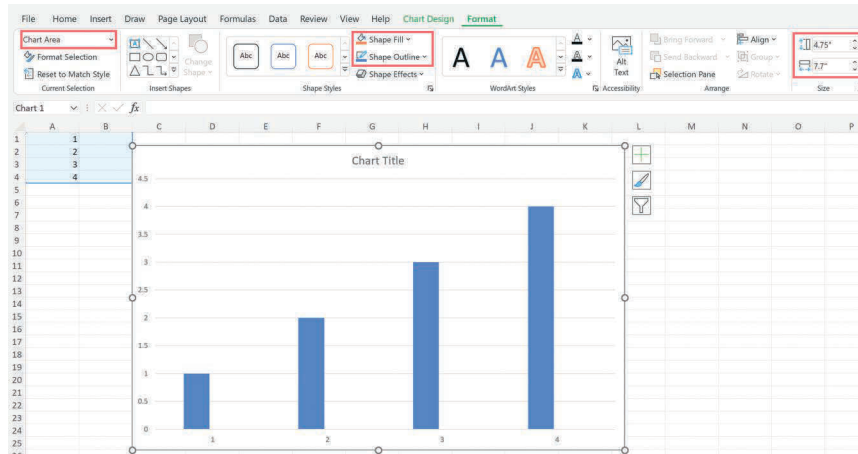
FIGURE 4.12

A few buttons and menus are particularly important here:

a. 🔲 The drop-down menu on the far left in the *Current Selection* area ("Chart Area" in the image) lists all of the elements in our chart, including axes, gridlines, and the data. If we have a chart with a lot of data or other elements in it, we can use this drop-down menu to select an element.

b. The *Shape Fill* and *Shape Outline* buttons near the center of the menu enable us to control the "fill" colors and "outline" colors of the elements in our chart.

c. On the far right of the tab are two fields that control the size of the chart. The default size for a chart is 3×5 inches. If we want to change the size and keep this ratio the same, we can check the box next to these fields and change the values by using the up/down arrows or typing new numbers.

3. The chart will appear near the center of the window and will have a series of default aspects, including a horizontal (x-) and vertical (y-) axis, default colors, a "Chart Title" box, and a default size of 3×5 inches.

4. When the chart is selected, the data included in the chart will be high-lighted in the worksheet.

Setting

Making

Moving

Office 365, which launched in 2013, introduced better and more current security features and allowed Microsoft to add features or fix bugs. One feature that Microsoft added to Excel was a new set of graphs. Some of these graphs—such as treemaps, sunburst diagrams, and box-and-whisker plots—require different data layouts than the standard one row per observation and one column per variable. 🦉 If you'd like to explore these graph types but don't know how to set up the data, you can insert the chart in PowerPoint (using the standard *Chart* area of the *Insert* menu). When you do so, PowerPoint will open an Excel file with placeholder data, which you can use as a template for your own data entry and graph creation.

Chart Parts and Properties

Many new data practitioners assume that a chart is nothing more than a horizontal axis, a vertical axis, and plotted data. But it's worth recognizing that a graph consists of many different parts that we can control and change.

Take this basic line chart, which shows the monthly unemployment rate in the US from 1960 to 2021 (Figure 4.13). At first glance, you may think, "Okay, the line goes up and down, up and down." You may also notice that there is a horizontal (x-) axis, a vertical (y-) axis, a title at the top, and so on.
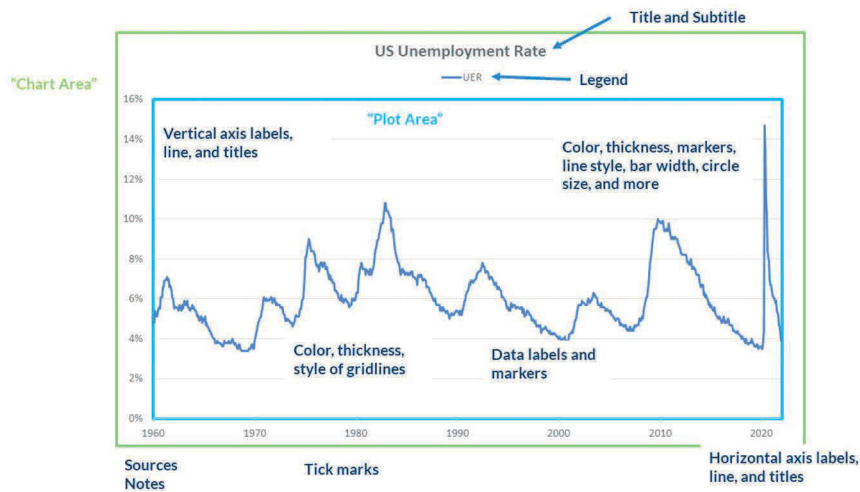

FIGURE 4.13

But there is a *lot* more going on in this graph than the basic line and title, and we can control and style all of it. We can set the font type, size, and orientation of the title, subtitle, sources, and notes. We can show or hide the vertical and horizontal axes, and style their color and thickness. We can turn tick marks on

or off; change the style, thickness, color, and frequency of the gridlines; and, of course, style the actual data shown in the graph, be it bars, lines, or circles.

In the example, I annotated (and added a line around) what Excel calls the *Plot Area* and the *Chart Area*. The styles, sizes, and arrangement of these areas can be adjusted—the *Plot Area* can be made smaller, for example, to make room for labels or other text—and is something that we will encounter in a few of the tutorials.

There are a few ways to modify the different chart elements. First, we can select the chart and use the *Add Chart Element* button in the *Chart Design* menu (which will often appear as *Chart Design > Add Chart Element* in the tutorials). Second, we can select the *plus* button that appears next to a selected chart (PCs only). Third, we can right-click on any element of the chart and select the appropriate *Format* menu, such as *Format Axis* or *Format Data Series*. Finally, we can use the CTRL+1 keyboard shortcut (CMD+1 on Macs) to modify the selected object.

Using any of these options will bring you to the *Format* menu. This menu has numerous options, but the top-line icons are as follows (Figure 4.14):

- The *paint can* icon lets us change the fill and outline colors of chart elements as well as change or style markers in the chart.

- The *pentagon* icon allows us to add styling options like a shadow or a glow around chart elements.

- The *arrow/square* icon contains options to control different text alignment features such as direction and margins.

- The *bar chart* icon differs for each part of the chart, but can include axis ranges, bar widths, and number formats.
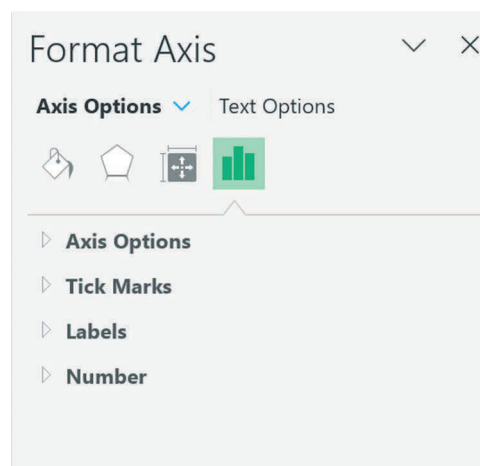
FIGURE 4.14

🦉All of these options are available for the chart as a whole as well, so we can change the border or background color of the entire chart. The *Format Chart Area* menu also includes a *Chart Options* area, which you can see in Figure 4.15. Here, we can control the dimensions of the entire chart using the fields under *Size* (though there are no direct controls for the *Plot Area*). We can also set whether the chart can be moved and/or sized with the cells in the *Properties* area of the menu. If we create a chart and insert or delete a row or column, the default option (*Move and size with cells*) in the *Properties* area dictates that the size of the chart will also change. But if we want to change things in the spreadsheet without changing the size of our chart, we can select the other options.
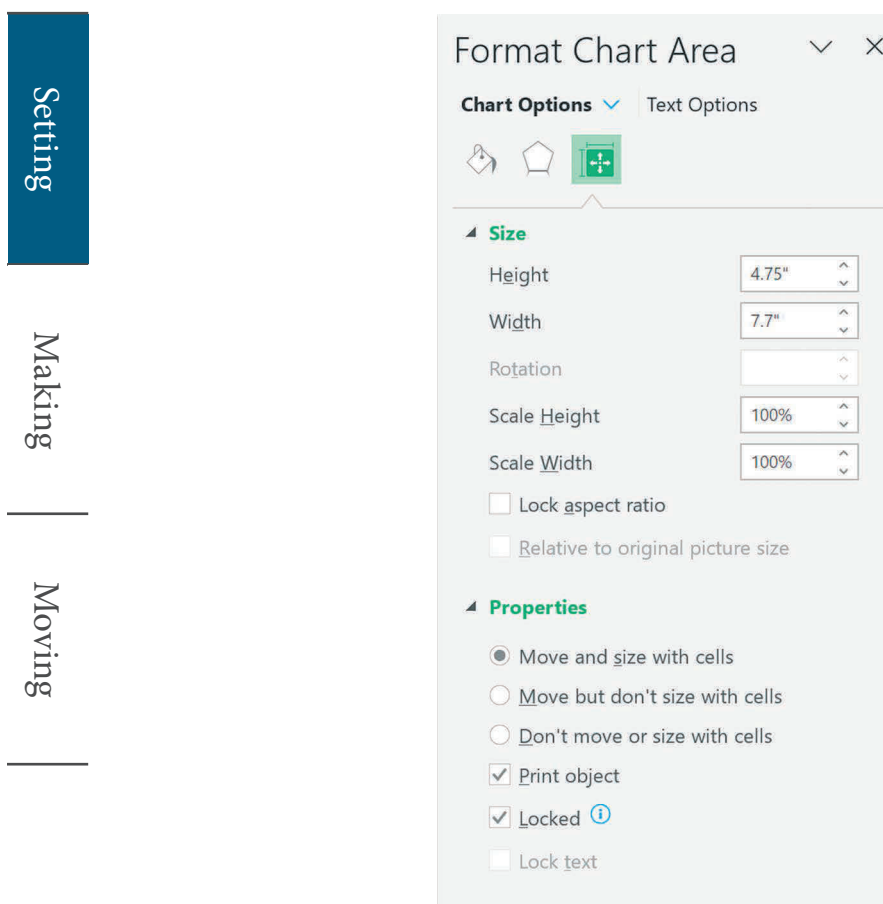
FIGURE 4.15

Excel is sensitive to using actual numbers in our charts, which means it won't know what to do with text or with spaces. Although it can be infuriating when we get an error and can't figure out why, it can also be used to our advantage. We can (and we will!) insert an "=NA()" formula in our worksheet

cells when making graphs because Excel will ignore those values. Let's say we want to add a data series to a chart in order to add labels (again, something we will do a lot!), we can insert NAs in the data series instead of 0s because Excel will ignore the NAs but will plot the 0s.

