Excel Primer

A Hands-on Introduction to Microsoft Excel
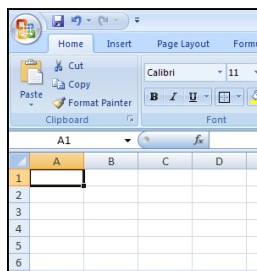

Table of Contents

Introduction

A.   Basics

   A-1. The Structure of Excel

   A-2. Formatting & Editing

      Enter numbers and text into cells,

         and change the formatting.

      Cut, copy, and paste cells.

      Alter the look of a whole spreadsheet.

   A-3. Formulas

      Write working formulas.

B.   Data Analysis

   Understand the data first.

   Import data into Excel and do basic clean-up.

   Derive new numbers from data.

   Sort, filter, graph, and pivot.

Files Included

This PDF:	Excel_Primer.pdf
Two Excel files:	Excel_Primer-Basics.xlsx
	Excel_Primer-Data_Analysis.xlsx
One more PDF:	geographic_diversification.pdf

Introduction

Who Should Use This Primer[1]

You should be one of the following people:

(a) **True beginner**—you've hardly ever opened Excel, and now you need to know everything.

(b) **Dabbling user**—you've got some "hook or crook" knowledge, but you've never been trained.

(c) **Curious expert**—you know Excel well, but you're always looking to learn new tricks.

If you are (b) or (c), just skim through this primer, especially the first parts. Use the whole thing as a refresher, and jump in whenever you spot something new.

If you are a true beginner, though, definitely take your time. Work your way forward methodically.

What is Excel?

Excel is a powerful tool to have in your toolkit, one that you'll use throughout your career. Part of the Microsoft Office suite of products, Excel is the dominant **spreadsheet** program on personal computers.[2] A spreadsheet program is the following four things, all at once:

1. **Table-maker**

   Excel gives you a big, stretchy piece of blank gridline paper.

   The program is not a true word-processor like Word, but you can just use Excel to make pretty little tables like this one:

My Portfolio	
Fund	Amount
VOO	$1,000
VWO	$300
VSS	$400

---

[1] You can pronounce the word *primer* (meaning an introductory textbook) either as "PRIMM-er" (short *i*) or as "PRIME-er" (long *i*). According to some folks, the short *i* is American, the long *i* is British. The long *i* pronunciation will probably take over, given the spelling of the word. As for the other meanings of *primer* (paint, explosive), always say "PRIME-er."

[2] The *first* spreadsheet program was called VisiCalc, and Lotus 123 was pre-eminent for a time. First-mover advantage can be illusory.

2. Calculator

Excel can add, subtract, multiply, divide, and do much, much more with numbers. If you want that column of figures to add up to a total, Excel can make it happen. In fact, Excel *should* make it happen. Alaskan sled dogs want to run. Excel wants to add up your numbers.

My Portfolio	
Fund	Amount
VOO	$1,000
VWO	$300
VSS	$400
Total	$1,700

3. Data-cruncher

Excel really shines when you use it to analyze lots and lots of numbers. With over 1 million rows and 16,000 columns, Excel has space to crunch almost any data you'll encounter. With Excel, you can turn a ton of marketing surveys, market prices, or machine readings into a few relevant figures—and in this way, draw sensible conclusions.

4. Model-maker

You can even use Excel to build a working model of the world—well, some limited portion of it. By running the model in different ways and studying the results, you can peer into the cloudy future. This might sound like psychic hooey. But you have to make business decisions in the *present*, and you need all the foresight you can muster. Excel can give you something of a crystal ball.

It's worth distinguishing these last two roles:

	Data Analysis	Modeling
What	• Draw insights from raw information	• Answer "what if?" • Predict the future
Business functions	• Statistics, Operations, Marketing	• Finance, New Product Development, Operations
How you'll recognize	• Rows & rows of numbers	• A blank spreadsheet—you have to build up from scratch
Skills you'll need	• Formatting • Formulas • Charts & graphs • Filter & sort • Pivot tables • Data cleaning	• Formatting • Formulas • Model layout • Goal Seek & Solver • Scenario and sensitivity analysis

Once you get the hang of Data Analysis and Modeling, you'll start using both together. For example, you might take insights from crunched data and feed them into a model as input. Or a model might spit out a bunch of numbers that you analyze as if they were data from the real world.

How to Use This Primer

The rest of this primer has two sections:

    A. Basics
    B. Data Analysis

As you work through each section in order, you should have the corresponding Excel file open[3]:

    A. Excel_Primer-Basics.xlsx
    B. Excel Primer - Data Analysis.xlsx

Within each Excel file (called a **workbook**), there will be several **worksheets**, like pages in a book. The different sheets are indicated by tabs at the bottom of the window.

In the Basics file, the first tab on the left is called "1. Start Here." Click on that tab to begin with. As you proceed, you'll move to other tabs. You can always check your work against example "Outcome" sheets.

Within each section of this primer, these icons will help you navigate:

Ⓣ **Task** to do
⌐ **Process** to follow

This is a hands-on primer. To learn how to do a task, simply follow the associated process, step by step. Overall, this primer will likely take you a few hours. Take breaks and don't do it all at once.

There is a little overlap between the content of this primer and that of the Excel section of our book *Case Studies & Cocktails*. However, this primer provides a more extensive and hands-on introduction to the basics of Excel and to data analysis. Modeling is not covered here, to keep the size of this primer reasonable.

Excel 2007 for PCs is the version of the program on which this primer is based. If you are using another version of Excel, in particular for Macs, the exact commands will vary. However, the principles and even certain specifics will be the same.

Finally, we assume that you're familiar with Microsoft Word. Some of the commands are the same in both Word and Excel.

In particular, if you ever make a mistake, just hit Undo (Ctrl-Z on PCs, Command-Z on Macs) and try again.

---

[3] We assume that you have Excel installed on your computer. Double-click any file to open it. If you have an older version of Excel and the .xlsx files don't work, try opening the .xls versions.
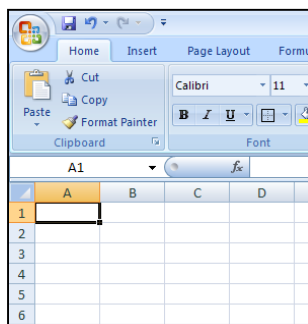
A. Basics

Contents

A-1. The Structure of Excel

When you open a new Excel file, you see a rectangular grid with lots of little boxes.

	A	B	C	D	…
1					…
2					…
3					…
4					…
…	…	…	…	…	…

Here's a close-up of the real thing, together with a bit of the **ribbon** at the top, where various controls and commands live.


Each little box is called a **cell**. Each cell is named by its **column** and **row**.

	A	**B**	C
1			
2			
3		Cell **B3**	
4			

Columns are labeled A through Z, then AA, AB, etc., going to the right. Rows are numbered from 1 on up, as you move down the spreadsheet.

So every cell has a natural name, starting with a letter or letters (for the column) and ending with a number (for the row). It's like the game Battleship, or an x-y coordinate plane. You can think of cell A1 as the origin at the upper left.

The whole grid constitutes a **worksheet**, a **spreadsheet**, or just a **sheet**. The name of the worksheet is on the active tab at the bottom of the window.
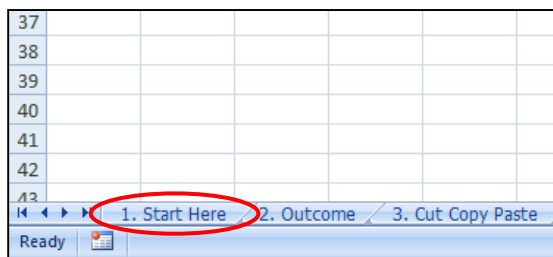
You should be on the blank sheet named "1. Start Here."

As mentioned in the introduction, a **workbook** typically contains several worksheets on different tabs. An Excel file *is* a workbook.

An individual cell can contain one of three things:

1)   A **number**.
2)   Some **text**.
3)   A **formula**, which computes something on the basis of other cells and shows you the result.

Formulas are what give Excel the power to crunch numbers. Whether you are analyzing data or building a model, your spreadsheet will have formulas in it.

We'll come back to formulas in part A-3. In the meantime, let's focus on how to format and edit worksheets.

A-2. Formatting & Editing

In this section, you'll enter data, change cell formats, cut/copy/paste data, and alter the look of a whole spreadsheet.

Ⓣ *Task:* **Enter numbers and text into cells, and change the formatting.**

Where to work: File – **Excel_Primer-Basics**
                       Sheet – **1. Start Here**   ← Do your work on this sheet.

Example outcome: Sheet – **2. Outcome**   ← Check your work against this sheet.

⌐ *Process*

Enter a number

1. Open the file *Excel_Primer-Basics* if you don't already have it open. If the sheet labeled "1. Start Here" isn't displayed, click on that tab at the bottom. You should see a blank spreadsheet.

2. Click directly on cell B3. You'll see the **active cell**, B3, surrounded by a black box. The column header (B) and the row label (3) are also highlighted, helping you see where you are.
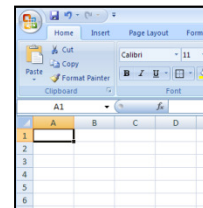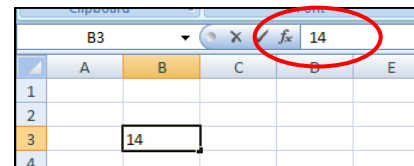

   From now on, whatever you type goes into cell B3, overwriting whatever's already there.

3. Type "14" but don't hit Enter yet.

   On a PC, you'll notice the number 14 appearing not only in cell B3 itself, but also above in the **formula bar**, which lives between the spreadsheet below and the ribbon above.
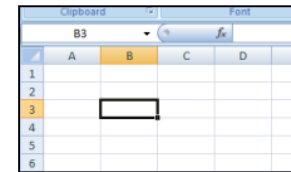
   The formula bar shows you the contents of a cell. It also shows you what cell you're on.

4. Hit Enter.

   The contents of cell B3 "lock" on 14, and the active cell becomes B4. This way, you can quickly enter a column of numbers without overwriting old numbers. The Enter key not only enters data; it takes you to the next row.

   Instead of hitting Enter, you could have clicked anywhere else on the spreadsheet. Excel would assume that you were done with cell B3 and would enter the number 14 into it. You could also hit Tab to enter and move right.
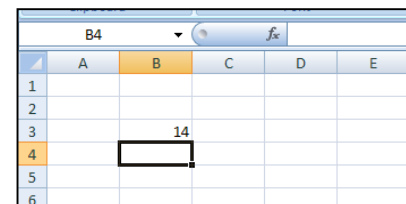
Format the contents of a cell

5. Click on cell B3 again. If the ribbon (at top) is not on the Home tab, click on Home at the top of the screen, so that the Home ribbon is showing.

   The Home ribbon contains various common commands, including formatting controls. Focus on the Font area first.

6. On the ribbon, click the buttons labeled **B** (**Bold**) and *I* (*Italics*), putting the 14 in **bold italics**.

   Notice that while the contents of the cell itself become bold italics (*14*), the formula bar above still displays a plain old 14. There is a difference between what a cell displays (*14*) and what it *actually* contains (14), which is revealed to you in the formula bar.

   In this case, the difference is cosmetic, but we'll see bigger differences down the road.

7. Hover over other Font and Alignment controls to see what they do.

   Click on cell B5. Give it a solid **border** all around and **fill** it with bright yellow, using the buttons circled to the right.

   Type "12" into the cell and turn the number red. Also **center** the 12, using the buttons circled to the right.

   Notice that, as always, the 12 in the formula bar stays unformatted. The true contents of the cell are not altered when you change the formatting.
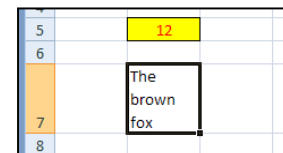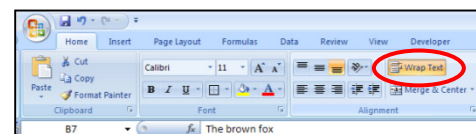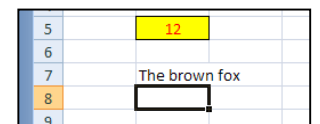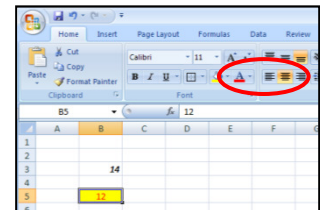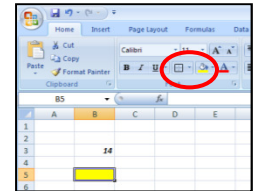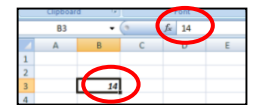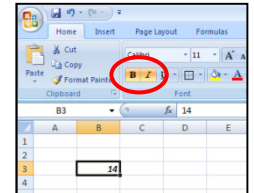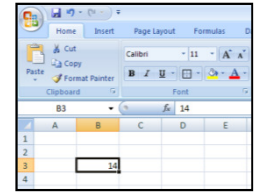
Wrap text

8. Click on cell B7. Type "The brown fox" and hit Enter.

   The contents of B7 spill over, covering up part of C7.

9. Click on B7 again.

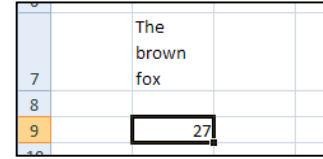   In the Alignment portion of the ribbon, click on **Wrap Text**.

   Cell B7 becomes much taller, as the words wrap around to fit within the cell.
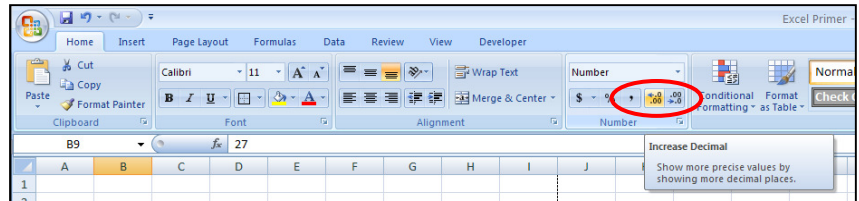
Adjust number formats

10. Click on cell B9. Type "27" and hit Enter.

    Click on B9 again. Alternatively, use the keyboard Up Arrow (↑) to move the active cell box back to B9.
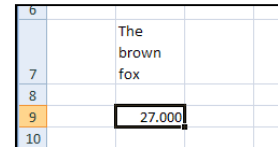
    Now look at the Number format controls in the ribbon.

    Click the **Increase Decimal** button (📈) to increase the number of decimal places.
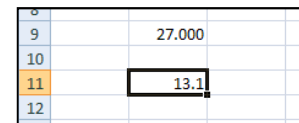
    Click three times, so that 27.000 is displayed in the cell.

    Notice yet again that the cell displays 27.000, but the cell's true value is simply 27, as revealed in the formula bar.
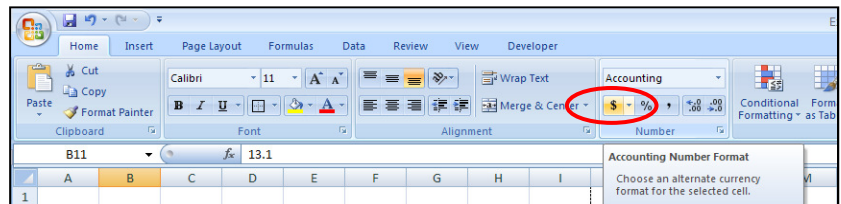
11. Click on cell B11, type in "13.1" and hit Enter.

    Click on B11 again, or hit the Up Arrow to get back to B11.

    Change the format to **dollars** ($) by clicking on the button labeled with a **$** sign.

    Notice that the **Number Format** in the pulldown menu above the $ becomes **Accounting**.

    The cell now displays "$13.10" (although it still *contains* plain old 13.1, as you can see in the formula bar).
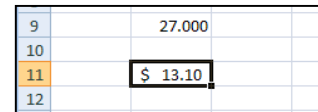
12. Click on cell B13, type in "25%" and hit Enter.

    Click on B13 again, or hit the Up Arrow to get back to B13.

    The format bar typically doesn't show formatting, but percents are an exception. 25% is displayed in *both* places— the cell *and* the formula bar. As you know, 25% is the same as the decimal 0.25, but it would be confusing to see 0.25 in one place and 25% in the other.
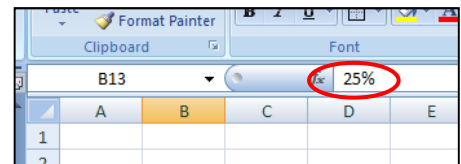
    Notice that the Number Format is **Percentage**.
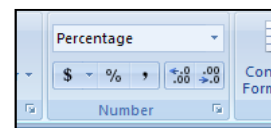
13. Click on cell B15, type in ".3" and hit Enter. **Select** B15 again (by clicking on it or by using Up Arrow).

    The decimal 0.3 is displayed in both the formula bar and the cell itself. Excel insists, quite properly, on showing a "leading zero" with decimals (0.3 rather than just .3)[4].

    Click the **%** button. The Number Format changes from General to **Percentage**.

    The cell display changes to 30%, as does the formula bar.

14. Click on cell B17, type in "3-1-12" and hit Enter. Select B17 again.
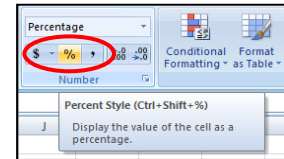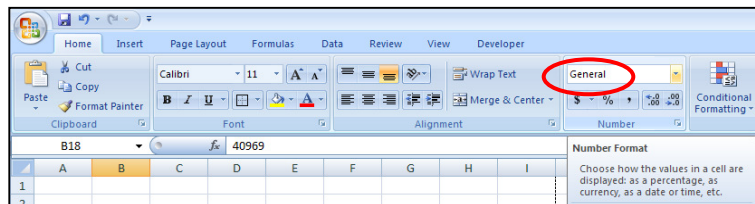
    Both the cell and the formula bar show "3/1/2012," the default format for a **date** (month/day/year). The Number Format has also become **Date**. In Excel, a date is actually a number—the number of days since January 1, 1900 (or since 1/1/1904 on Macs).

    Type "3/1/12" into cell B18, hit Enter, then change the Number Format from "Date" to "General" using the drop-down menu.

    Cell B18 now displays the number 40969. That's how many days there have been since 1/1/1900.[5]

    There are many possible date formats; you can even customize. Experiment with other Number Formats on your own by clicking on the drop-down menu and choosing More Number Formats at the bottom.

Merge cells

15. Type the text "Merged cell" into cell B20. Select both B20 and C20 by click-dragging across the two cells, as if you were highlighting a portion of text in Word.

    You've now selected a **range**—a rectangular block of cells. In this case, the selected range just has two cells in it, namely B20 and C20. Excel calls this range **B20:C20**. The colon is just the word "to," so you can read B20:C20 as "B20 to C20." A black box surrounds the selected range.

    Any formatting that you can do to one cell at a time, you can do to a whole range, too.

---

[4] Decimals without leading zeros can be misread far too easily. This error has actually killed people (imagine getting 5 milligrams of morphine in a hospital instead of .5 milligrams). Don't kill anybody—write 0.5 milligrams.

[5] Approximately. Microsoft Excel erroneously thinks that there was a leap day (February 29) in 1900. In fact, there was not. The normal rule for leap days (one every 4 years) does not apply to century years, which must be divisible by 400 to have a leap day.

You can also **merge** several cells into one cell. Do this judiciously—it can be hard to adjust a spreadsheet that has a lot of merged cells, because the grid is no longer uniform.
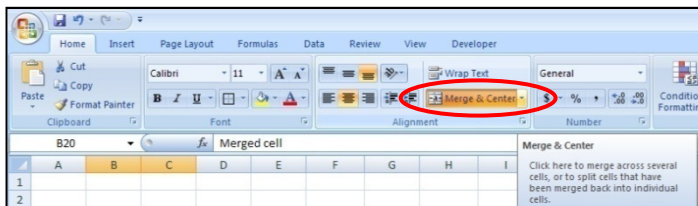
In the ribbon, click "Merge & Center."

The two cells join into one, and the text is centered. The name of the new cell is B20. The old cell C20 has vanished, becoming part of B20.

Center text across a range of cells without merging

16. Type "Centered across selection" into B22 and hit Enter. Select cells B22 to D22 by click-dragging.

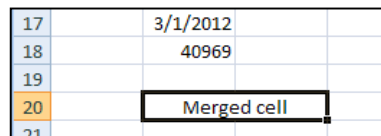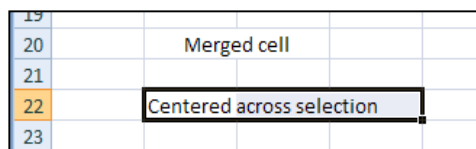This range is called B22:D22. You just indicate the first cell and the last cell when you're talking about a range.

Click on the **Format Cells: Alignment** button, which is located at the bottom right of the Alignment part of the ribbon. The button looks like this: .
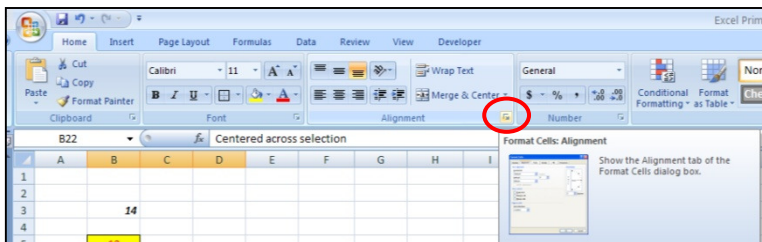
A "Format Cells" dialogue box opens up, with the Alignment tab activated.

17. Under the Horizontal drop-down menu, choose "Center Across Selection" and click "OK."

The text is now centered across B22: D22. It looks as if you've merged the three cells, but C22 and D22 still exist, as you can tell by clicking on them.
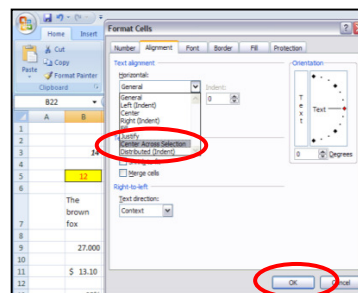
Centering Across Selection gives you much of the benefit of merged cells without making cells disappear (which can be dangerous).
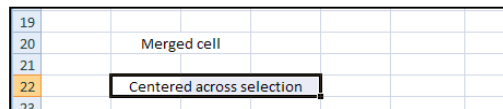
Ⓣ *Task:* **Cut, copy, and paste cells.**

Where to work: File – **Excel_Primer-Basics**
Sheet – **3. Cut Copy Paste**   ← Do your work on this sheet.

Example outcome: Sheet – **4. Outcome**   ← Check your work against this sheet.

⌐ *Process*

Copy and paste cells.

1. Open the file *Excel_Primer-Basics* if you don't already have it open. If the sheet labeled "3. Cut Copy Paste" isn't displayed, click on that tab at the bottom. You should see a blank spreadsheet.

2. Type "1" into cell B2, "2" into B3, and "3" into B4. Select the whole range from B2 to B4 by click-dragging.
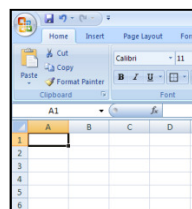
    Keyboard shortcut: click on B2, then hold down **Shift** and hit Down Arrow (↓) twice. By holding down Shift, you expand the selection.

    Second shortcut: click on B2, then hold down Shift and click on B4. The selection expands to include both B2 and B4 (so B3 is included as well).
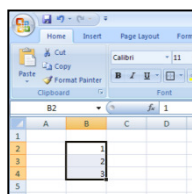
3. Tell Excel you want to **copy** these 3 cells. How? The same way you do in Word.
    Keyboard: Ctrl-C (PC) or Command-C (Mac)
    Ribbon: Click on the Copy button in the Clipboard area.
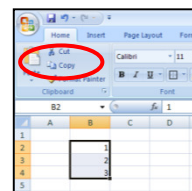
    A moving dotted border appears around the three cells, as if to say: "These cells are ready to be copied—just tell us *where*."

4. Click on cell D2 and hit Enter.

    You've now **pasted** the contents of B2:B4 into D2:D4, which are now selected. You didn't have to select the whole range D2:D4—all you had to do was indicate where the first cell would go, and smart Excel figured out the rest.
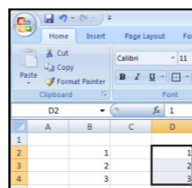
5. With D2:D4 still selected, copy the cells (by hitting Ctrl-C or however you prefer). The moving dotted border reappears, this time around D2:D4.

   Click on cell F2 and hit Ctrl-V (PC) or Command-V (Mac), or click the Paste button on the ribbon.
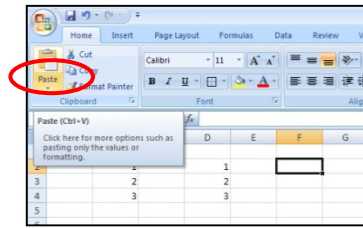
   You've now pasted again, but this time, the source cells stay activated. The moving dotted border is still around D2:D4. Using Ctrl-V or the Paste button is handy if you want to copy the same cells multiple times. If you don't, use Enter.
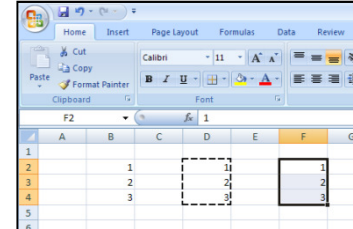
   To cancel the activation on D2:D4 (so you don't accidentally copy the cells), hit Escape (Esc) on your keyboard.

Copy and paste part of a cell.

6. Sometimes you don't want to copy and paste a whole cell. Here's how to copy and paste just *part* of a cell. Click on B6 and type "Copy and paste" into the cell. Hit Enter.
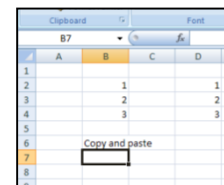
7. Select the word "paste" *inside* B6. How? You have at least 3 options:
   Double-click: Double-click on B6 to edit inside B6. Then click-drag over the word "paste" inside the cell to highlight it.
   Function key F2 (on PCs): Click on B6, then hit F2 to edit inside B6. Click-drag over the word "paste" inside the cell.
   Formula bar: Click on B6, then click-drag over the word "paste" in the formula bar.

   Hit Ctrl-C (or Cmd-V on Macs) to copy. There will be no moving dotted border, because you're not copying a whole cell or range of cells.

8. Click on cell D6 and hit Ctrl-V.

   The word "paste" appears in the cell.
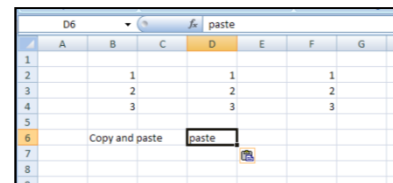
   A small floating button appears at the bottom right of D6, allowing you to change certain Paste options if you want. Ignore this for now.

9. Click on D7 and type "within" into it. Hit Enter.

   Activate editing inside D7 in any of these 3 ways:
   Double-click on D7.
   Click on D7 and hit F2.
   Click on D7 and click on the formula bar.

   Put the cursor at the beginning of the word "within," then hit Ctrl-V. You have now pasted the word "paste" in front of "within." Hit the spacebar to put a space between the words.
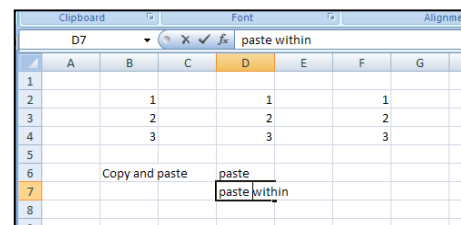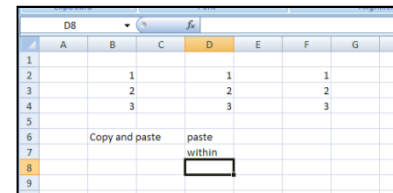
Drag & drop.