Recognizing the Defaults

Throughout all of these options and menus, we must recognize Excel's defaults and when we need to change them. Excel simply plots the data how we specify—it's not doing anything right or wrong, so being aware of its default behaviors and how to change them is crucial.

There are six main default behaviors in Excel that are worth keeping in mind because they will come up again and again in the tutorials.

1. **Adjusting the Vertical Axis Range and Units.** As a default, Excel will not let you plot data to the top of the vertical axis. If you need to change the vertical axis limits, you can do so in the *Format Axis* menu.

   The scatterplot in Figure 4.16 shows per capita gross domestic product (GDP) on the x-axis and life expectancy on the y-axis. The default range for the vertical axis is 0 to 100, but we should adjust that minimum for two reasons. First, 0 is (fortunately) not an average life expectancy for any country. Second, unlike in a bar chart where a baseline of 0 is important and necessary, in a scatterplot, we can use a nonzero axis. In this case, we can type "50" in the *Minimum* box to manually fix the starting point. If you add more data, the *Maximum* value might change, but the *Minimum* will stay at 50 until you reset it.
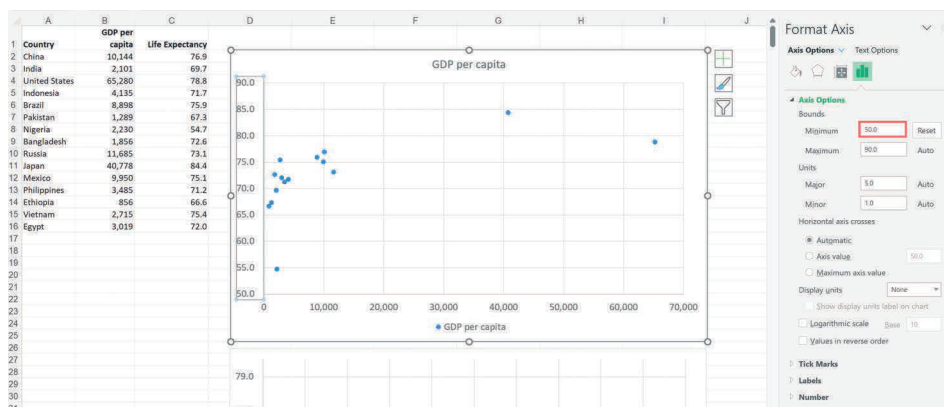

FIGURE 4.16

2. **Length of Error Bars.** Error bars are useful for our graphs because they allow us not only to denote uncertainty or ranges but also to add annotations or pointers to specific data points. Excel has preset defaults when we insert error bars into our charts. 🦉 For scatterplots, Excel will insert two error bars: one along the vertical dimension and one along the horizontal dimensions. For line charts and bar charts, Excel will only insert an error bar along one dimension.

We can also adjust the length and appearance of the error bars, but it's worth noting that the default value of all error bars is set at a *Fixed Value* equal to one. If I zoom into the GDP-life expectancy scatterplot and add error bars, the vertical error bars are visible but the horizontal error bars are not (Figure 4.17). The horizontal bars exist, but because they have a length of 1, they cannot be seen because the range on the x-axis (0–5,000) is much, much larger than that of the y-axis (65–80).
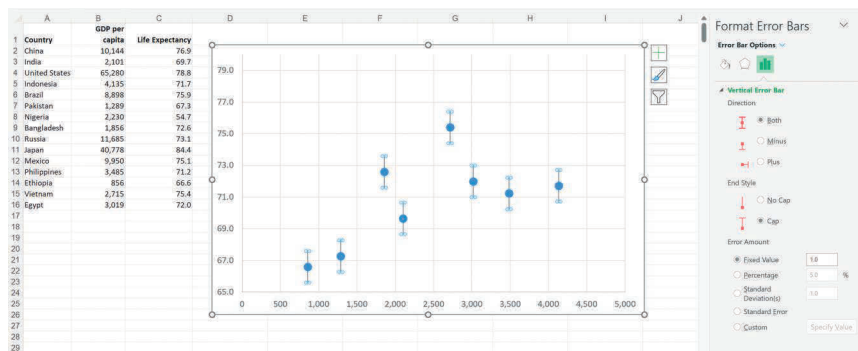
FIGURE 4.17

3. **Creating the graph.** There are some basic defaults that apply when we insert a chart in Excel. Take a scatterplot—when we select two data series and insert a scatterplot (or what Excel calls a *Scatter*), Excel will, by default, place the first series in the *X* (horizontal) position and the second series in the *Y* (vertical) position. We can always change the organization of the data in the worksheet or in the chart by using the options in the *Select Data Source* menu (Figure 4.18).

4. **Color transparency.** We will sometimes have bars, lines, or dots that overlap one another. In some cases, we might need to winnow down our data to show a smaller number of series, but in other cases, we can change the transparency of the colors to help show the data behind the object. In the *Format Data Series* menu, we can change the transparency of the fill and border colors by selecting a color and adjusting the *Transparency* slider (Figure 4.19).

Setting | Making | Moving
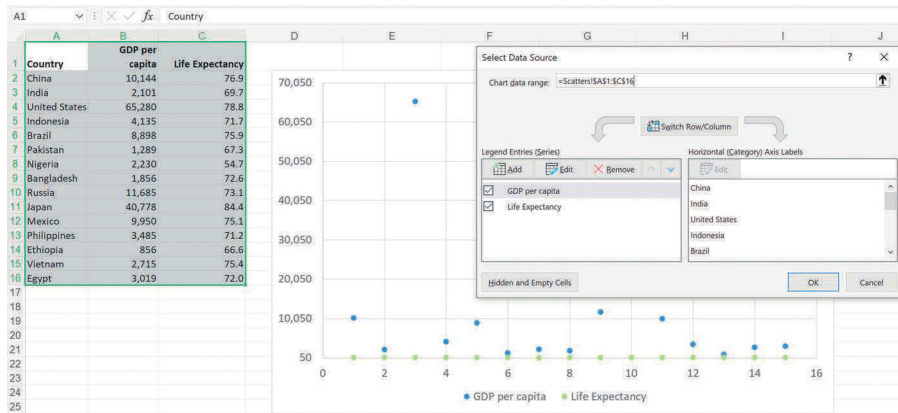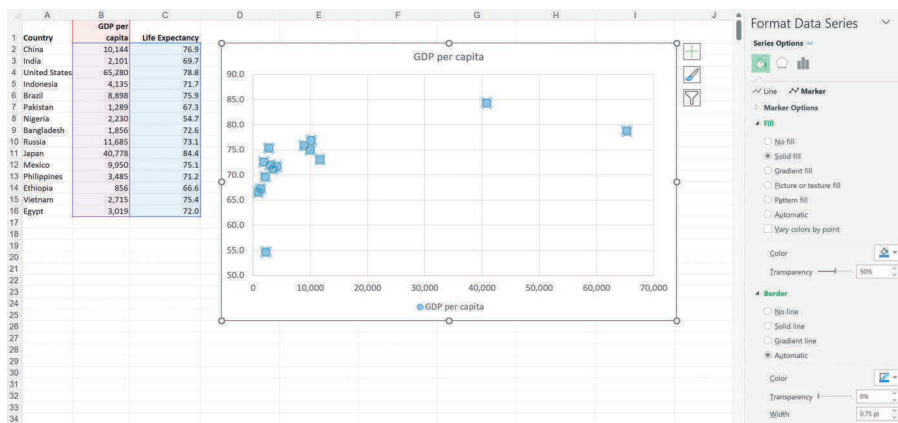
FIGURE 4.18

FIGURE 4.19

5. **Bars and gap width.** For bar charts, Excel sets the default widths of the bars and the gaps between bars to –27% and 219%, respectively. We can adjust these widths by right-clicking on the bars and selecting the *Format Data Series* menu. Here, we can use the *Series Overlap* and *Gap Width* options to adjust. *Series Overlap* refers to having bars sit on top of each other instead of next to each other. *Gap Width* refers to the space between the bars, so adjusting the width of the gap changes the width of the bars. The chart on the left of Figure 4.20 uses the default settings for these options, while the chart on the right uses edited values of 0% and 100%. Personally, I prefer the 0% and 100% combination because the ratio of bar width to white space looks better to me, but neither setting is objectively better than the other.