10. Select the range F2:F4 by click-dragging over those 3 cells.

    Hover over any edge of the selection (any edge of the thick black border). The cursor, normally a thick white cross (⊹), changes to a thin black cross with arrows (✛) under a white cursor arrow ( ).
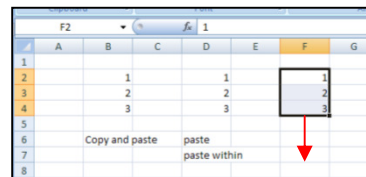
    Click, and then drag the range of cells to F6:F8. A dotted border will show you where you're dragging. When you release the button, the cell contents will vanish from the old place (F2:F4) and appear in the new place (F6:F8). You have cut and pasted using **Drag & Drop.**
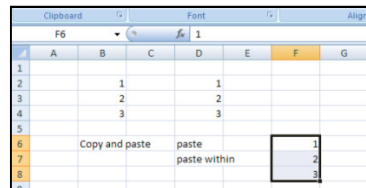
    Alternatively, you can use the **Cut** command (Ctrl-X on PCs or Command-X on Macs) and then Paste, but for many jobs, Drag & Drop is faster and easier.

11. To *copy* and paste using Drag & Drop, hold down the Ctrl key (on PCs) as you drag. Try this:

    Leave F6:F8 highlighted. Hover over any edge of the selection, then hold down the Ctrl key.

    The cursor changes from a thin black cross with arrows (✛) to a typical white cursor arrow with a plus sign ( ). The plus sign indicates that you are copying, not cutting.
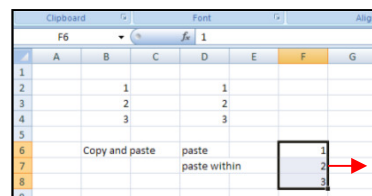
    Drag the selection to H6:H8. The "1-2-3" contents appear there, but they don't vanish from F6:F8.

    Drag & Drop works essentially the same way in Word.
        Plain Drag & Drop = Cut & Paste.
        Ctrl Drag & Drop = Copy & Paste.
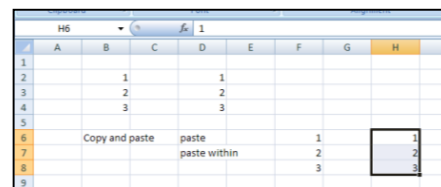
Auto-Fill a range.

12. Select the range F6:F8 again.

    Hover over the little black square (▪) at the bottom right of the black border surrounding the range. The cursor changes to a black cross (✚).
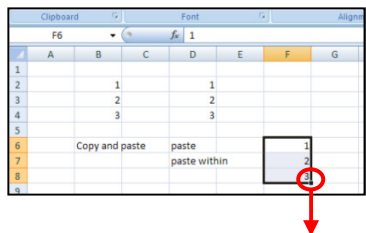
13. Drag the black square downward, extending the series of numbers up to 10 (in cell F15).

    Excel is smart enough to figure out this pattern of increasing numbers. If that *isn't* the pattern you want, then click on the small pop-up button below the black square. You can choose various Auto-Fill options, such as simply repeating 1-2-3 on down.
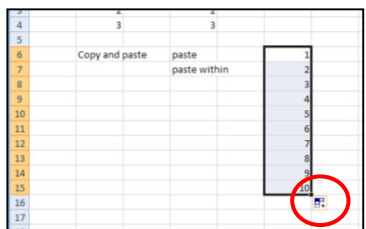
Auto-Erase within a range.

14. With the range F6:F15 still highlighted, drag the black square back *upwards*. You'll see cells being erased.

    Stop when you get to 6 (in F11).

    When you drag the black square back into the selected range of cells, you start erasing.
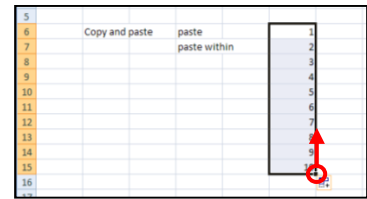
Use Format Painter to copy formats.

15. Put the word "fancy" into D13. Make it bold italics.

    Put the word "dull" into D15.

16. Click on D13 again (where you put "fancy"), then click **Format Painter** in the Clipboard segment of the Home ribbon.

    Cell D13 becomes surrounded by a moving dotted border—just as if you were cutting or copying it. But all you're doing is copying the fancy *format*, not the contents.

    The normal cursor (⊕) gets a little paintbrush added on, to remind you that wherever you click next is about to get painted with D13's format.

17. Click on D15. The word "dull" becomes bold italics.

    Notice that Format Painter happens as soon as you click—there's no separate Paste.
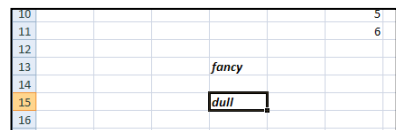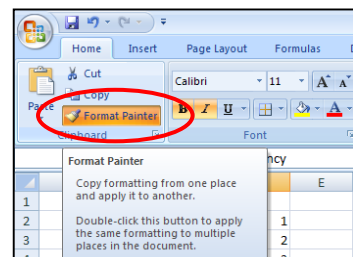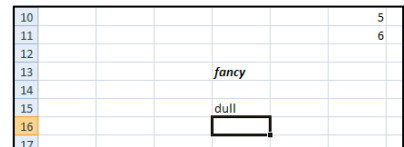
Paste without formatting.

18. Here's how to copy the content but *not* the formatting of a cell.

    Leaving D15 highlighted, choose Copy (in whichever way you prefer). Observe the moving dotted border around D15.

    Click on D17, selecting that cell as the destination.

19. Click on the drop-down arrow underneath the Paste button in the ribbon. Choose **Paste Values**.

    The word "dull" appears in D17 *without* bold italics. You've done the opposite of Format Painter. Under **Paste Special**, you can play with other Paste options if you like.

    Hit the Escape key (Esc) to remove the moving dotted border around D15, so you don't accidentally copy the cell again.
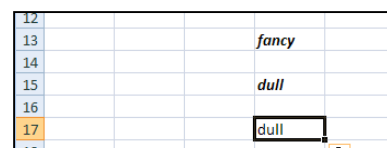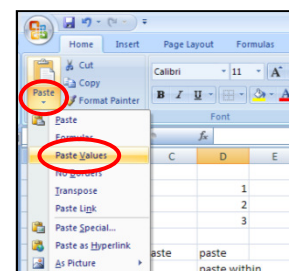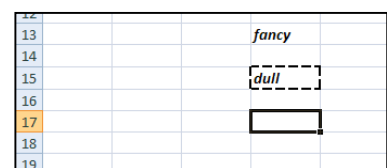
Add a comment.

20. Put the number "100" into D19. Right-click on D19 to call up a floating menu.

    These floating menus are very useful—you can change formatting, cut/copy/paste, and so on.

    For now, choose **Insert Comment**.

    A floating comment box appears. Your name (or whoever the software's registered to) is automatically included.

21. Type "Hello" into the box, then click anywhere else on the spreadsheet.

    The comment disappears, leaving only a red triangle in the upper right corner of D19 as a ghostly indicator of the comment's existence.

22. Hover over D19 to see the comment again.

    When you're group-editing a spreadsheet, comments are great, but even then use them sparingly.

    To edit or delete a comment, right-click the cell that the comment is attached to. Choose **Edit Comment** or **Delete Comment**. You can also choose to show comments permanently.
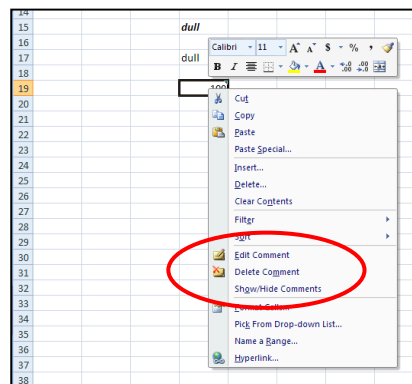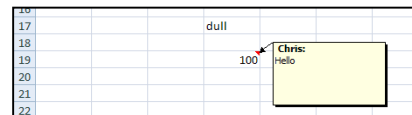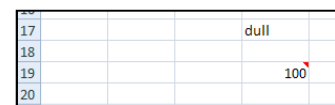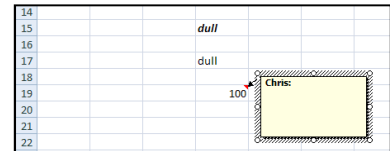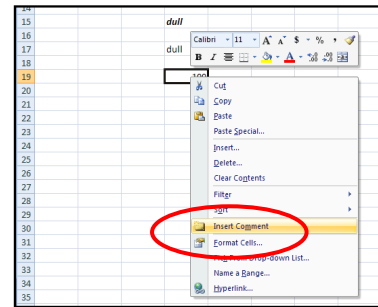
Ⓣ *Task:* **Alter the look of a whole spreadsheet.**

Where to work: File – **Excel_Primer-Basics**
                        Sheet – **5. Alter Worksheet**   ← Do your work on this sheet.

Example outcome: Sheet – **6. Outcome**   ← Check your work against this sheet.

⌐ *Process*

Adjust row heights and column widths.

1. Open the file *Excel_Primer-Basics* if you don't already have it open. If the sheet labeled "5. Alter Worksheet" isn't displayed, click on that tab at the bottom. You should see a blank spreadsheet.

2. Hover over the "3" that labels row 3. Notice that the cursor changes to a thick black arrow pointing right (➡).
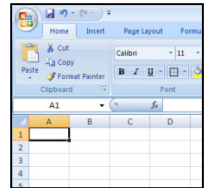
   Click on the "3." You've selected the whole row, and now you can change its formatting, you can copy & paste it, etc.
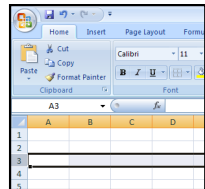
   However, to alter a row's height, you *don't* have to select it.

3. Hover *between* row label 5 and row label 6. The cursor becomes a black cross with vertical arrows (✛).
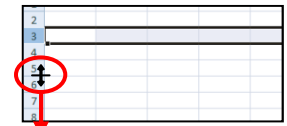
   Drag downward to make row 5 taller.

   Make the row 50 pixels tall. The measurement appears as you drag.
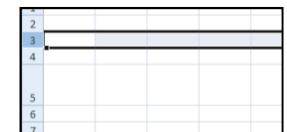
4. Using the same technique, make row 7 100 pixels tall.

   Now double-click in exactly the same spot, between row label 7 and row label 8.

   Row 7 returns to its natural height.

5. You change column widths the same way.
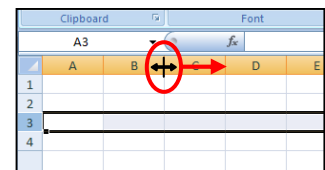
   Make column B 100 pixels wide by clicking between column headers B and C and dragging to the right.

   The cursor becomes a black cross with horizontal arrows (↔).
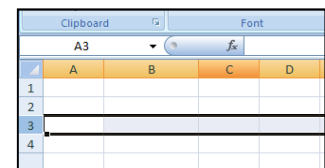
Hide and unhide columns and rows.

6. Select column E and call up a floating menu by right-clicking directly on column label E.

    In the menu, choose **Hide**.

    Column E disappears from view. Column D is now next to Column F.

    The hidden column (E) still exists, and in fact right now it's still selected, as you can tell by the black vertical line. That's kind of dangerous, because you could accidentally turn it bold, etc. A good practice is to click somewhere else.
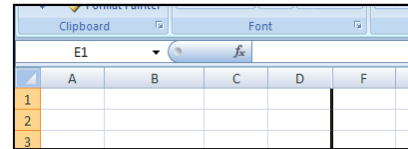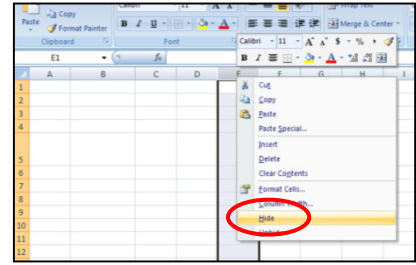
    The existence of the hidden column is indicated by the gap in the sequence (D to F), of course. But since you might miss that gap, on PCs there's a *slightly* thicker line placed between the column labels D and F. Mac Excel does a better job of alerting you to the gap: the two column headers are colored differently.
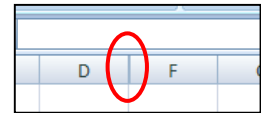
7. Using the same process, hide row 7: right-click and choose Hide.

    To **unhide** row 7, select both row labels 6 and 8 by clicking directly on 6 and dragging down to 8. Alternatively, you could click on 8 and drag up to 6.

    Rows 6 through 8 are now selected, including hidden row 7.

8. Right-click on either row label (6 or 8) and select **Unhide**.

    Row 7 now reappears.
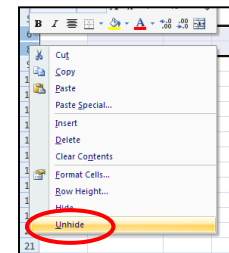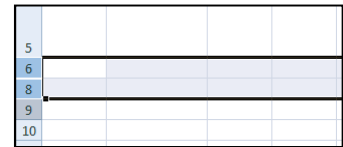
Hide gridlines.

9. At the top of the screen, click on View to switch to the View ribbon.

    In the Show/Hide part of the ribbon, uncheck the Gridlines option.

    The light gridlines on the spreadsheet vanish. The cell structure is still there, of course. Dropping gridlines has a positive psychological effect on some folks.