Setting

Making

Moving
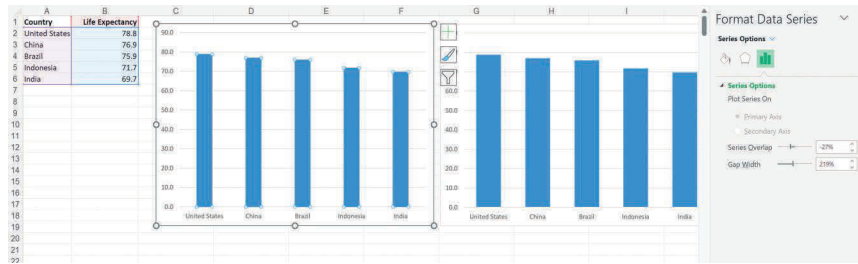
FIGURE 4.20

6. **Primary and Secondary Axes.** The final styling Excel default to be aware of is the ability to place data on secondary axes. When we insert a column chart, for example, Excel places the data along a horizontal (x-) axis at the bottom of the chart and a vertical (y-) axis along the left side of the chart. We could add more data and apply them to the "secondary" axes—a horizontal axis at the top of the chart and a vertical axis along the right side of the chart. 🦉 In most cases, Excel will not show the *secondary horizontal axis* by default. Instead, we need to make that axis visible by using the *Chart Options* menus. If we add a data series to a chart on the secondary axis but it doesn't look quite right, we may need to turn on and adjust those axis values (Figure 4.21).
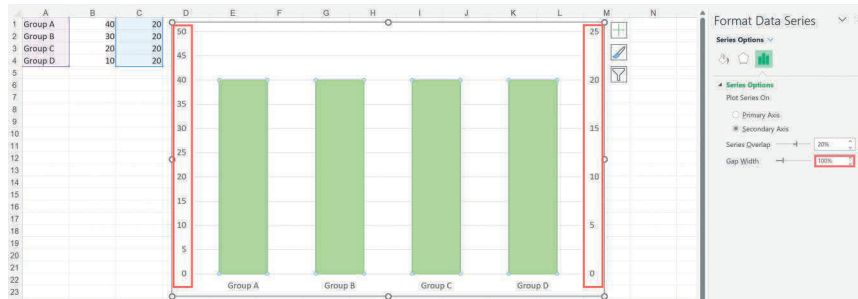
FIGURE 4.21

Worksheet Formatting

Although this book focuses on data visualization, the formatting options in worksheets, which are generally used for text elements, can still be useful for us. As you'll see, there are lots of charts we can make directly in the worksheet by using the formatting options.

The primary formatting tools in Excel are located in the *Home* tab of the ribbon, and we'll discuss each area, delineated by vertical lines, from left to right along the first half of the tab (Figure 4.22).

Setting | Making | Moving
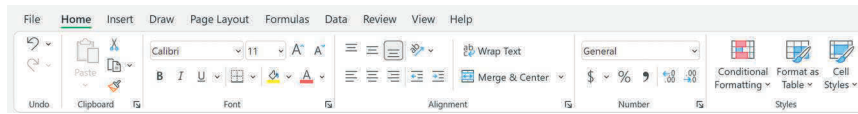
FIGURE 4.22

Copying and Pasting

You probably know the basic keyboard shortcuts for copying and pasting (CTRL+C and CTRL+V), but the *Paste* drop-down menu is important because it can allow us to paste data or calculations as formulas with or without the cell formatting, as values (just the numbers, no formulas), or as a variety of other options. 🗿 In addition to the *Paste* menu options and the well-known keyboard shortcuts, Excel also lets you can drag the contents of a cell—values or formulas—across rows or columns by grabbing the green square in the bottom-right part of the selected cell. You can double-click on that part of the square and the cell contents (including formulas) will fill up to the next blank cell (Figure 4.23).
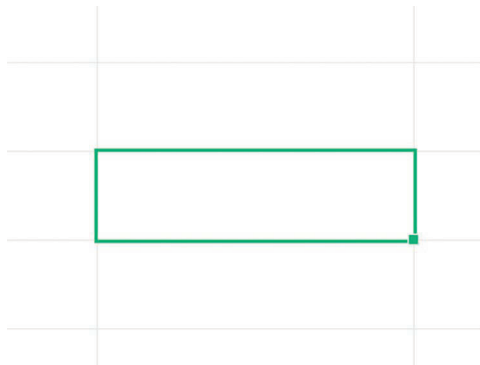
FIGURE 4.23

   In the tutorials, when you see something like, "Copy-and-paste the formula down the column," I am referring to three steps: select the cell with the original formula and copy (CTRL+C), select the cells where you want to add the formula, and paste (CTRL+V). Alternatively, you can select the cell with the original formula and drag it down the length of the column using the green square in the bottom-right part of the cell.

Text Formatting

In the text formatting area, we can control the font type and size by using the two drop-down menus, and we can add boldface, italics, or underline.

The button with the grid is the *Borders* area of the tab, which allows us to add borders around cells. The color and thickness of the border can be controlled at the bottom of that menu. The default is a black border, but we might have cells filled with color and want to use a white border so changing the color of the border can be helpful. Finally, we can change the color fill of a cell in the menu with the paint can icon and the color of the text in the menu with the capital letter *A* icon.

Cell Alignment

Here, we can control the vertical alignment of the cells using the top three icons and the horizontal alignment using the bottom three icons. We can also rotate the text by using the icon with the letters "ab" on top of an arrow. Finally, we can indent text with the two icons that show arrows and three lines, a tool that can be useful when we want to align numbers in tables (see Chapter 29).

Although a vertical line separates the indent icons from the next area of the tab, that area is also part of the alignment section. The *Wrap Text* tool lets us expand a cell so text appears on multiple lines instead of overlapping with other cells. The *Merge & Center* tool merges selected cells together and offers a few formatting options. These tools can be used together to save time when creating data visualizations by merging cells and wrapping the text so the text isn't entered separately in multiple cells.

🦉One last thing worth knowing: we can force a carriage return in our cells by placing the cursor where we would like the return and using the ALT+ENTER keyboard combination (Option+ENTER on Macs). If we want to insert a carriage return in a formula, we can use the CHAR(10) formula (with the *Wrap Text* option turned on)—see Chapter 22 to see how this function works in practice.

Data Sorting

Another aspect of formatting data in your worksheet is understanding how to arrange and sort your data. The menu to sort your data is not located in the *Home* tab of the ribbon but in the *Data* tab. There are four primary buttons in this area of the ribbon:

1. The *A → Z shortcut* automatically sorts the data from smallest to largest.

2. The *Z → A shortcut* automatically sorts the data from largest to smallest.

3. The *Sort* button lets us specify the direction to sort the data, which variable (or variables) to sort by, and whether to sort the data from left-to-right rather than up-to-down.

4. The *Filter* button lets us add sorting and filter options to the cells at the top of our data columns. This button can be especially useful if we want to make some quick sorting and filtering changes to our data (Figure 4.24).
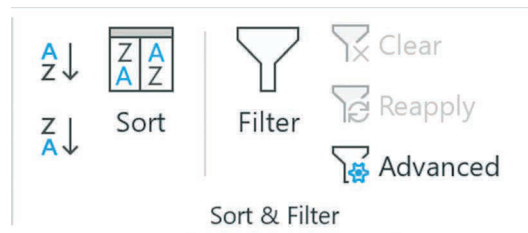
FIGURE 4.24

Number Formats

The last part of this section of the *Home* tab is the number formatting area. The drop-down menu with the word "General" in it has default formatting options included. The buttons below ($, %, and ,) format the numbers in your cells. The buttons to the right with the zeros add or subtract decimals from the data.

We can access more number formatting options by using the formal *Number Format* menu. Right-click on a cell and select the *Format Cells* option. In the first area of the menu (under *Number*), a list of formatting options, including "Number," "Currency," "Percentage," and "Scientific," appears. Each area lets us apply various formatting options to include different symbols, decimal places, commas, and so on.

🗿 The *Custom* area of the *Format Cells* menu is particularly powerful. Here, we can add custom formats to the entries in the spreadsheet cells or graphs. With this tool, we can change how the number looks without changing any data. Custom number formats have a specific structure with four arguments, separated by semi-colons:

   a. Positive values

   b. Negative values

   c. Zero values

   d. Text values

We don't need to include all four arguments for every formula, and we can skip a format by simply adding another semi-colon. (Where we enter the custom number format will be called different things depending on the operating system and use case: on PCs, it will be a box called *Type* when working in a

Setting

Making

Moving

spreadsheet but *Format Code* when working in a graph; on Macs, the box is always called *Type*.)

I'll demonstrate how a custom number format works with the highlighted example in Figure 4.25.

$$\text{\#, \#\#0\_);[Red](\#, \#\#0)}$$

This format will style positive and negative values. The pound sign (#) is a place-holder for optional digits. The "#, ##0" format tells Excel to use a comma in the number and not to include decimals. The underscore tells Excel to add a space the width of whatever character comes next, so the "_)" character tells Excel to add a space the width of a closed parenthesis. The format for negative numbers is included after the semicolon and will color the numbers red ("[Red]") and wrap them in parentheses.
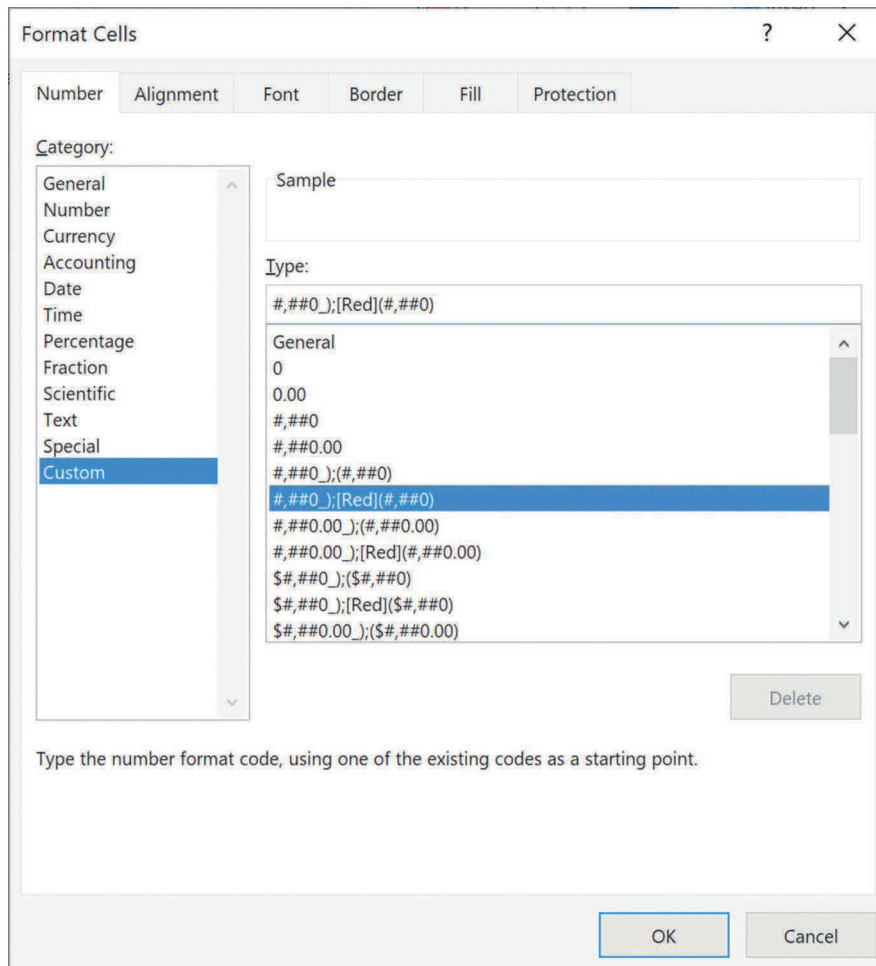
FIGURE 4.25

You can see a variety of formatting results in the image below. A much longer list as well as excellent details about how to create custom number formats can be found on the ExcelJet website (https://exceljet.net/custom-number-formats) (Table 4.2).