    Other options on this ribbon give you the power to control the view in various ways.
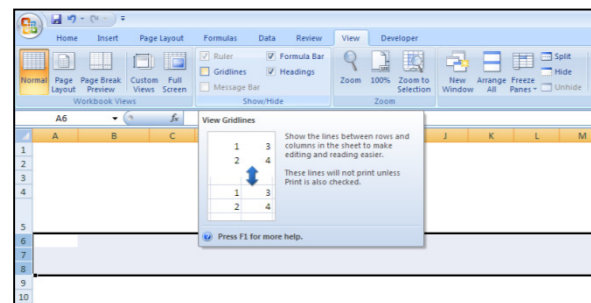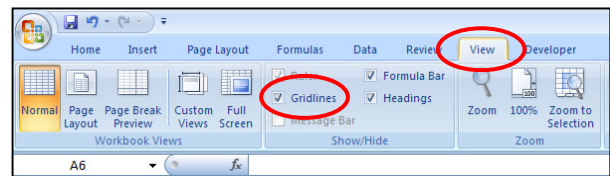
Freeze panes.

10. One of the most important options on the View ribbon is Freeze Panes. To appreciate its power, you'll need to put some data into the sheet.

    Put "test1" into H5, then drag the black square right until you have filled in "test5" in cell L5.

    Excel assumes that you want the number at the end to increase sequentially.

11. Type "1" into H6 and "2" into H7. Select both cells, then drag the black square down to H25, filling the range with numbers up to 20.

    Next, drag the black square right to copy the numbers into the other four columns, through "test5."

    Notice that numbers default to right-justification, while text defaults to left-justification. Let's center all the data to make the text and numbers line up more clearly.

12. Select columns H through L by clicking on column header H and dragging to column L.

    Go to the Home ribbon. Click on the Center button in the Alignment segment of the ribbon. The sample data is now clearly lined up.

    Now imagine that the columns of numbers extend to the right and down far off the screen—you have dozens of columns and hundreds of rows. If you scroll around using the **scroll bars** on the right and bottom of the window, you'll lose your place. Freeze Panes helps you keep track of your place in a giant spreadsheet by "freezing" the left and/or upper parts of the spreadsheet.

    Which parts? You have to tell Excel by splitting the window first.

13. To split the window horizontally, look for the tiny white rectangle toward the upper right of the whole window, just above the vertical scroll bar and below the formula bar.

    When you hover over the rectangle, the cursor looks like ⬍ .

    Click-drag down to just below row 5. The window splits into an upper part and a lower part.
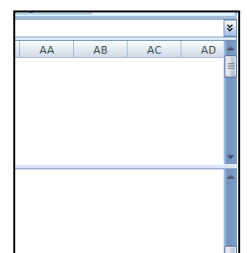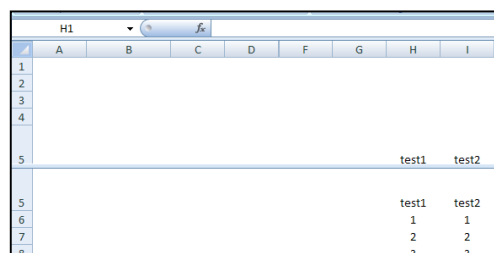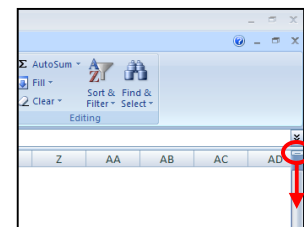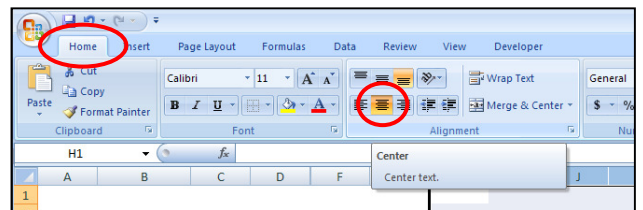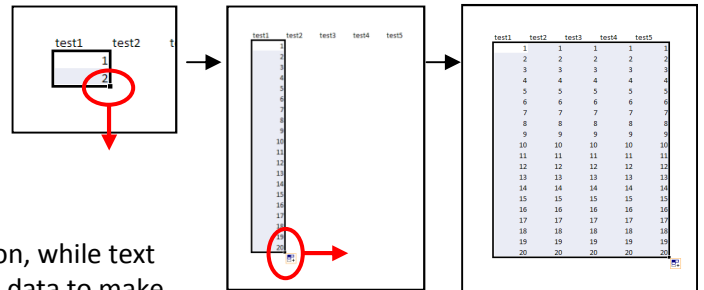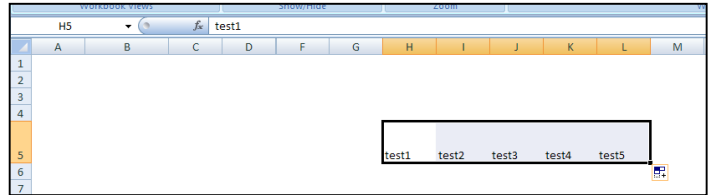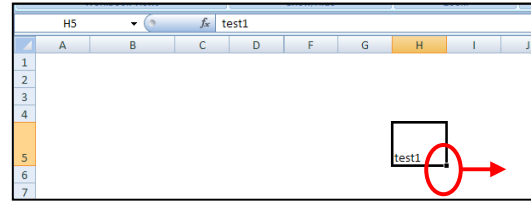
14. Now split the window vertically. To do so, grab one of the rectangles on either side of the bottom scroll bar. The cursor should look like ◀||▶.
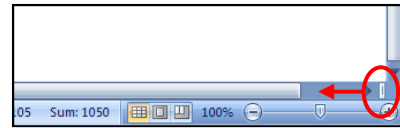
    Drag the vertical split to the spot between column H and column I.

    The window is now split into four panes.
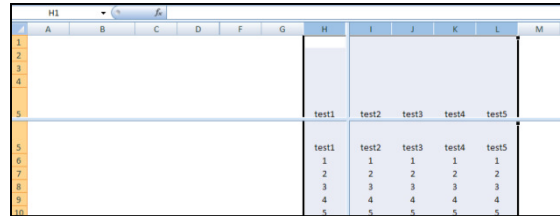
15. Click on the View ribbon.

    Click on the Freeze Panes button in the Window part, then choose Freeze Panes in the drop-down menu.
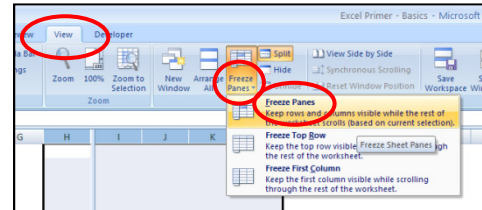
    Three of the four panes of the window are now "frozen." The boundaries between the four panes are now black lines.

    If you scroll down or to the right, using the scroll bars or the mouse wheel, you'll see that the left columns (A through H) and the top rows (1 through 5) remain in place. For instance, in the picture to the right, the spreadsheet has been scrolled both down and right, hiding some of the original data but revealing more rows and columns.
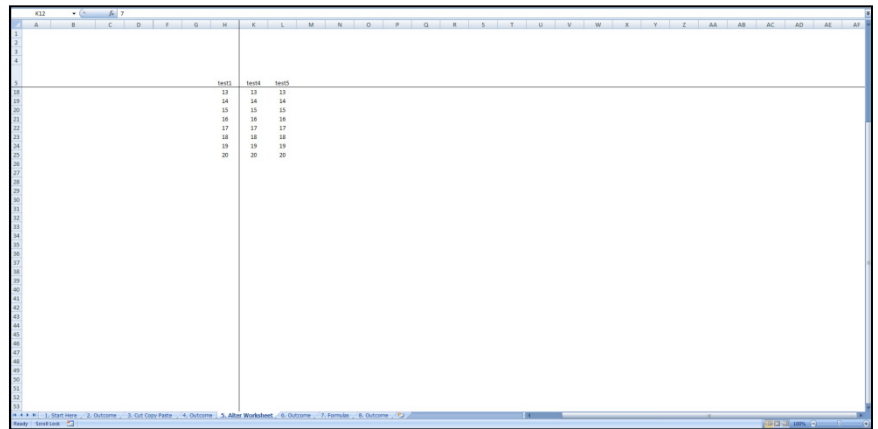
For a small spreadsheet, Freeze Panes does you little good, but for a giant sheet of numbers, Freeze Panes is a lifesaver.

Use conditional formatting.

16. This section closes with a glimpse at **conditional formatting**, which gives different formats to various cells according to their contents. Conditional formatting can help you spot patterns or outliers in a set of numbers.

    Select cell L24 (which contains the number 19) and replace the 19 with the number 11.

    Select the range K6:L25, then click on the Conditional Formatting button in the Home ribbon.

    Choose **Data Bars**, then the Red Data Bar option.

    Data Bars are essentially bar graphs built right into cells. With these bars, you can quickly see that that 11 is different from its neighbors.
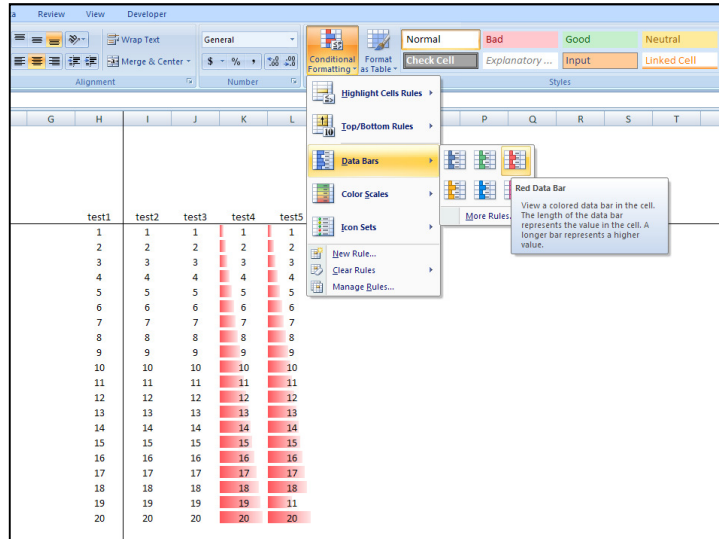
17. Select the range J6:25, then click on the Conditional Formatting button again.

    Choose **Color Scales**, then the Green-Yellow-Red Color Scale option. With this option, high numbers are green, middle numbers are yellow, and low numbers are red.

    You can customize Color Scales or any other kind of conditional formatting by selecting the range and choosing Manage Rules.

    Several other kinds of conditional formats exist. Explore the possibilities, but don't go crazy— after all, too many colors or other format variations can be psychedelic.
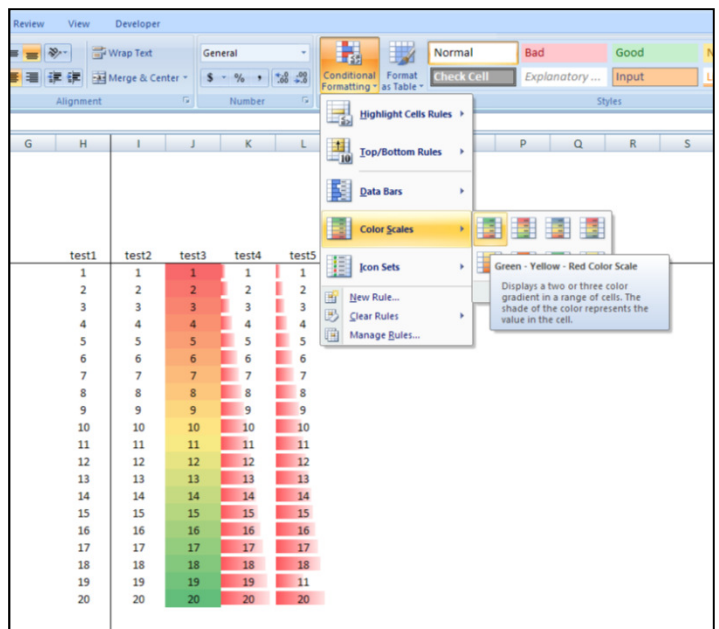
A-3. Formulas

It's time to unlock more of Excel's talents. So far, you've only put data—numbers and text—into cells. If that were all a cell could contain, then Excel would just be a table-maker.

However, a cell can contain a **formula**, which takes information from other cells, does a computation, and displays the result.

In this example, cell B3 contains the formula "=B1 + B2", as you can see in the formula bar.

This formula is a command: add the contents of B1 and B2 and display the result in B3.

B1 contains the number 7, while B2 contains the number 13, so B3 displays 20 (which is 7+13).

The power of this feature is that if you change the 7 or the 13, the sum automatically updates. The beauty of this feature is that your math—your thinking—becomes visible.

Again, you have to distinguish between what a cell *contains* and what it *displays*. You've seen the difference with number formats—a cell might contain 13.1 but display *$13.10*. With formulas, though, the difference is radical. B3 *contains* a formula (= B1+B2) but *displays* the result of that formula (20).

Think of inputs and outputs. The formula in B3 contains **references** to other cells, B1 and B2, which are the inputs to the formula. The output of the formula is displayed right in the cell—and that output can become the *input* to a formula in *another* cell, if that cell contained a formula referring to B3.

All right, let's start doing.

ⓣ *Task:* **Write working formulas.**

Where to work: File – **Excel_Primer-Basics**
                     Sheet – **7. Formulas**   ← Do your work on this sheet.

Example outcome: Sheet – **8. Outcome**   ← Check your work against this sheet.

⌐ *Process*

Create arithmetic formulas.

1. Open the file *Excel_Primer-Basics* if you don't already have it open. If the sheet labeled "7. Alter Formulas" isn't displayed, click on that tab at the bottom. You should see a blank spreadsheet.
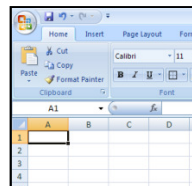
2. Put "7" in B1 and "13" in B2. In B3, type "=b1+b2" and wait to hit Enter.

    The equals sign (=) at the very start is what tells Excel that you're entering a formula. After that, as you type in any cell reference (b1 or b2—capital letters don't matter), the cell you're referring to gets highlighted in some color. This helps you avoid errors. (By the way, you probably won't see the "BINOMDIST" that appeared in the picture above, on the left side of the formula bar—we'll come back to what that is later.)
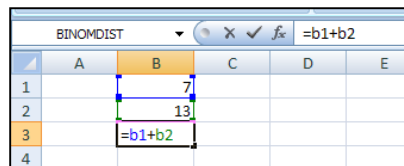
3. Hit Enter. B3 displays 20, which of course equals 7 + 13.

    Click again on cell B3 to see that it doesn't contain a **hard-coded** 20 (a 20 you literally typed in) but a "soft" 20, one produced by a formula. How can you tell? Look at the formula bar. You see "=B1+B2", not "20."
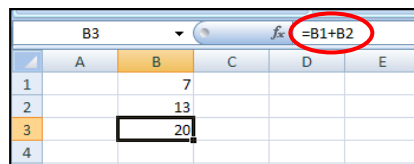
    If you haven't selected a cell, you can't tell whether what it's displaying is hard-coded or not. One way to decipher someone else's spreadsheet is to select different cells and study the formula bar.

    Get "inside" B3 again by double-clicking the cell (or hitting F2, or clicking in the formula bar). Again, Excel color-codes all the input cells for you, so you can trace cause and effect in the sheet. It's also capitalized the cell references for the sake of a consistent look.
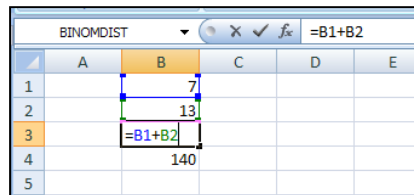
4. Click on B4. Hit "=" to start entering a formula. Now, instead of *typing* cell references such as "b1", you can click on cell B1 itself.

    Click on B1. The formula you're building is now "=B1". A dotted moving border shows up around B1. A moving border like that means "unfinished business," whether you're copying a cell or you're building a formula.
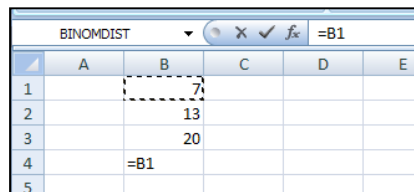
Type an asterisk (*, which is Shift-8 on the typical American QWERTY keyboard; it also has its own key on number pads). The asterisk means multiplication, because most keyboards lack the true multiplication sign (×), which is not the same thing as the letter "x" by any means.

5. Finally, click on B3 to enter "B3" into your growing formula, and hit Enter. B4 displays 140, which equals 7 × 20 (or 7 * 20, as you get used to Excel-arithmetic).

    Click again on B4 to see the formula you built (=B1*B3) in the formula bar.

    If you double-click to get inside the cell (or hit F2 or click in the formula bar), you'll see the color-coded input cells.

    The color-coding is more than cosmetic. You can change the formula by dragging those colored borders around.

6. Hover over the edge of the green border around B3. Click and drag it up to B2.

    The formula in B4 changes to "=B1*B2". Hit Enter. Cell B4 now displays 91, which is 7 times 13.[6]

    As you can see, typing "=" into a blank cell is not like typing a letter or a number, or in fact most symbols. The equals sign means "this is a formula." Without it, Excel thinks you're typing data.

7. Enter "B1*B2" (*without* =) into B5. The chunk of text is dead.

    Enter "Not a formula without =" into C5, so you remember how to start a formula. There *are* two other ways to begin formulas: the plus sign (+) and the minus sign, or hyphen (-). These symbols are not as general as the equals sign, though, so make it a habit to start with =.

8. Enter the following formulas however you like (typing the full formula, or typing part of it and clicking to enter cell references).

    1. Subtraction: In B7, put in "=B1-B2" and hit Enter. You get -6 (= 7 – 13). You can use either the hyphen above the "P" key or the dedicated "–" key on a number pad.
    2. Division: In B8, put in "=B2/B3" and hit Enter. You get 0.65 (= 13 ÷ 20, or 13/20 in decimal form). The forward slash (/) is below the question mark; it also has its own key on a number pad.
    3. Squaring: In B9, put in "=B1^2" and hit Enter. You get 49 (= 7^2, or 7^2) The caret symbol (^, or Shift-6) indicates that you're taking an exponent.
    General exponents: In B10, put in "=B1^B2" and hit Enter. You get a big old number: 96889010407, which equals 7^{13} or 7^13. Ah, the raw power of Excel!

---

[6] 91 is one of those numbers that *looks* prime, but as you can see, it's not. The multiples of 7 (above 70) and of 13 are not as solidly known as they should be.

9. Parentheses work as you'd expect to control the order of operations.[7] Put "=B1*(B2+B3)" into B11 and hit Enter. You get 231, which equals 7×(13+20), or 7×33.

   By the way, you always have to make multiplication explicit with an asterisk (*). You can't just bump parentheses up against each other, as you can in algebra when you write (x + y)(x – y).

Calculate a CAGR (Compound Annualized Growth Rate).

10. Here's a tough but useful formula that uses only the arithmetic covered so far. First, the set-up: type "Start Value" into B13, "End Value" into B14, and "Years" into B15.

    Type "$100" into C13 and "$150" into C14. Notice how you've given the Currency format to both cells by putting $ in front of the number. Type "3" into C15.

    Here's what we're trying to figure out. Say you put $100 into some investment, which pays you back $150 exactly 3 years later. What is the equivalent *annual* growth rate? In other words, what interest rate, compounded each year, would yield a 50% growth in the value of your investment after 3 years? (It's not 50% ÷ 3, by the way.) Some folks call this the CAGR (pronounced "CAGG-er"), the Compound Annualized Growth Rate.

    Let's do the algebra first. If *r* is the CAGR, then $100*(1 + r) would be the value of $100 after 1 year of growth, using the * to mean multiplication.[8] To grow at that rate for 3 years, you'd multiply by (1 + r) two more times. So the final value would be $100*(1 + r)^3. Here's the equation to solve for *r*, then:

    $$\$100*(1 + r)^3 = \$150$$

    $(1 + r)^3 = \$150/\$100$ — Divide both sides by $100.

    $(1 + r) = (\$150/\$100)\verb|^|(1/3)$ — Take the cube root of both sides to get rid of the cube.

    (Taking the cube root of something is the same as raising it to the 1/3 power.)

    $r = (\$150/\$100)\verb|^|(1/3) - 1$

So now we can figure out an algebraic formula involving the three inputs: Start, End, and Years. Here's the general formula for the CAGR (*r*), given the Start value, the End value, and the number of Years:

    $r = (End/Start)\verb|^|(1/Years)-1$

---

[7] If it's been a while since your last math class or quant-based standardized test, remember PEMDAS: Parentheses; Exponents; Multiplication or Division; Addition or Subtraction. The order is also known as "Please Excuse My Dear Aunt Sally," leaving unspoken the act she committed to require the forgiveness of millions of schoolchildren.

[8] If you're not sure why, imagine that r = 15%, or 0.15. Then $100*(1 + r) = $100*(1.15) = $115, which is 15% more than $100. When you multiply $100 by 1 + r, you get $100 + $100*r. The $100 by itself is the principal. The $100*r is the growth or interest, which in this example is 15% of $100, or $15.

11. Type "CAGR" into cell B17. In C17, put this formula: "=(C14/C13)^(1/C15)-1" and hit Enter. Change the Number Format of C17 to Percentage and increase the number of decimal places to 2.

    You should see 14.47% in C17. This means that your initial $100 would have to grow at an annual rate of 14.47%, approximately, to get you to $150 in 3 years.

    Change the End Value in C13 to $200. The CAGR goes up to 25.99%—to double your money in 3 years, you need a higher CAGR. Change the number of years in C15 to 10. The CAGR falls to 7.18%—you have much more time to double your money, so it doesn't have to grow as fast.

Rename cells.

12. To make difficult formulas more readable, you can rename cells.

    Click on C13 (where $100 is). Now click on the *left* part of the formula bar, which contains the reference "C13" itself.

    The reference "C13" becomes highlighted, as shown.

13. Type the word "Start" and hit Enter. The left side of the formula bar now reads "Start". That's the **name** that C13 will naturally have in future formulas.

    Using the same process, rename C14 "End" and C15 "Years". The renaming of C15 is shown to the right.

    Unfortunately, these new names don't update in any formulas you've already created. And you can *still* refer to those cells as C13, C14, and C15 if you wish. But you can type their new names into any new formulas, and if you *click* on those cells as you build a formula, you'll get the new names as well.

14. Click on C17 and hit Delete to clear its contents. Type "=(" and click on C14, where the $200 is. This action puts the name "End" into the formula, which reads "=(End" at this point.

    Type "/" and then click on C13, where the $100 is. This puts the name "Start" into the formula, which reads "=(End/Start" at this point.

15. Type ")" to close the parentheses, then "^(1/" and click on C15 (where 10 is). The name "Years" appears. Finish the formula with ")-1" and hit Enter. You get the same result as before (7.18%), but now the formula is closer to your algebra.

    C17:   =(End/Start)^(1/Years)-1

Use built-in functions.

16. Excel goes way beyond arithmetic with hundreds of built-in **functions** that you can use in formulas. Functions act on one or more inputs and produce an output.

    For example, the function SQRT takes the square root. Here's how to use it.

    Click on cell D9 and type "=sqrt". Excel confirms that it understands you (caps don't matter).
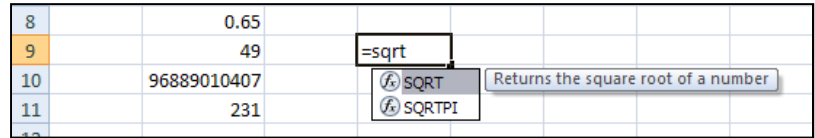
    Hit "(", the open parentheses. This is how you tell Excel what to take the square root of. Whatever is in the parentheses is the input.
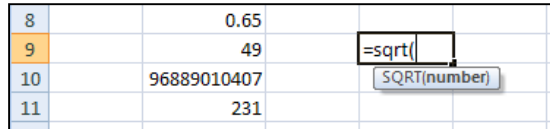
    This notation is the same as you learned in algebra. The function *f*(*x*) acts on the input *x* and gives you a result.

17. Either type "B9" or actually click on cell B9 to enter that reference into the function.

    Now close the parentheses with ")" and hit Enter. The cell displays 7, the square root of 49. The cell *contains* a formula that makes use of the function SQRT.
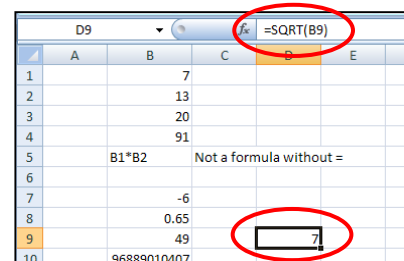
    Excel has all kinds of built-in functions. Some take just one input or **argument** inside the parentheses. Others take more than one argument; you separate arguments with commas.

18. Click on E1, type "=sum(" and click on B1. You're now going to use the built-in function SUM, but you need to tell Excel to add more than one number.
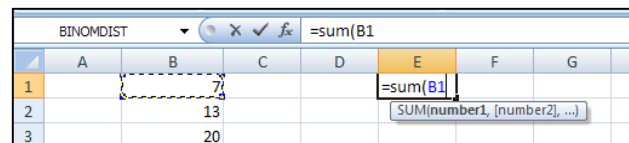
    Type a comma "," and then click on B2. You could keep going, adding more commas, but let's stop here.
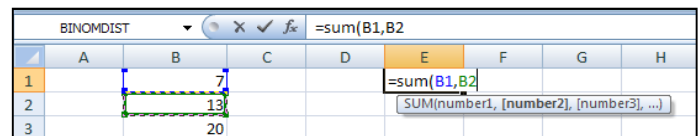
    Close the parentheses with ")". Hit Enter. You get 20, as you'd expect: the sum of 7 and 13.
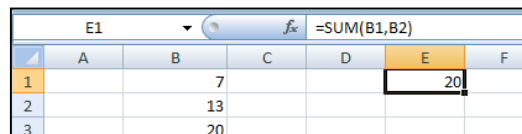
    Is that more cumbersome than typing "=B1+B2"? Sure. But what if you want to add up a whole range of cells? Using the plus sign would be a pain, while the SUM function can handle big ranges easily, just as a python can unhook its jaws to swallow a pot-bellied pig.

19. Type "2" into G1 and "4" into G2. Select the range G1:G2, then extend the series up to 20 by dragging the little black square down to G10.
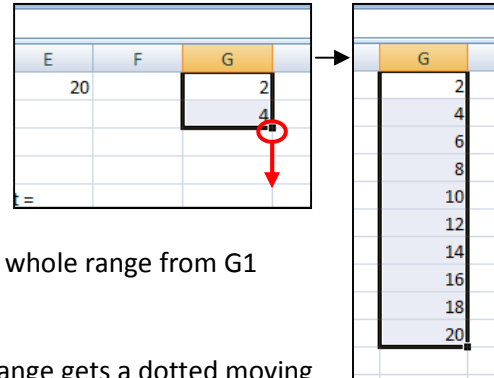
    Excel figures out that you want to go up by 2's.

20. Click on G12 and type "=sum(" into it. Now drag over the whole range from G1 through G10.

    The formula becomes "=sum(G1:G10", while the whole range gets a dotted moving border (the "unfinished business" border).

    Finally, close the parentheses with ")" and hit Enter. Cell G12 displays 110, the sum of the positive even numbers from 2 through 20.