TABLE 4.2

Value	Code	Result	Notes
123456	#,##0_)	123,456	Positive number format
-123456	;[Red](#,##0)	(123,456)	Negative number format
18-Jan-21	yyyy	2021	Year only
18-Jan-21	mmmm-yyyy	January-2021	Month and year
15.5	##.#" years old"	15.5 years old	Text for units
155	#,##0	155	Positive number format
1550	#,##0	1,550	Positive number format
15500	#,##0	15,500	Positive number format
12000	0,"K"	12K	Number in thousands
65	;;;		Hide the content
100	\VA	VA	Escape character
125	[Blue][>100]#,##0	125	Conditional format

Pulling It All Together

Now that we've gone over all of these menus, let's create a basic column chart. We're not going to add fancy formatting or complex parts, but we'll walk through a few of the basic chart formatting options so you can become familiar with how the more advanced tutorials will look.

For this example, we'll use data from the World Bank to make a paired column chart of gross domestic product (GDP) measured in thousands of dollars for seven countries in 2010 and 2020.

First, we select the data in cells A1:E8 and use the *Insert* tab to create a *Clustered Column Chart*, which is the first column chart in the top-left menu and is labeled as *2-D Column* (Figure 4.26).

Notice the space between the columns? There's nothing objectively wrong with that space, but personally, I'm not a fan. So let's tighten it up. Select the bars, right-click, select the *Format Data Series* option in the drop-down menu, and change the *Series Overlap* to 0% and the *Gap Width* in to 100% in the menu. In the tutorial chapters, these directions will look something like: right-click on the bars > *Format Data Series* > *Series Overlap* to 0% and *Gap Width* to 100% (Figure 4.27).

Setting
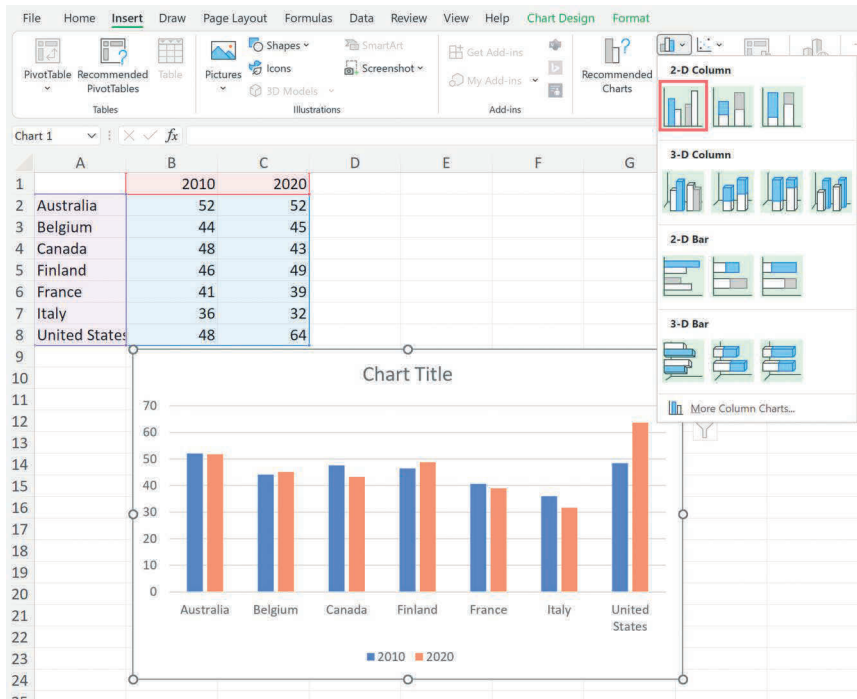
Making
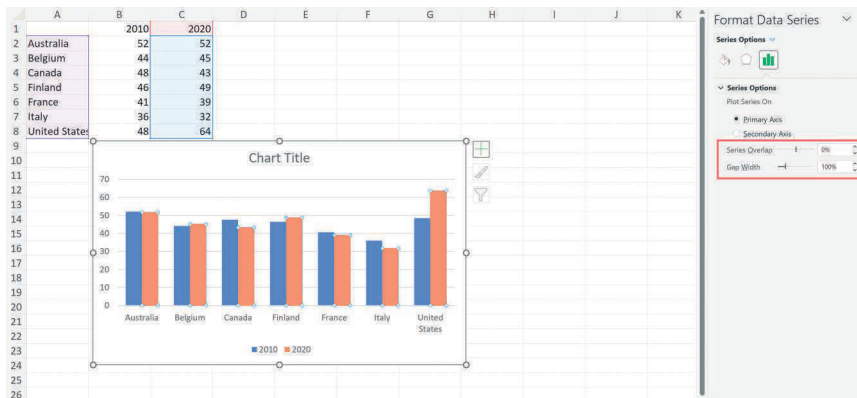
Moving

FIGURE 4.26

FIGURE 4.27

Let's adjust the formatting of the numbers on the y-axis. Select the y-axis labels and right click: *Format Axis > Number > Currency.* Again, this notation is intended to lead us from one step to the next. In the *Currency* menu, select the dollar sign ($) in the drop-down menu (if it wasn't automatically selected) (Figure 4.28).
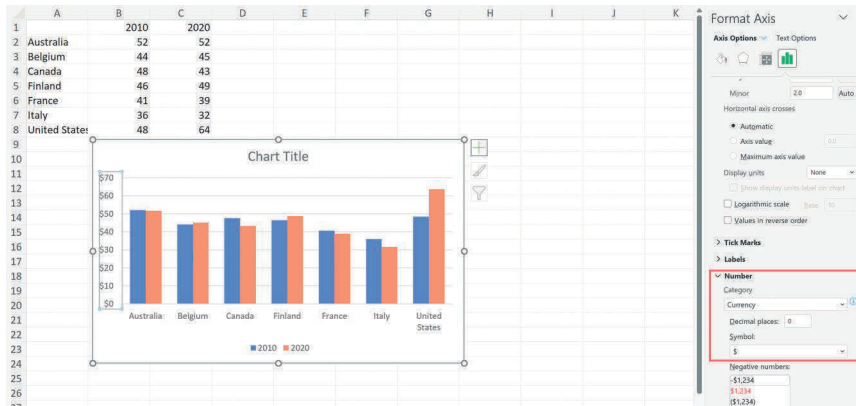
FIGURE 4.28

Just for fun, let's add some labels to each set of bars: right-click on the bars and select the *Add Data Labels* option in the menu. Because the numbers in the worksheet cells are shown as whole numbers, the data labels will appear as whole numbers. Let's modify the format in the chart by selecting the data labels, right-clicking, and *Format Data Labels > Number*. Now, set the number of decimal places to 1 in the box (Figure 4.29).
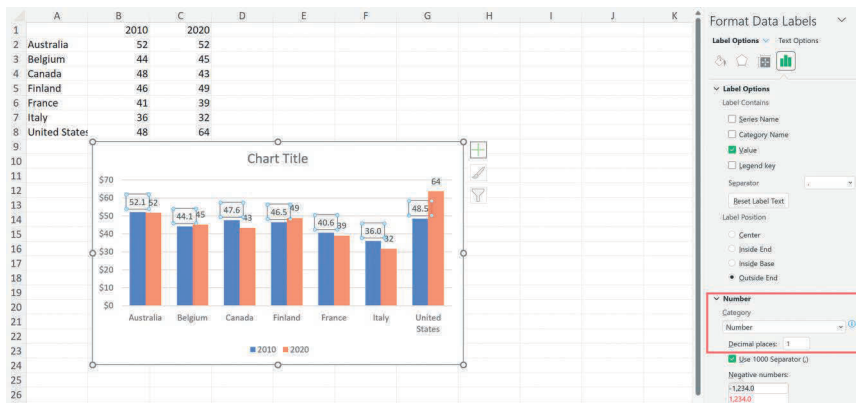
FIGURE 4.29

Let's move the legend to the top: select the legend, right-click, and in the *Legend Options* menu, select the *Top* option (Figure 4.30).

Sorting the bars based on the data values is usually better than the default alphabetical sorting. We can make this change in the worksheet by selecting the data (cells C1:C8), navigating to *Data > Sort*, and sorting on the 2020 per capita GDP values in column C (Figure 4.31).