    The formula in G12 is "=SUM(G1:G10)". There's only one argument in the function, but that argument is an entire range, G1:G10. The function SUM knows to add up every number in the cells in that range. If you wanted to add another range *as well*, you could put in a comma, then the reference to that other range inside the parentheses.
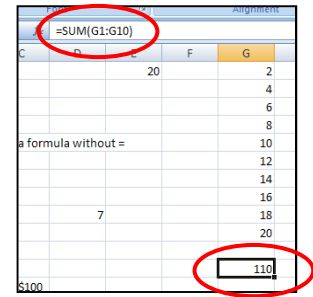
    A very useful function that takes more than one argument is IF. As the name implies, IF gives you one result *if* some condition is true, and a different result if that condition is not true.

    Here is the structure of the IF function, which takes three arguments.

    IF(*condition,   desired result if condition is true,   alternative result if condition is false*)

21. Let's build a simple IF function that will say whether a number is positive or negative.

    Type "Test" into F14. Put "4" into G14, put a solid border around the cell, and fill it with yellow. This cell will be our input. Next, type "Sign" into F16.
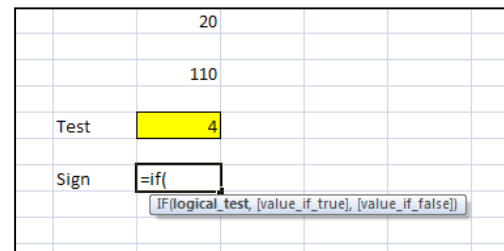
    Click on G16, which is where we'll put the IF function. Type "=IF(" and observe how Excel tries to help you, prompting you with what goes where.
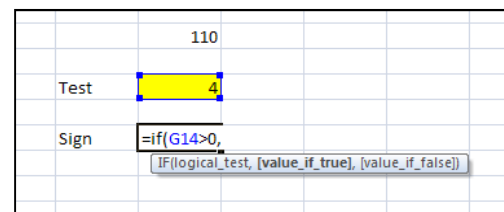
    The prompt says that the first argument is *logical_test*—that's the condition. Type in "G14>0" and then a comma to move on to the next condition, which is *value_if_true*. That's the desired result *if* G14 is in fact greater than 0.

    Type in the word "Positive" in quotation marks. By using quotes, you're indicating that you want Excel to take that as a bit of text, *not* as the name of some cell you renamed *Positive* for some reason. Type a comma to move on to the last argument.
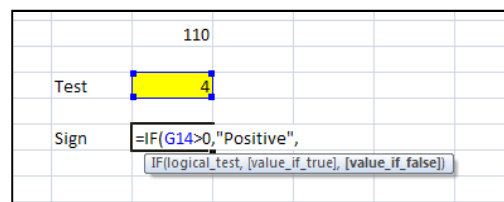
The final argument is *value_if_false*, which is the alternative result *if* the condition is not true. Type the word "Negative" in quotation marks. Close the parentheses with ")" and hit Enter. The formula should read thus:

=IF(G14>0,"Positive","Negative")

		110	
Test		4	
Sign	=IF(G14>0,"Positive","Negative")		

Test	4
Sign	Positive

The cell displays the word "Positive," as you'd expect. After all, 4 is a positive number.

22. Test different numbers in G14. Type "-5" into G14 and hit Enter. Cell G16 displays "Negative," just as you'd expect.

Test	-5
Sign	Negative

Now type "0" into G14. Uh oh. The formula produces "Negative," but 0 is not negative—it's neither positive nor negative. To deal with this third case, we're going to **nest** another IF function inside the first one.

You can use a flowchart to think through the cases.

```
                               Yes        ┌──────────────┐
                              ───────────>│    Show      │
                                          │  "Positive"  │
┌──────────────┐                          └──────────────┘
│  Is G14>0?   │                                          Yes   ┌──────────────┐
└──────────────┘            No  ┌──────────────┐        ──────> │    Show      │
                              ──>│  Is G14<0?   │               │  "Negative"  │
                                 └──────────────┘               └──────────────┘
                                                    No    ┌──────────────┐
                                                   ──────>│    Show      │
                                                          │ "Equals Zero"│
                                                          └──────────────┘
```
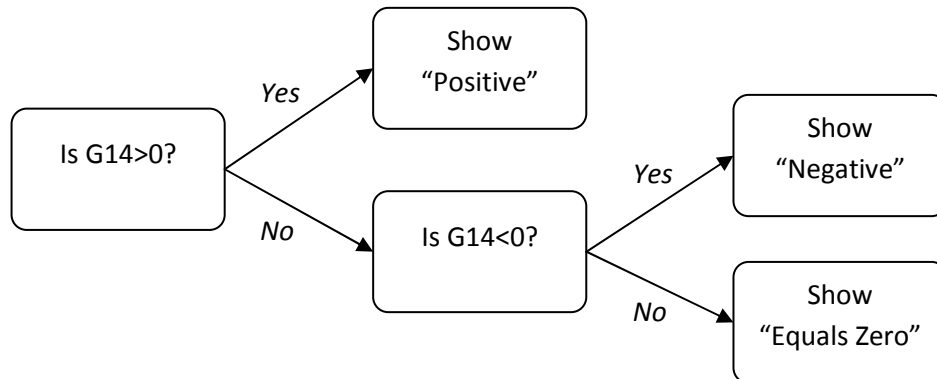
So you need to replace the current alternative outcome ("Negative") with another IF function—one that will test whether the number is negative. If it isn't, you know that the number is zero, because it's the only number that's neither positive nor negative.[9]

23. Click on G16 and get "inside" to edit the contents directly. Drag over the third argument "Negative", including the quotation marks, and hit Delete to erase.

Test	0
Sign	=IF(G14>0,"Positive",)
	IF(logical_test, [value_if_true], [value_if_false])

In the position of the third argument, type in the following:

IF(G14<0,"Negative","Equals Zero")

Be sure to keep the commas outside the quotation marks, because the commas are there to separate the arguments.[10] The full formula should look *exactly* like this:

=IF(G14>0,"Positive",IF(G14<0,"Negative","Equals Zero"))

---

[9] At least, it's the only "real" number that's neither positive nor negative. By default, Excel stays in the realm of the real. Not bad advice for the rest of us.

[10] In American English, one is always supposed to put commas and periods inside quotation marks, logic be darned. At times this document violates that prescription to try to prevent mistakes as you type things into Excel.

Notice the *two* closing parentheses at the end, because you had two open parentheses. An open parenthesis is like a ball thrown in the air—you have to catch it with a closed parenthesis.

	G16		▼		*fx*	=IF(G14>0,"Positive",IF(G14<0,"Negative","Equals Zero"))

	A	B	C	D	E	F	G	H	I
1		7				20		2	

13		Start Value	$100				
14		End Value	$200			Test	0
15		Years	10				
16						Sign	Equals Zero
17		CAGR	7.18%				

How far should you go with nesting functions? Not too far, because they become mind-scrambling. The formula above nests one IF inside another; that's acceptable. But too many layers down and it's like the deep levels of the movie Inception: your brain becomes boggled.[11]

More generally, Excel excels at keeping your logic explicit. If you write formulas that are too complex, you won't understand them later, and neither will anyone else. Break the computation into steps, and have each cell just do part of the work.

IF is a **logical** function. Other logical functions include AND, OR, and NOT. Excel has ten categories of functions, as well as good context-sensitive help that can assist you in choosing the right function and using it correctly.

24. Click on G18. In the formula bar, click on the **Insert Function** button, which is labeled *fx*.
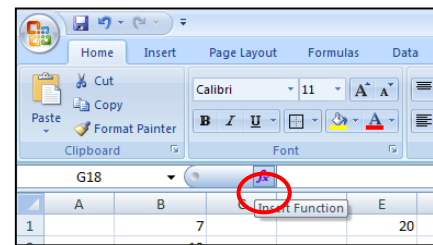
In the pop-up window, select the Logical category, then click on IF. Click OK.

Another pop-up window, Function Arguments, guides you through the entry of the three arguments of IF:
- Logical_test
- Value_if_true
- Value_if_false
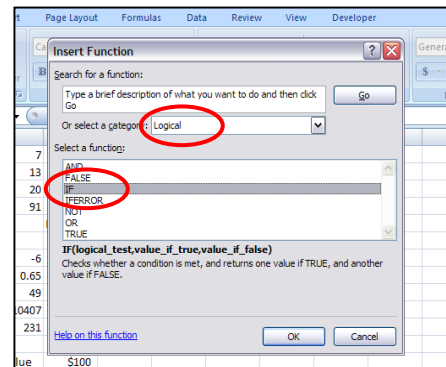
You can type directly into the boxes. The small blue buttons next to the entry boxes allow you to click-enter cell references from the worksheet.

As you build the function, the Function Arguments window gives you intermediate results, as well as quick access to further help on the function. Use this guidance whenever you're working with an unfamiliar function.

---

[11] But not in a good way.

Recognize relative addresses and create absolute addresses.

25. Click on cell D9 (which displays 7). Notice that the formula bar reads "=SQRT(B9)", indicating that cell D9 is taking the square root of the value displayed in B9 (which is 49).

    What if you *copy* cell D9 somewhere else?

26. Drag the little black square in the bottom right corner down two cells, filling in D10 and D11.

    D10 displays 311270, while D11 displays 15.19868. Evidently, neither cell is taking the square root of B9! Instead, D10 is taking the square root of B10 (that big number beginning with 9), while D11 is taking the square root of B11, which displays 231.

    Click on D10 directly to confirm. Indeed, D10 contains the formula "=SQRT(B10)". The reference to B9 inside D9's formula changed when you copied D9 elsewhere.

    This all happened because of **relative addressing** or **relative references**. When you put "=SQRT(B9)" into D9, you *thought* you were telling D9 to take the square root of B9. In fact, you were really telling it to take the square root of the cell *two columns away, to the left.* So, when you copied the formula down, the formula kept taking the square root of the cell two spots to the left.

Excel assumes you want relative addressing, because it makes big spreadsheets much easier to build. Most of the time, you *want* relative addressing.

The opposite of relative addressing is **absolute addressing**, which means "refer back to such-and-such a cell, no matter where I copy this formula."

Absolute addresses are useful when you're taking percents of a whole.

27. Put the formula "=G1/G12" into cell H1. Apply the Percentage number format to H1 and increase the decimal point by one place.

    The cell reads 1.8%, indicating that 2 is 1.8% of 110. What about the rest of the numbers in that column, which add up to 110?

28. Drag the little black square down to fill in through H10. (Alternatively, you can *double-click* on the black square, and it will fill down exactly as far as you want. How does Excel know? You're right next to a column of data that goes to G10.)

Ouch! You get all kinds of error readings, most of them saying that you're trying to divide by zero.

Why? Think about the formula you wrote, "=G1/G12". When you copy that down, you'll get "=G2/G13", "=G3/G14" and so on. You don't want the bottom of the fraction to shift. The top of the fraction changes appropriately, but you want to *lock* the bottom of the fraction. You want the denominator to be the number in G12, no matter what.

And of course, the *last* thing you'd want to do would be to type 110 (the actual number in G12) directly into the formula! Avoid hard-coding numbers into formulas as much as you can—particularly if those numbers are the results of computations in your spreadsheet.

Here's how to create an absolute address for G12—put **dollar signs** ($) directly in front of the column letter and row number, like so: $G$12. Think of those dollar signs as nails—you're nailing down the reference, so it doesn't change as you copy the formula.

First, clear the erroneous cells by hitting Delete.

29. Put the formula "=G1/$G$12" into cell H1. There are two options on how to do this:
    - Type the formula in exactly as you see it.
    - (PCs only) Type in "=", then click on G1 to enter the G1 reference, then type "/", then click on G12—and hit the F4 function key on the keyboard once. The G12 reference in the formula changes to $G$12. The F4 key cycles through the four possibilities of relative and absolute addressing:

G12	$G$12	G$12	$G12
all relative	all absolute	lock row only	lock column only

Hit Enter. Now copy the formula down by dragging the black square or by double-clicking (or even by copying and pasting).

The formulas copy and update as you want them. For example, H2 contains "=G2/$G$12", H3 contains "=G3/$G$12", and so on.

Long story short: use dollar signs to make absolute addresses in formulas—that is, to lock down the references to particular cells.

By the way, the names of renamed cells (such as Start and End, the new names you gave to C13 and C14) are automatically absolute. So you could just give a new name to G12, such as Total, then write your first formula as "=G2/Total". When you copy that cell elsewhere, you will always be dividing by Total—the value displayed in G12.

Create cumulative sums.

30. Imagine you want to compute the running or cumulative sum of the percents in column H. Here's a useful pattern to learn.

Put "=H1" in cell I1.

Now put "=I1+H2" in cell I2.

Notice that I2 is the sum of the cell above it (I2, the *previous* cumulative sum) and the cell next to it (H2, the percent in the *current* row). That gives you the current cumulative sum.

Even better, this formula is a pattern that extends perfectly. All the addresses are relative.

Copy cell I2 down to I10. The column fills with the cumulative sum, topping out at 100% as you'd expect.

There are other ways to create a cumulative sum formula that you can copy down. For instance, you can use the SUM formula as follows.

31. Type "=SUM(" into J1, then click on H1, then type ":", then type ")". The formula should look like "=SUM(H1:H1)" at this point. Notice that you have a one-cell range—not very useful yet.

Now you're going to "pin" one of the end of the range to H1 by changing one reference to $H$1, using absolute addressing. Meanwhile, you leave the other end of the range relative, so it will stretch the range as you copy the formula.

Click on the first H1 in the formula and change it to "$H$1", either by overtyping it or by hitting function key F4 once. Hit Enter.

Now copy the cell down. You get cumulative sums, because the range you're summing expands appropriately. For instance, cell J6 contains "=SUM($H$1:H6)", summing the cells boxed in red to the right.

Which technique should you use? Either works. The first uses a principle called *recursion*—you figure out each cumulative sum using the previous cumulative sum. Notice, though, that your first formula had to be different. The second method keeps all the formulas the same, but it uses a mix of absolute and relative addressing, which is a bit more cryptic.

All right, you're almost done with this section of the primer. One last miscellaneous task: adding worksheets to a workbook, renaming them, moving them around, and deleting them.

Add, rename, move, and delete a worksheet.

32. Click on the tiny tab just to the right of the "8. Outcome" tab, at the bottom of the window.
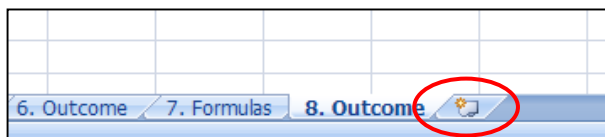
A new worksheet will appear, called "Sheet14" or "Sheet15" or even higher (depending on how many times someone has created sheets in this workbook). That's just to guarantee a new name for the sheet. Every sheet in a workbook must have a different name.
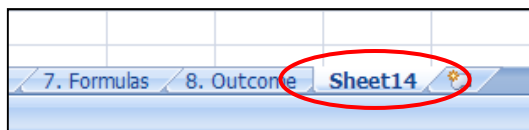
33. Double-click directly on the name of the sheet to rename it almost anything you want. There are certain characters you can't use, but Excel will prevent you from making mistakes.
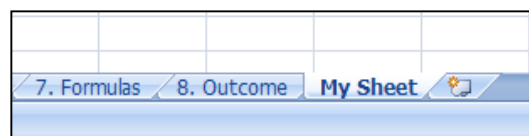
34. Move the sheet by dragging the tab to the left and dropping it between "7. Formulas" and "8. Outcome".
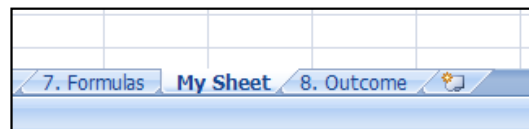
By the way, if you had held down Control as you dragged & dropped, you would have made a copy of the sheet.

35. Delete the new sheet entirely by right-clicking on the tab and choosing Delete from the floating menu.

If you had anything on that sheet, Excel would confirm with you first before deleting.
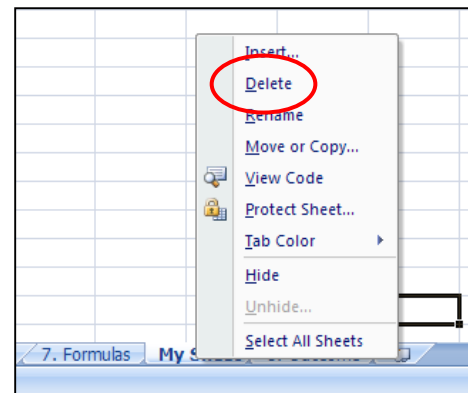
B. Data Analysis

In this section, as the name says, you'll use Excel to analyze data.

However, the initial task has little to do with Excel. If you're going to do anything intelligent with the software, then you'd better start here:

ⓣ *Understand* **the data first.**

Data has meaning. Make sure you get the broad storyline before you dive into number-crunching.

Here are some real numbers from a real company—in fact, one called Realty Income. This company is a "real estate investment trust," or REIT (pronounced "reet"). REITs typically own properties that produce income. For instance, Realty Income owns over 2,600 commercial properties in 49 US states. These properties are rented out to various long-term tenants, such as Taco Bell, Jiffy Lube, Rite Aid, and Wawa[12]. As a REIT, Realty Income must pay through at least 90% of its income as dividends to shareholders. To grow, Realty must not only increase the rents on properties it currently owns but also acquire new properties.

Let's say you're doing a little project for Realty: helping them identify opportunities to expand. Specifically, you're asked to look at some of their geographic data. Which states are doing well and which aren't?

Realty provides you with the following state-level data, which is published on their website (as of spring 2012, the link is http://www.realtyincome.com/pdfs/geographic_diversification.pdf). The file *geographic_diversification.pdf* is also included with this primer.

Geographic Diversification

The following table sets forth certain state-by-state information regarding Realty Income's property portfolio as of March 31, 2012 (dollars in thousands).

---

[12] The best convenience store on the planet. Wawa milk is a delight.

State	Number of Properties	Percent Leased	Approximate Leasable Square Feet	Rental Revenue for the Quarter Ended March 31, 2012 [1]	Percentage of Rental Revenue
Alabama	62	95%	420,200	$ 1,767	1.5%
Alaska	2	100	128,500	307	0.3
Arizona	87	98	619,500	3,069	2.7
Arkansas	17	94	92,400	285	0.2
California	123	99	2,670,100	15,337	13.4
Colorado	59	95	504,200	1,868	1.6
Connecticut	23	87	269,100	1,092	1.0
Delaware	17	100	33,300	433	0.4
Florida	184	96	1,881,000	7,754	6.8
Georgia	143	94	1,251,000	4,908	4.3
Hawaii	--	--	--	--	--
Idaho	12	92	80,700	320	0.3
Illinois	101	98	1,335,900	6,043	5.3
Indiana	81	96	799,000	3,610	3.2
Iowa	21	100	290,600	1,019	0.9
Kansas	37	95	642,900	1,350	1.2
Kentucky	23	96	134,700	653	0.6
Louisiana	34	100	344,200	1,260	1.1
Maine	3	100	22,500	139	0.1
Maryland	29	100	384,000	2,194	1.9
Massachusetts	64	91	575,400	2,286	2.0
Michigan	54	96	287,200	1,197	1.0
Minnesota	150	100	1,003,600	6,765	5.9
Mississippi	72	99	360,700	1,549	1.4
Missouri	76	95	1,027,500	3,802	3.3
Montana	2	100	30,000	81	0.1
Nebraska	19	95	196,300	503	0.4
Nevada	15	100	325,800	1,042	0.9
New Hampshire	15	93	217,200	944	0.8
New Jersey	33	91	260,400	1,909	1.7
New Mexico	9	100	58,400	200	0.2
New York	42	93	776,200	4,213	3.7
North Carolina	93	100	570,100	2,863	2.5
North Dakota	6	100	36,600	63	0.1
Ohio	134	96	1,122,100	3,941	3.4
Oklahoma	35	94	752,400	1,426	1.2
Oregon	20	100	384,200	1,236	1.1
Pennsylvania	103	98	907,200	4,137	3.6
Rhode Island	3	100	11,000	37	*
South Carolina	98	98	371,400	2,294	2.0
South Dakota	10	100	89,800	186	0.2
Tennessee	127	96	737,500	2,901	2.5
Texas	213	98	3,122,600	10,488	9.2
Utah	6	100	121,700	285	0.2
Vermont	4	100	12,700	130	0.1
Virginia	105	95	1,519,400	4,582	4.0
Washington	35	97	298,100	1,035	0.9
West Virginia	2	100	23,000	125	0.1
Wisconsin	27	93	269,200	938	0.8
Wyoming	1	0	5,400	0	0.0
Totals/Average	2,631	97%	27,376,900	$ 114,566	100.0%

* Less than 0.1%

(1) Includes rental revenue for all properties owned by Realty Income at March 31, 2012, including revenue from properties reclassified as discontinued operations of $162. Excludes revenue of $64 from properties owned by Crest.

Before you break out Excel, **figure out what you have**. Look over the table, focusing on one column at a time.

State: Each row corresponds to a different US state, obviously, with a total or average row at the bottom.

Number of Properties: How many properties Realty owns in each state. As you glance down the list, notice the "-- " in the Hawaii row. Compare this observation with what you might remember from a couple of pages ago: "Realty Income owns over 2,600 commercial properties in 49 US states."

The "-- " in the Hawaii row should make sense: since Realty has properties in 49 states, and there are 50 US states, there must be one state with no Realty properties. Now you know that state is Hawaii.

This particular "click-together" of two jigsaw pieces is minor, but it highlights a crucial principle of data analysis: *strive to reconcile disparate pieces of data—or at least, understand their differences.* As much as you can, try to build *one* coherent picture of the world you're studying. Of course, two pieces of data from the same source (as in this case) will usually agree, but not always—and that discrepancy can be important. Different data sources will often be mismatched, but look for the reasons why.

Take note of the total: 2,631 properties.

Percent Leased: What proportion of the properties is leased (to tenants). A question you might ask is whether partial leases of a property are possible. In other words, if Realty owns 23 properties in Connecticut and the Percent Leased is 87%, does that mean that 87% of 23 properties, or 20 properties, are leased, while 3 are not? Or is it that 87% of the *space* in the 23 properties is leased, while 13% is not?

The only way to resolve this question is to learn more about how the business works. From Realty Income's website, you discover that Realty Income owns free-standing buildings that it almost always leases to single tenants at a time. So you can conclude that Percent Leased is a percent of properties, not of square footage.

When you encounter a percent in data, ask yourself what it's a percent *of*—and whether it makes sense for it to be a percent of something other than your assumption. Most of the time, your immediate intuition will be right, but you don't want to get caught assuming the wrong thing.

The average Percent Leased is 97%. Connect that with the previous column: 97% of the 2,631 properties are leased out to tenants.

Approximate Leasable Square Feet: The total area of the properties in that state. The unit here (square feet) tells you the quantity (area) that is being measured. Again, note the total: 27,376,900 square feet. As you glance over more and more data, it can be helpful to round off key numbers and articulate them to yourself within a full sentence: *"Realty has about twenty-seven million square feet of leasable space."* That's more memorable than the raw figure.

Rental Revenue for the Quarter Ended March 31, 2012: The rental revenue for the first quarter of 2012. Notice that the dollars are in thousands and that the revenue is *not* annual. It's easy to miss the latter point—you'd never think that Realty Income was only making $37 a year from renting 3 properties in Rhode Island, but you might think that they made $37,000 a year. Remember to annualize these

revenues if you need to. In this industry, you can just multiply by 4 to get a very good estimate of annual revenue (in highly seasonal businesses, annualization needs to be more carefully done).

Also read the footnote. Some footnotes in company data are unimportant; others are super-important, because that's where the company squirrels away bad news that it must legally release. As you check out any footnote, think about the size of the issue. Here, the inclusion ($162 thousand) and exclusion ($64 thousand) are far less than 1% of the total revenue across all states ($114,566 thousand, or almost $115 million). Perhaps the impact differs by state, but on first pass, you can ignore this footnote.

Percentage of Rental Revenue: Rental Revenue is the previous column, so in this column you're taking a percent *of* the total rental revenue. To confirm this thinking, recognize that the total percent in the final row is 100.0%.

Here's another way to avoid "drowning in data" as you scan this sort of table of numbers: *pick a particular case and articulate it.* "In Ohio, they made $3,941, almost $4,000 thousand or $4 million in the first quarter of 2012. That was 3.4% of the total revenue they made that quarter."

Now that you've given the numbers a serious look-over, let's get them into Excel.

ⓣ *Task:* **Import data into Excel and do basic clean-up.**

Where to work: File – **Excel_Primer-Data_Analysis**
                              Sheet – **1. Start Here**   ← Do your work on this sheet.

Example outcomes: Sheet – **2. Cleaned-up Data**   ← Check your work against this sheet.
                              Sheets **1a**, **1b**, and **1c** illustrate intermediate results.

⌐ *Process*

Cut and paste data from a PDF.

1. Open the file *Excel_Primer-Data_Analysis* and click on the empty sheet "1. Start Here".

    Open the PDF called *geographic_diversification.pdf*. The file will open in Adobe Acrobat Reader on PCs or in Preview on Macs.

    Click-drag over *just* the table to highlight it. Leave out the text at top and bottom for now—that's easy to cut and paste. Choose Copy (Control-C or Command-C) to copy the table.

    Switch to Excel, click on cell A5, and choose Paste.

    If you're lucky, you'll get what's in sheet "1a. Direct from PDF (a)", or something close to it. The data copied in *roughly* correctly. Each entry in the table copied into a separate cell in Excel. In fact, some entries copied into *more* than one cell—for instance, "New York" became "New" and "York". That's because Excel chose here to think of the *space* between words as a **delimiter**—the separator between entries. So there's some clean-up to do.

13 Rental	Revenue	for				
14 the	Quarter	Ended				
15 March	31,	2012(1)				
16 Percentag of						
17 Rental						
18 Revenue						
19 Alabama	62	95%	420,200	$	1,767	1.50%
20 Alaska	2	100	128,500	307	0.3	
21 Arizona	87	98	619,500	3,069	2.7	
22 Arkansas	17	94	92,400	285	0.2	
23 California	123	99	2,670,100	15,337	13.4	
24 Colorado	59	95	504,200	1,868	1.6	
25 Connectic	23	87	269,100	1,092	1	
26 Delaware	17	100	33,300	433	0.4	
27 Florida	184	96	1,881,000	7,754	6.8	
28 Georgia	143	94	1,251,000	4,908	4.3	
29 Hawaii	--	--	--	--	--	
30 Idaho	12	92	80,700	320	0.3	
31 Illinois	101	98	1,335,900	6,043	5.3	

If you're *less* lucky, you'll get what's in sheet "1b. Direct from PDF (b)". Here, each row became a single cell. During this copying procedure, Excel didn't interpret spaces in the main row as delimiters. Rather, the line break became the only delimiter for the table itself.

16	Percentage of
17	Rental
18	Revenue
19	Alabama 62 95% 420,200 $ 1,767 1.5%
20	Alaska 2 100 128,500 307 0.3
21	Arizona 87 98 619,500 3,069 2.7
22	Arkansas 17 94 92,400 285 0.2
23	California 123 99 2,670,100 15,337 13.4
24	Colorado 59 95 504,200 1,868 1.6
25	Connecticut 23 87 269,100 1,092 1.0
26	Delaware 17 100 33,300 433 0.4
27	Florida 184 96 1,881,000 7,754 6.8
28	Georgia 143 94 1,251,000 4,908 4.3
29	Hawaii -- -- -- -- --
30	Idaho 12 92 80,700 320 0.3

Which way will Excel do it for you? It's a bit unpredictable, depending on the exact boundaries of what you've highlighted, which hidden character you picked up or not, etc. So try copying a few different ways to find the one that works best:
- Select slightly different boundaries. Try copying only part of the table.
- Choose Paste Special and try pasting "as" different options (Text, Unicode Text, etc.)