Setting

Making

Moving

FIGURE 4.30

Setting

FIGURE 4.31

Making

Moving
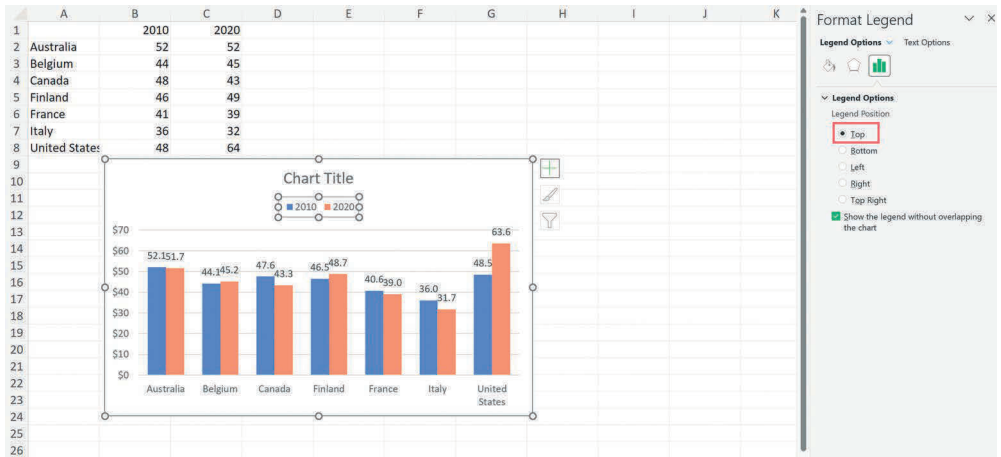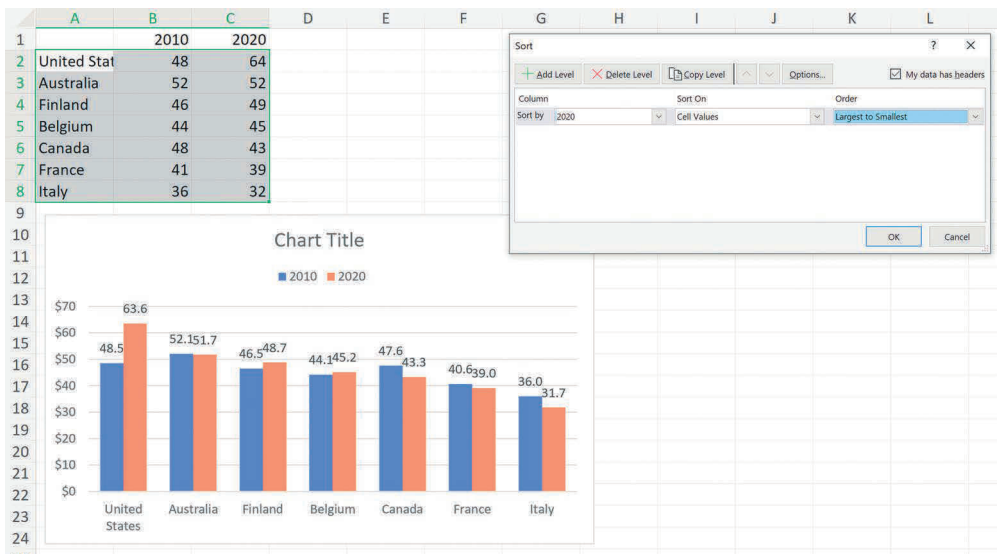
Finally, we can make a few finishing changes to the chart. I might change the size of the chart (3x5 inches is pretty small) and insert a title by editing the default *Chart title* area. Personally, I would also change the colors of the bars and use a different font to get away from the default settings. If I were including labels above each bar, I would delete the y-axis labels and the horizontal gridlines (by simply selecting and deleting) because they are redundant with the visible data labels (Figure 4.32).

Per capita GDP in the United States is more than double per capita GDP in Italy in 2020
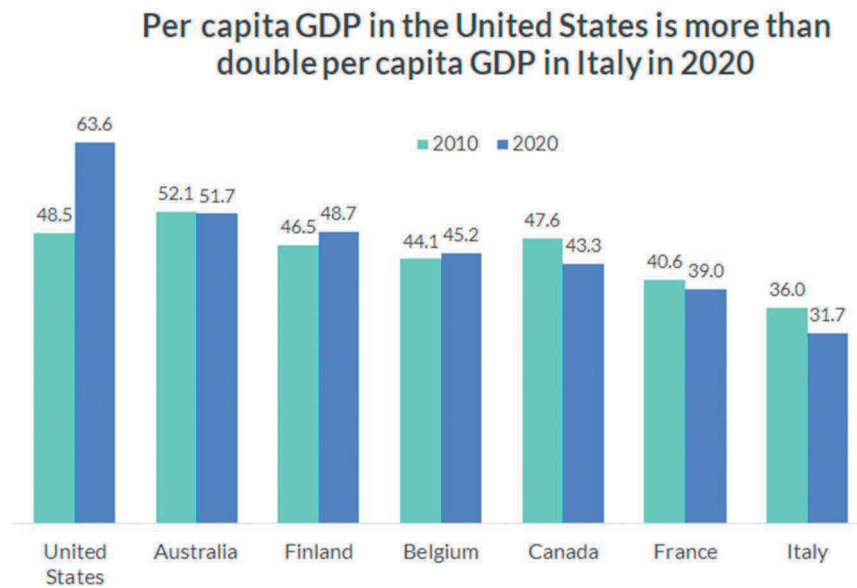
FIGURE 4.32

Going Forward

We've only just scratched the surface of what Excel can do. If you're new to Excel, this chapter has hopefully given you a peek into the power of how you can use the tool to analyze, organize, format, and visualize your data. If you've been using Excel for a while, you've likely seen a few techniques and strategies that can make your work in Excel more productive and efficient. Of course, no single chapter—or even a single book—can teach you everything there is to learn about Excel. A lot of Excel knowledge comes with time and practice (and patience). I've included a list of some of my favorite data visualization websites in Appendix 3 at the end of the book to help you continue your learning journey.

Setting

Making

Moving

Fundamental Formulas Used in Building Your Graphs

Throughout the tutorials in this book, you'll see the heavy use of formulas to arrange and organize the data for your graphs. I will often read in my full dataset to Excel before pulling the specific data points or series I need for the graph, but the full dataset remains for me to update, reference later on, pull additional data points or series, or make other changes.

In this section, I define and show examples for 17 Excel formulas that I use regularly in my work and that you will see throughout the tutorials. There is no way to show every Excel formula here—nor is it really necessary—but these formulas should serve you well when using the tool.

For each formula, I demonstrate how it works using a simple dataset of population estimates for 14 countries around the world (Table 5.1).

Before we get into specific formulas, it's worth understanding how formulas are copied and pasted within an Excel spreadsheet. To insert a formula in a cell, start by typing an equal sign (=). From there, Excel will prompt different formula options as we type. In the pop-up that appears, we can click on the linked (blue) text and a separate window will open with an Excel help window (Figure 5.1).

We can edit the formulas in the "formula bar," which is the white text box above the worksheet. Excel has a number of built-in reference tools to help us find the right formula for our purposes.

TABLE 5.1

Region	Country	Population, 2020
Africa	Ethiopia	114,963,583
Asia	Vietnam	97,338,583
Asia	Turkey	84,339,067
Asia	Thailand	69,799,978
Africa	South Africa	59,308,690
Europe	Spain	47,351,567
South America	Argentina	45,376,763
South America	Peru	32,971,846
Asia	Nepal	29,136,808
Africa	Chad	16,425,859
Europe	Belarus	9,379,952
Europe	Denmark	5,831,404
Asia	Singapore	5,685,807
Asia	Mongolia	3,278,292
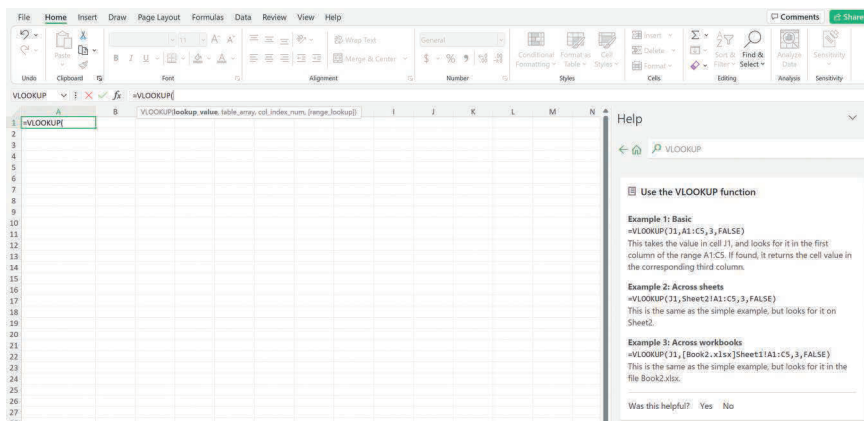
FIGURE 5.1

Absolute and Relative References

Once we have inserted a formula, we can copy and paste it anywhere in our Excel workbook (or even another workbook). The elements of the formula "follow" their position to the next location. Take a simple example: type a number in cell B1 and another cell C1. In cell A1, insert a simple formula that multiplies the values in those cells: =B1*C1.

If we copy the formula in cell A1 (CTRL+C) and paste it in cell A2 (CTRL+V), the formula updates automatically to the following: =B2*C2. In other words,

Excel is "looking" at the inputs to the formula relative to its position, so we call these inputs *relative references*.

This functionality can be problematic if we want to reference the input in a formula elsewhere in a worksheet. Say we want to multiply a whole bunch of numbers by the value in cell C1. Copying our original formula won't reference that cell unless we convert it to an *absolute reference*. To do so, we can "lock" the cell reference by inserting a dollar sign ($) before the row reference (number), column reference (letter), or both. (We can also use a keyboard shortcut to insert the absolute reference rather than typing it in—F4 or FN+F4 on PCs and CMD+T on Macs.)

Let's demonstrate with our population data. We have the region of the world in column A, the country name in column B, and the 2020 population in column C. We can convert the numbers to millions by dividing each number in column C by 1,000,000. To do so, we type 1,000,000 in cell E2. In cell D2, we type =C2/E2, which converts the population of Ethiopia to 115.0. If we copy and paste that formula, we get an error because the formula in cell D3 for Vietnam is =C3/E3, and cell E3 is empty. Again, the formula inputs are always relative to their initial position (Figure 5.2 and Table 5.2).

FIGURE 5.2

TABLE 5.2

	A	B	C	D	E	F
1	Region	Country	Population, 2020	Millions		Millions
2	Africa	Ethiopia	114,963,583	115.0	1,000,000	115.0
3	Asia	Vietnam	97,338,583	#DIV/0!		97.3
4	Asia	Turkey	84,339,067	#DIV/0!		84.3
5	Asia	Thailand	69,799,978	#DIV/0!		69.8
6	Africa	South Africa	59,308,690	#DIV/0!		59.3
7	Europe	Spain	47,351,567	#DIV/0!		47.4
8	South America	Argentina	45,376,763	#DIV/0!		45.4
9	South America	Peru	32,971,846	#DIV/0!		33.0
10	Asia	Nepal	29,136,808	#DIV/0!		29.1
11	Africa	Chad	16,425,859	#DIV/0!		16.4
12	Europe	Belarus	9,379,952	#DIV/0!		9.4
13	Europe	Denmark	5,831,404	#DIV/0!		5.8
14	Asia	Singapore	5,685,807	#DIV/0!		5.7
15	Asia	Mongolia	3,278,292	#DIV/0!		3.3

Setting

Making

Moving

If we change the formula to =C3/$E$3, we "lock" that cell reference. Now, when we copy the formula down the column, we still reference cell E3 in the denominator. For Thailand (in row 5), the formula is =C5/$E$3. We could have also typed =C3/E$3 because we are not copying across columns, so the column reference does not change.

You will see these *relative* and *absolute references* throughout the tutorials. I use them liberally to copy-and-paste formulas without worrying about making manual changes.

IF Formulas

```
=IF(Evaluation, if TRUE, if FALSE)
```

We start with one of the easiest formulas in Excel, the IF formula. The IF formula has three arguments: what we are going to evaluate, what happens if the evaluation is true, and what happens if the evaluation is false. As an example, let's create a binary variable that tells us whether a country has more than or fewer than 50 million people.

```
=IF(C2>50000000,1,0)
```

When we copy and paste this formula down column D, we'll get 1s for countries with more than 50 million people and 0s for countries with fewer than 50 million people.

We can also use this formula with text by placing what we want to come out of the formula in quotation marks. For example:

```
=IF(C2>50000000,"More than 50 million","Fewer than
                 50 million")
```

When we copy and paste this formula down column E, we get the relevant text in each cell (Figure 5.3 and Table 5.3).
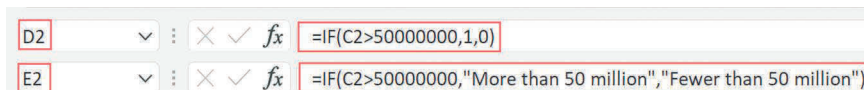
FIGURE 5.3

TABLE 5.3

	A	B	C	D	E
1	**Region**	**Country**	**Population, 2020**	**IF-Number**	**IF-Text**
2	Africa	Ethiopia	114,963,583	1	More than 50 million
3	Asia	Vietnam	97,338,583	1	More than 50 million
4	Asia	Turkey	84,339,067	1	More than 50 million
5	Asia	Thailand	69,799,978	1	More than 50 million
6	Africa	South Africa	59,308,690	1	More than 50 million
7	Europe	Spain	47,351,567	0	Fewer than 50 million
8	South America	Argentina	45,376,763	0	Fewer than 50 million
9	South America	Peru	32,971,846	0	Fewer than 50 million
10	Asia	Nepal	29,136,808	0	Fewer than 50 million
11	Africa	Chad	16,425,859	0	Fewer than 50 million
12	Europe	Belarus	9,379,952	0	Fewer than 50 million
13	Europe	Denmark	5,831,404	0	Fewer than 50 million
14	Asia	Singapore	5,685,807	0	Fewer than 50 million
15	Asia	Mongolia	3,278,292	0	Fewer than 50 million

```
=AND(argument 1, argument 2,…)
```

Another simple formula: when using the AND function, Excel returns the word "TRUE" if every argument in the formula is true. For example, =AND(C2>50000000,A2="Asia") will return the word "TRUE" for Vietnam, Turkey, and Thailand as each country has more than 50 million people and is located in Asia, but "FALSE" for all other countries (Figure 5.4 and Table 5.4).

| D2 | ∨ ⋮ ✕ ✓ *fx* | =AND(C2>50000000,A2="Asia") |
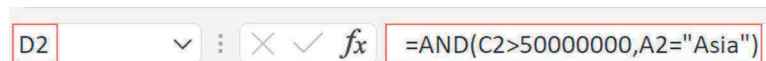
FIGURE 5.4

```
=OR(argument 1, argument 2,…)
```

Similar to the AND function, the OR function returns TRUE if *any* of its arguments are true and returns FALSE if *all* of its arguments are false. Here, we can insert =OR(C2>50000000,A2="Asia") and see TRUE for countries that have more than 50 million people *OR* are located in Asia (Figure 5.5 and Table 5.5).

```
SUMIF(range, criteria, sum _ range)
```

The SUMIF formula lets us sum a series of data values if it satisfies a certain condition. The three arguments in the SUMIF formula are:

TABLE 5.4

	A	B	C	D
1	Region	Country	Population, 2020	AND
2	Africa	Ethiopia	114,963,583	FALSE
3	Asia	Vietnam	97,338,583	TRUE
4	Asia	Turkey	84,339,067	TRUE
5	Asia	Thailand	69,799,978	TRUE
6	Africa	South Africa	59,308,690	FALSE
7	Europe	Spain	47,351,567	FALSE
8	South America	Argentina	45,376,763	FALSE
9	South America	Peru	32,971,846	FALSE
10	Asia	Nepal	29,136,808	FALSE
11	Africa	Chad	16,425,859	FALSE
12	Europe	Belarus	9,379,952	FALSE
13	Europe	Denmark	5,831,404	FALSE
14	Asia	Singapore	5,685,807	FALSE
15	Asia	Mongolia	3,278,292	FALSE

D2 f_x =OR(C2>50000000,A2="Asia")

FIGURE 5.5

TABLE 5.5

	A	B	C	D
1	Region	Country	Population, 2020	OR
2	Africa	Ethiopia	114,963,583	TRUE
3	Asia	Vietnam	97,338,583	TRUE
4	Asia	Turkey	84,339,067	TRUE
5	Asia	Thailand	69,799,978	TRUE
6	Africa	South Africa	59,308,690	TRUE
7	Europe	Spain	47,351,567	FALSE
8	South America	Argentina	45,376,763	FALSE
9	South America	Peru	32,971,846	FALSE
10	Asia	Nepal	29,136,808	TRUE
11	Africa	Chad	16,425,859	FALSE
12	Europe	Belarus	9,379,952	FALSE
13	Europe	Denmark	5,831,404	FALSE
14	Asia	Singapore	5,685,807	TRUE
15	Asia	Mongolia	3,278,292	TRUE

- *range*, which refers to the list of values that reference the condition;
- *criteria*, which is the value we set a condition on; and
- *sum_range*, which is the range of values across which we want to sum.

For this example, let's sum the population across the six Asian countries. We insert this formula in cell C2:

```
=SUMIF(A2:A15,"Asia",C2:C15)
```

The first argument is the *range*, which points us to the *Region* column. The second argument, the *criteria*, is what we are looking for in the range ("Asia"). And the final argument, *sum_range*, is the column of data we want to add up (C2:C15). I typed "Asia" in the formula for this example, but I could have typed it in a cell somewhere and referenced that cell in the formula (Figure 5.6 and Table 5.6).

| D2 | | ⋮ | ✕ ✓ | *fx* | =SUMIF(A2:A15,"Asia",C2:C15) |

FIGURE 5.6

TABLE 5.6

	A	B	C	D
1	Region	Country	Population, 2020	SUMIF
2	Africa	Ethiopia	114,963,583	289,578,535
3	Asia	Vietnam	97,338,583	
4	Asia	Turkey	84,339,067	
5	Asia	Thailand	69,799,978	
6	Africa	South Africa	59,308,690	
7	Europe	Spain	47,351,567	
8	South America	Argentina	45,376,763	
9	South America	Peru	32,971,846	
10	Asia	Nepal	29,136,808	
11	Africa	Chad	16,425,859	
12	Europe	Belarus	9,379,952	
13	Europe	Denmark	5,831,404	
14	Asia	Singapore	5,685,807	
15	Asia	Mongolia	3,278,292	

Setting

Making

Moving
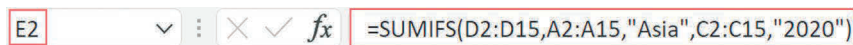
The SUMIF formula is valuable for adding data values automatically, especially if you might update or sort your data in different ways.

```
SUMIFS(sum _ range, criteria _ range1, criteria1,
[criteria _ range2, criteria2], ...)
```

Unlike the SUMIF formula, which only lets us apply one condition, the SUMIFS formula lets us apply multiple conditions. Let's expand our dataset to include 2 years of data and see what happens when we apply a revised formula:

```
=SUMIFS(D2:D15,A2:A15,"Asia",C2:C15,"2020")
```

Here, we are going to sum the population estimates in column D for countries in Asia *and* in 2020. Notice that the order of the arguments is slightly different than in SUMIF—we start with the column that we are going to sum, then the condition range, then the actual condition (Figure 5.7 and Table 5.7).

| E2 | | | ✕ ✓ | *fx* | =SUMIFS(D2:D15,A2:A15,"Asia",C2:C15,"2020") |

FIGURE 5.7

TABLE 5.7

	A	B	C	D	E
1	Region	Country	Year	Population	SUMIFS
2	Asia	Nepal	1960	10,105,060	280,614,436
3	Asia	Nepal	2020	29,136,808	
4	Asia	Vietnam	1960	32,670,048	
5	Asia	Vietnam	2020	97,338,583	
6	South America	Argentina	1960	20,481,781	
7	South America	Argentina	2020	45,376,763	
8	Asia	Thailand	1960	27,397,208	
9	Asia	Thailand	2020	69,799,978	
10	Asia	Turkey	1960	27,472,339	
11	Asia	Turkey	2020	84,339,067	
12	South America	Peru	1960	10,155,011	
13	South America	Peru	2020	32,971,846	
14	Europe	Spain	1960	30,455,000	
15	Europe	Spain	2020	47,351,567	

Setting

Making

Moving

You can see these arguments more clearly if I highlight the three parts of the formula:

```
=SUMIFS(D2:D15,A2:A15,"Asia",C2:C15,"2020")
```

The first part of the formula in green is the series we want to add up; the second section in blue defines that we want to search for "Asia" in cells A2:A15; and the third section in pink defines that we want to look for "2020" in cells C2:C15.

Using SUMIF rather than a simple SUM formula is valuable because it makes the task less prone to error (i.e., selecting the wrong cells) and is more flexible to the format of our data. If we were to sort these data by population, the result of the SUMIFS formula will remain the same, but the results of a regular SUM formula would be incorrect.

- AVERAGEIF(range, criteria, average _ range)
- AVERAGEIFS(average _ range, criteria _ range1, criteria1, [criteria _ range2, criteria2], ...)

AVERAGEIF and AVERAGEIFS work the same way as SUMIF and SUMIFS. In this example, we calculate the average population for all Asian countries in both 1960 and 2020 and for all Asian countries in 2020 only (Figure 5.8 and Table 5.8).

| E2 | ✓ | ⋮ | ✕ ✓ | *fx* | =AVERAGEIF(A2:A15,"Asia",D2:D15) |
| F2 | ✓ | ⋮ | ✕ ✓ | *fx* | =AVERAGEIFS(D2:D15,A2:A15,"Asia",C2:C15,"2020") |
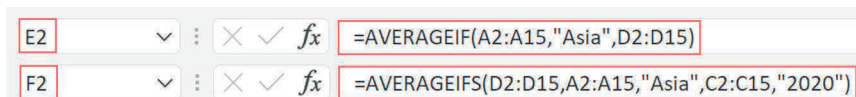
FIGURE 5.8

```
      Cell E2: =AVERAGEIF(A2:A15,"Asia",D2:D15)

  Cell F2: =AVERAGEIFS(D2:D15,A2:A15,"Asia",C2:C15,"2020")
```

- COUNTIF(criteria _ range, criteria1)
- COUNTIFS(criteria _ range1, criteria1, [criteria _ range2, criteria2]…)

TABLE 5.8

	A	B	C	D	E	F
1	**Region**	**Country**	**Year**	**Population**	**AVERAGEIF**	**AVERAGEIFS**
2	Asia	Nepal	1960	10,105,060	47,282,386	70,153,609
3	Asia	Nepal	2020	29,136,808		
4	Asia	Vietnam	1960	32,670,048		
5	Asia	Vietnam	2020	97,338,583		
6	South America	Argentina	1960	20,481,781		
7	South America	Argentina	2020	45,376,763		
8	Asia	Thailand	1960	27,397,208		
9	Asia	Thailand	2020	69,799,978		
10	Asia	Turkey	1960	27,472,339		
11	Asia	Turkey	2020	84,339,067		
12	South America	Peru	1960	10,155,011		
13	South America	Peru	2020	32971846		
14	Europe	Spain	1960	30455000		
15	Europe	Spain	2020	47351567		

The COUNTIF and COUNTIFS formulas are slightly different than SUMIF and AVERAGEIF because we only need to look at one condition and not the sum or average of the values in another column (Figure 5.9 and Table 5.9):

E2	⌄ ⋮ ✕ ✓ *fx*	=COUNTIF(A2:A15,"Asia")
F2	⌄ ⋮ ✕ ✓ *fx*	=COUNTIFS(A2:A15,"Asia",C2:C15,"2020")

FIGURE 5.9

```
Cell E2: =COUNTIF(A2:A15,"Asia")

Cell F2: =COUNTIFS(A2:A15,"Asia",C2:C15,"2020")
```

TABLE 5.9

	A	B	C	D	E	F
1	**Region**	**Country**	**Year**	**Population**	**COUNTIF**	**COUNTIFS**
2	Asia	Nepal	1960	10,105,060	8	4
3	Asia	Nepal	2020	29,136,808		
4	Asia	Vietnam	1960	32,670,048		
5	Asia	Vietnam	2020	97,338,583		
6	South America	Argentina	1960	20,481,781		
7	South America	Argentina	2020	45,376,763		
8	Asia	Thailand	1960	27,397,208		
9	Asia	Thailand	2020	69,799,978		
10	Asia	Turkey	1960	27,472,339		
11	Asia	Turkey	2020	84,339,067		
12	South America	Peru	1960	10,155,011		
13	South America	Peru	2020	32,971,846		
14	Europe	Spain	1960	30,455,000		
15	Europe	Spain	2020	47,351,567		

Find the Top and Bottom

- MIN(number1, [number2], ...)
- MAX(number1, [number2], ...)

The MIN (minimum) and MAX (maximum) formulas let us pull out the largest and smallest values in a data series (Figure 5.10 and Table 5.10).
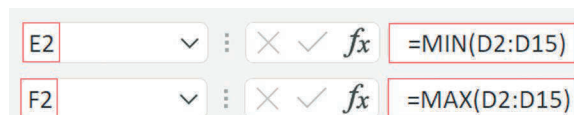
Cell E2: =MIN(D2:D15)

Cell F2: =MAX(D2:D15)

FIGURE 5.10

TABLE 5.10

	A	B	C	D	E	F
1	**Region**	**Country**	**Year**	**Population**	**MIN**	**MAX**
2	Asia	Nepal	1960	10,105,060	10,105,060	97,338,583
3	Asia	Nepal	2020	29,136,808		
4	Asia	Vietnam	1960	32,670,048		
5	Asia	Vietnam	2020	97,338,583		
6	South America	Argentina	1960	20,481,781		
7	South America	Argentina	2020	45,376,763		
8	Asia	Thailand	1960	27,397,208		
9	Asia	Thailand	2020	69,799,978		
10	Asia	Turkey	1960	27,472,339		
11	Asia	Turkey	2020	84,339,067		
12	South America	Peru	1960	10,155,011		
13	South America	Peru	2020	32,971,846		
14	Europe	Spain	1960	30,455,000		
15	Europe	Spain	2020	47,351,567		

Combine Pieces of Information

- CONCATENATE(text1, [text2],…)
- CONCAT(text1, [text2],…)
- Ampersand (&) symbol

We can combine (concatenate) data fields in Excel using a few methods. The CONCATENATE formula was the original way to combine data fields but has been replaced with the CONCAT function. Both work in the same way—the arguments inside the formula are joined together. A faster and a slightly easier way is to use the ampersand (&) symbol.

Using these techniques, we can combine data fields to create data labels or to use in another formula. With our country population dataset, we can create a label for each country in "country name, year" format by using:

```
Cell E2: =B2&", "&C2
```

This formula joins the value in cell B2 ("Nepal") with a comma and space (", ") and the value in cell C2 (1960). I often use this approach to create custom labels or specific search identifiers for a data value (Figure 5.11 and Table 5.11).

| E2 | | ▾ | ⋮ | ✕ ✓ | fx | =B2&", "&C2 |

FIGURE 5.11

TABLE 5.11

	A	B	C	D	E
1	Region	Country	Year	Population	Label
2	Asia	Nepal	1960	10,105,060	Nepal, 1960
3	Asia	Nepal	2020	29,136,808	Nepal, 2020
4	Asia	Vietnam	1960	32,670,048	Vietnam, 1960
5	Asia	Vietnam	2020	97,338,583	Vietnam, 2020
6	South America	Argentina	1960	20,481,781	Argentina, 1960
7	South America	Argentina	2020	45,376,763	Argentina, 2020
8	Asia	Thailand	1960	27,397,208	Thailand, 1960
9	Asia	Thailand	2020	69,799,978	Thailand, 2020
10	Asia	Turkey	1960	27,472,339	Turkey, 1960
11	Asia	Turkey	2020	84,339,067	Turkey, 2020
12	South America	Peru	1960	10,155,011	Peru, 1960
13	South America	Peru	2020	32,971,846	Peru, 2020
14	Europe	Spain	1960	30,455,000	Spain, 1960
15	Europe	Spain	2020	47,351,567	Spain, 2020

LOOKUP Functions

- VLOOKUP(lookup _ value, table _ array, col _ index _ num, [range _ lookup])

- HLOOKUP(lookup _ value, table _ array, row _ index _ num, [range _ lookup])

- XLOOKUP(lookup _ value, lookup _ array, return _ array, [if _ not _ found], [match _ mode], [search _ mode])

Setting

Making

Moving

We now get into the LOOKUP formulas, which are extremely help-
ful for simplifying or organizing our data. There are three primary
LOOKUP formulas in Excel:

– VLOOKUP is a vertical lookup function that looks up and down
columns.

– HLOOKUP is a horizontal lookup function that looks across
rows.

– XLOOKUP is a newer function that looks across rows and
columns.

The formulas all work similarly, and the VLOOKUP formula
appears throughout the tutorials, so we'll focus on it here. There are
four arguments in VLOOKUP:

– The *lookup_value* is what we are looking for in our dataset, such
as a country name.

– The *table_array* is the table or set of columns where we are look-
ing to find the data value or values we want. The thing we are
looking for (e.g., country name) must exist in the first column of
the table.

– The *col_index_num* is the column number we want to look
through in our table. The column in which we are looking for
the *lookup_value* counts as the first column.

– The *range_lookup* has two options:

– Exact match (0 or FALSE) searches for the exact value in the
first column.

– Approximate match (1 or TRUE) assumes the first column
in the table is sorted and searches for the closest value. This
option is the default in Excel, but we don't always want to find
an approximate match so it's always worth specifying.

Admittedly, there are a lot of parts to the VLOOKUP formula,
so let's break it down:

```
=VLOOKUP("Peru",A2:B15,2,0)
```

We've got our four pieces:

– The first argument is the thing we are looking for: the word
"Peru".

Setting   Making   Moving

– The second argument is the data table range *where* we are look-ing (A2:B15). The thing we are looking for ("Peru") must appear in the first column of this range.

– The third argument is the column number from which we want to pull our result—the first column contains the list of things we are looking for (country names) and extends to the right. In this case, we want to get the population value for Peru, so we want to pull the data from the second column in the range (column B).

– The last argument is the *exact* or *approximate* match identifier. In this case, we want to find the value for Peru *exactly*, so we insert the number 0 (or the word FALSE if you like) (Figure 5.12 and Table 5.12).

C2   ✓ ⋮ ✕ ✓ *fx*   =VLOOKUP("Peru",A2:B15,2,0)
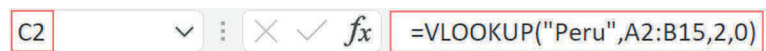
FIGURE 5.12

TABLE 5.12

	A	B	C
1	Country	Population, 2020	VLOOKUP
2	Ethiopia	114,963,583	32,971,846
3	Vietnam	97,338,583	
4	Turkey	84,339,067	
5	Thailand	69,799,978	
6	South Africa	59,308,690	
7	Spain	47,351,567	
8	Argentina	45,376,763	
9	Peru	32,971,846	
10	Nepal	29,136,808	
11	Chad	16,425,859	
12	Belarus	9,379,952	
13	Denmark	5,831,404	
14	Singapore	5,685,807	
15	Mongolia	3,278,292	

Setting

Making

Moving

What about using the *approximate* match? Let's say we want to add an indicator to our dataset that classifies countries as "small," "middle," or "large." We can accomplish this task by using a VLOOKUP formula.

First, we must set up a smaller "lookup" table to the side, which defines our conditions for what constitutes a "small," "medium," or "large" country. Then, in column C (starting in cell C2), we insert this formula:

```
=VLOOKUP(B2,$E$2:$F$4,2,1)
```

Our formula starts by looking at cell B2, the population of Ethiopia. With the second argument (i.e., $E$2:$F$4), we instruct our formula to compare that population number in cell B2 to the thresholds we created in the blue-shaded lookup table in columns D and E (Figure 5.13 and Table 5.13). (Remember, the thing we are looking for must be in the first column.) Here, we use *absolute references* (the dollar signs) so the reference to the lookup table doesn't change when we copy the formula down column C. The third argument, 2, specifies that we are going to extract the word in the second column from our lookup table) using an *approximate match* (the 1 in the last argument).


```
D2          ∨  ⋮  ✕ ✓ ƒx  =VLOOKUP(C2,$E$2:$F$4,2,1)
```

FIGURE 5.13

But what does the *approximate match* do? In short, it takes the data (here in column E) and builds ranges with it. Countries with populations between 0 and 50 million people are considered "small," countries with 50–100 million people are considered "middle," and countries with 100 million or more people are classified as "large." When the formula looks in the table for the population of Ethiopia (cell C2), it compares the 115 million data point to each of those thresholds: to 0 million—greater than that; to 50 million—greater than that; and finally 100 million or more. The formula sees that Ethiopia's population exceeds the 100 million threshold, so it inserts the text value in the second column ("Large") into the cell.

Setting

Making

Moving

TABLE 5.13

	A	B	C	D	E	F
1	Region	Country	Population, 2020	VLOOKUP		
2	Africa	Ethiopia	114,963,583	Large	0	Small
3	Asia	Vietnam	97,338,583	Middle	50,000,000	Middle
4	Asia	Turkey	84,339,067	Middle	100,000,000	Large
5	Asia	Thailand	69,799,978	Middle		
6	Africa	South Africa	59,308,690	Middle		
7	Europe	Spain	47,351,567	Small		
8	South America	Argentina	45,376,763	Small		
9	South America	Peru	32,971,846	Small		
10	Asia	Nepal	29,136,808	Small		
11	Africa	Chad	16,425,859	Small		
12	Europe	Belarus	9,379,952	Small		
13	Europe	Denmark	5,831,404	Small		
14	Asia	Singapore	5,685,807	Small		
15	Asia	Mongolia	3,278,292	Small		

I use the VLOOKUP formula a lot in this book, so it's worth familiarizing yourself with how it works. Absolute references can be added to any of its arguments, and we can embed many of the formulas shown above—IF, AND, OR, and the ampersand symbol—inside the VLOOKUP. To demonstrate, let's extend our previous example with the following formula in cell F2:

```
=VLOOKUP(G2&" "&G3,A2:E15,5,0)
```

Here, we have a table with a new *Label* column that combines the country name and the year of data. Our new formula combines the values in cells G2 and G3 and separates them with a space (e.g., "Vietnam 2020"). It then looks through our *Label* column for the same combined value and pulls out the data point from the fifth column of the designated range for the specific country-year combination specified in cells G2 and G3 (Figure 5.14 and Table 5.14).

F2	∨ ⋮ ✕ ✓ *fx*	=VLOOKUP(G2&" "&G3,A2:E15,5,0)

FIGURE 5.14

Setting

Making

Moving

TABLE 5.14

	A	B	C	D	E	F	G
1	Label	Region	Country	Year	Population	VLOOKUP	
2	Nepal 1960	Asia	Nepal	1960	10,105,060	97,338,583	Vietnam
3	Nepal 2020	Asia	Nepal	2020	29,136,808		2020
4	Vietnam 1960	Asia	Vietnam	1960	32,670,048		
5	Vietnam 2020	Asia	Vietnam	2020	97,338,583		
6	Argentina 1960	South America	Argentina	1960	20,481,781		
7	Argentina 2020	South America	Argentina	2020	45,376,763		
8	Thailand 1960	Asia	Thailand	1960	27,397,208		
9	Thailand 2020	Asia	Thailand	2020	69,799,978		
10	Turkey 1960	Asia	Turkey	1960	27,472,339		
11	Turkey 2020	Asia	Turkey	2020	84,339,067		
12	Peru 1960	South America	Peru	1960	10,155,011		
13	Peru 2020	South America	Peru	2020	32,971,846		
14	Spain 1960	Europe	Spain	1960	30,455,000		
15	Spain 2020	Europe	Spain	2020	47,351,567		

```
XLOOKUP(lookup _ value,lookup _ array,return _
array,[if _ not _ found], [match _ mode], [search _ mode])
```

XLOOKUP works similarly to VLOOKUP, but it's worth defining because I expect it to become the go-to formula in the coming years. XLOOKUP is more flexible than VLOOKUP for two main reasons. First, the value you are looking for does not need to appear in the first column of the data table. Second, XLOOKUP can return an *array* with multiple items, so one formula can return multiple fields.

Let's say our data table is rearranged so the country names are in the far-right column. With XLOOKUP, we can conduct a simple lookup:

```
=XLOOKUP(B2,D8:D14,B8:B14)
```

The formula looks for the value in cell B2 (Turkey) in the moved country column, cells D8:D14. It then looks for the value in cells B8:B14 that corresponds to our lookup value, Turkey.

Setting

Making

Moving

In row 5, we have another XLOOKUP formula, which pulls both population estimates into new cells. That formula is:

=XLOOKUP(B5,D8:D14,B8:C14)

By setting the last argument as cells B8:C14, Excel returns the corresponding values from each column when we click the Enter or Return key. VLOOKUP, by comparison, would require us to add a separate column reference that would need to change from column to column (Figure 5.15 and Table 5.15).
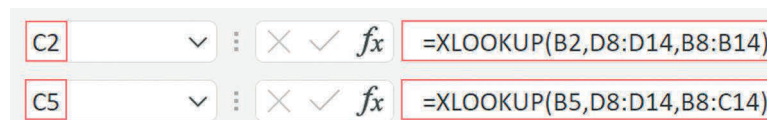

FIGURE 5.15

TABLE 5.15

	A	B	C	D
1		**Country**	**Population, 1960**	
2		Turkey	27,472,339	
3				
4		Country	Population, 1960	Population, 2020
5		Peru	10,155,011	32,971,846
6				
7	**Region**	**Population, 1960**	**Population, 2020**	**Country**
8	South America	20,481,781	45,376,763	Argentina
9	Asia	10,105,060	29,136,808	Nepal
10	South America	10,155,011	32,971,846	Peru
11	Europe	30,455,000	47,351,567	Spain
12	Asia	27,397,208	69,799,978	Thailand
13	Asia	27,472,339	84,339,067	Turkey
14	Asia	32,670,048	97,338,583	Vietnam

Wrap-Up

There are two other lookup formulas that many people like to use in Excel: INDEX and MATCH. Both formulas work similarly to the LOOKUP formulas but are faster (in terms of computer processing speed) and more flexible in terms of how the data table can be organized. But they also have more

arguments, which can make them difficult to build and understand. I consider myself a "LOOKUP" person, so I don't present these formulas here.

Excel has a myriad of formula types and a variety of ways to construct and combine formulas. Some formulas can pull data directly from the web (e.g., STOCKHISTORY), and others are *arrays* that can hold a collection of values, not just in a single cell. The Microsoft Office suite also houses the Visual Basic for Applications (VBA) programming language, which can be used to automate certain tasks or expand Excel's general capabilities. In short, the formulas and approaches that I use in this book may not be the right formulas or even what the most experienced Excel users would choose to use. But I have used these formulas for many years, and they have served me well to analyze, extract, organize, and visualize my data.

Setting

Making

Moving