Let's assume for now that you are able to do the 1a version. We'll come back to the 1b version shortly. If you aren't close to 1a in your copying, pick up from the 1a version at this point.

2. Scroll to assess the damage—how off are you? A quick scan reveals the following:
   - Screwy column headers (every word in its own cell)
   - Alabama's off by one column because the dollar sign ($) got its own cell.
   - Two-word states are off by one column.

Widen columns (by dragging between column headers) to show cut-off text and numbers. In particular, cut-off numbers show up as "######", as in cell D69. By widening column D, you see the number.

62	Utah		6	100	121,700	285	0.2	
63	Vermont		4	100	12,700	130	0.1	
64	Virginia		105	95	1,519,400	4,582	4	
65	Washingt		35	97	298,100	1,035	0.9	
66	West	Virginia	2	100	23,000	125	0.1	
67	Wisconsin		27	93	269,200	938	0.8	
68	Wyoming		1	0	3,100	0	0	
69	Totals/Av		2,631	97%	######## $		114,566	100.00%
70								

Clean up data.

Now you have to get the data into the right columns. You have at least two choices: *manual* clean-up or *formula* clean-up. Manual clean-up means cutting and pasting cell by cell or even within cells. For a small table, manual clean-up is totally reasonable. Formula clean-up means leaving the data mostly as is, but putting formulas in other cells that read in the right data. As long as there's a pattern in the chaos, there's a formula that can decode it. For a giant table, you'll probably need to do some kind of formula clean-up in addition to manual clean-up.[13]

3. First, let's get the column headers right. Since there are only a few columns and the Paste outcome for those column headers was so terrible, you're going to recopy them one header at a time into cells J18, K18, and so forth to the right.

So that Excel doesn't read the spaces as delimiters, copy the text of each header "inside" the target cell. For instance, "Number of Properties" should go into K18. Select the header "Number of Properties" in the PDF and hit Copy. Now *double*-click on cell K18 (or click into the

---

[13] Before you start manipulating, it can be a good idea to make a copy of the raw, ugly data by copying the whole worksheet to an "Archive Raw Data" sheet, just in case you make a mistake in the process somewhere.

formula bar, or hit function key F2) so that you're editing *inside* K18, and hit Paste. You'll have to manually remove the line break after the word "of," unfortunately.

Once you're done copying the column headers, adjust the column widths, center the text, and use Wrap Text to prettify what you've got.

4. Completely delete rows 5 through 17 now, since they're distracting. How? Select those rows by click-dragging over the row labels. Then right-click and select Delete.

The rows are completely deleted from the spreadsheet. Everything that's further down moves up, so that what *was* in row 18 is now in the new row 5.

While we're at it, Freeze Panes so that rows 1 through 5 are frozen. This way, as you scroll up and down, you don't lose your place. (See pages 19-20 to refresh yourself if necessary.)

5. The name of each state should go in column J. How can a formula figure out what to put in there correctly? Here's the pattern to notice:
   o If a state has one word in its name, column A has the state name, while column B has a *number* in it (except for Hawaii, which has some pesky dashes in column B).
   o If a state has two words in its name, column A and column B together contain the state's name. So column B has *text* in it (again, except for Hawaii).

This is enough to give Excel clear directions. Let's first get rid of the Hawaii exception, since those dashes are really the number zero.

Type "0" into B16 and copy that 0 across to F16.

6. Next, here are the functions and operations we're going to need.
   o IF
     ▪ Tests a condition and gives you one outcome or another
     ▪ The structure is IF(*test, result if true, result if false*)
   o ISNUMBER
     ▪ Tells you whether a cell contains a number or not
     ▪ This function just takes one argument (the cell you are wondering about)
   o &
     ▪ Lets you "glue" the text contents of cells together
     ▪ Goes between the cell references

Click on cell J6 and try writing a general formula that will work, not just for Alabama, but for every other state as well.

Here the first draft:

=IF(ISNUMBER(B6),A6,A6&B6)

Looks good so far—you get "Alabama" in J6.

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4											
5	Revenue									State	Number of Properties
6	Alabama	62	95%	420,200	$		1,767	1.50%		Alabama	
7	Alaska	2	100	128,500	307	0.3					
8	Arizona	87	98	619,500	3,069	2.7					

Now copy down. Whoops—you're missing a space between the two words in two-word states. So adjust your original formula as follows:

=IF(ISNUMBER(B6),A6,A6&" "&B6)

Nebraska
Nevada
NewHampshire
NewJersey
NewMexico
NewYork
NorthCarolina
NorthDakota
Ohio
Oklahoma

You're manually putting a space in between the two words. Recopy this formula down, and you're all set. Widen the column appropriately.

7. Since the $ issue is only in rows 6 (Alabama) and 56 (Totals/Average), do some manual clean-up. Drag and drop F6 on top of E6. Since E6 is not currently empty, you'll be asked "Do you want to replace the contents of the destination cells?" Click OK.

			State	Number of Properties	Percent Leased	Approximate Leasable Square Feet	Rental Revenue for the Quarter Ended March 31, 2012
420,200 $		1,767	1.50%	Alabama			
128,500	307	0.3		Alaska			
619,500	3,069	2.7		Arizona			
92,400	285	0.2		Arkansas			
2,670,100	15,337	13.4		California			
504,200	1,868	1.6		Colorado			
269,100	1,092	1		Connecticut			
33,300	433	0.4		Delaware			
1,881,000	7,754	6.8		Florida			
1,251,000	4,908	4.3		Georgia			
0	0	0		Hawaii			
80,700	320	0.3		Idaho			
1,335,900	6,043	5.3		Illinois			
799,000	3,610	3.2		Indiana			
290,600	1,019	0.9		Iowa			
642,900	1,350	1.2		Kansas			
134,700	653	0.6		Kentucky			

Microsoft Office Excel

Do you want to replace the contents of the destination cells?

[OK]   [Cancel]

As for the 1.50%, notice that this number is a percent, while the other numbers are not technically percents. So just type "1.5" into F6, to be consistent with the other numbers in the column. Delete the contents of G6.

Do the same things in row 56, so that the value 114,566 is in E56, 100 is in F56, and G56 is empty.

Finally, change C6 from "95%" to "95" (first, change the Number Format to General, then type in "95"). Do the same for C56, so that it contains 97.

8. Set up formulas to read in all the other numbers. For instance, in each cell of column K, you want a formula that will copy in the corresponding cell of column B if that cell contains a number, and if not, you want the cell in column C.

Here's a first take on what the formula in K6 would be.

=IF(ISNUMBER(B6),B6,C6)

This formula will copy down in its column correctly. But you can write just one formula and copy it in *all* the columns. The formula above is *almost* correct, but if you copy it to the right, the

ISNUMBER test is done on the wrong cell. So you want to lock the column (but *not* the row) of the cell tested in ISNUMBER. The way you do that is by putting the dollar sign lock ($) on the column only. This is called **mixed addressing** (part relative, part absolute).
Type the following formula into K6:

=IF(ISNUMBER($B6),B6,C6)

By locking the B column for the cell you're testing but *not* the row, you get the row to change correctly (but not the column) as you copy this formula across to column O and down to row 56.

Spot-check against the original PDF. Do the numbers show up right? In particular, check boundaries of different cases—Pennsylvania (one word) and Rhode Island (two words), for instance.

North Carolina	93	100	570100	2863	2.5
North Dakota	6	100	36600	63	0.1
Ohio	134	96	1122100	3941	3.4
Oklahoma	35	94	752400	1426	1.2
Oregon	20	100	384200	1236	1.1
Pennsylvania	103	98	907200	4137	3.6
Rhode Island	3	100	11000	37 *	
South Carolina	98	98	371400	2294	2
South Dakota	10	100	89800	186	0.2
Tennessee	127	96	737500	2901	2.5
Texas	213	98	3122600	10488	9.1

Again, you might prefer to do manual clean-up, but it's just a question of size. If the original table is 500 rows, not 50, then you definitely want to use formulas to clean up your data.

9. Improve the number formats of columns M and N by adding in commas as thousand separators.

Here's how. Select columns M and N, then click on the little arrow at the bottom of the Number part of the Home ribbon. In the pop-up window, select the Number tab, then Number, then set Decimal places to 0 and check the Use 1000 Separator box.
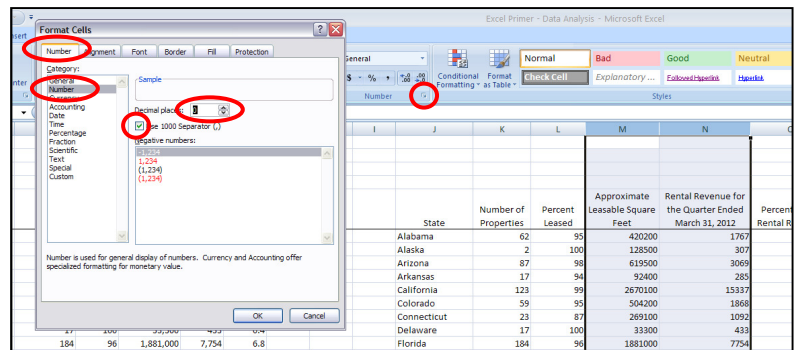
10. Now, save this whole sheet as a backup. Control-drag and drop the sheet after the "1b" sheet. Sheet "1c. Backup" contains this version of the sheet as well.

Why back up now? It's just a preference, perhaps, but I (the author) prefer to do data analysis on cleaned-up source data that has been hard-coded, with its messy origins consigned to the archives. So while I like to keep copies of the nice formulas I wrote to extract and line up the numbers correctly, I don't want to always have the extraction formulas around in the next stage. So I like to turn those formulas into numbers.

11. Go back to sheet "1. Start Here" and select cells J6 through O56. (Keyboard shortcut: click on J6, then hold down both Shift and Control and hit the right arrow key and then the down arrow key. Shift extends the selection, Control jumps to the end of the data.)

Hit Control-C to copy the cells (you get the dotted moving border), then hit Paste Values (*not* regular Paste). This action copies the values right over the formulas. You can't tell, because the values were already being displayed, but something big has happened underneath.
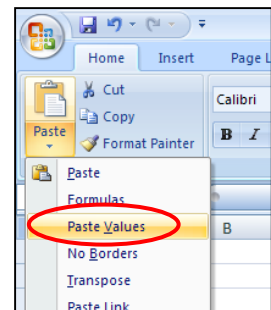
Hit Escape and check the contents of any of the cells you just messed with by clicking on it. Notice that now you have only numbers.

Obviously, only take this course of action when you're good and ready!

Delete columns A through I, since you don't need them anymore.

12. Next, consider column F, which contains Percentage of Rental Revenue. Those numbers are heavily rounded—and in theory, they are derived from the numbers in column E. Consider cell F27, which contains 1. That 1 indicates the 1%, rounded, that 1,197 (in cell E27) represents of 114,566 in cell E56.

Thus, it would be good to replace the rounded numbers in column F with more precisely computed numbers—that is, formulas. This may seem to clash with what we just did (killing off the extraction formulas), but the purpose is to make derived numbers as accurate as possible. First, you make the fundamental source numbers (in columns B through E) as clean and airtight as possible. Then you re-derive any derived numbers. This re-derivation confirms and reconciles what you know about the data. (Also, you get rid of pesky anomalies like the asterisk in cell F44.)

Put "=E6/$E$56" in cell F6. Change the Number Format to Percentage and increase the number of decimal places to 2.

Copy this cell down the column, replacing all the other cells down through F56.

13. We should also turn column C into real percentages. Here's a quick way to do so.

Put "=C6/100" into cell G6. Change the Number Format to Percentage. Notice that since these percents are fundamental (they're not derived from any other numbers in the data), there's no benefit to increasing the decimal places, so leave them alone.

Copy cell G6 down to G56. Double-clicking the black square is the fastest way to do this copying job.

With those cells still highlighted, hit Copy. Click on cell C6 and—be careful here!—hit Paste Special, then choose "Values and number formats".

You've now replaced numbers that weren't technically percents with true percent values.
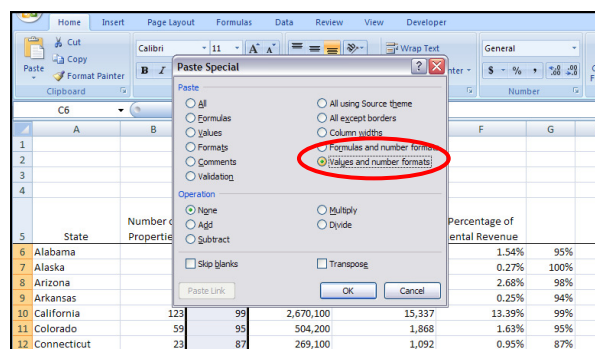
Now erase G6:G56, which was just your scratch work anyway.

14. The Total/Average row should be derived where possible too.

    Replace the hardcoded 2631 in B56 with the formula "=SUM(B6:B55)". The number 2631 is still displayed—no error there.

    Since a thousands separator would be nice, Format Paint cell D56 onto B56, which should now display 2,631.

    Copy cell B56 to D56 and to E56, since they are sums as well. Again, the displayed numbers do not change at all, verifying that those computations in the original PDF were correct.[14]

    For now, leave alone the 97%, although we will be able to compute that number as well and increase its precision somewhat.

    One other quick task: insert a blank row between the Wyoming row and the Total/Average row. Do this by right-clicking on row 56 and choosing "Insert". Why? Later, when you are running Sorts and Filters and creating Graphs and Pivot Tables, you won't want to include the totals by accident. A blank row between the body of the table and the total row helps prevent that.
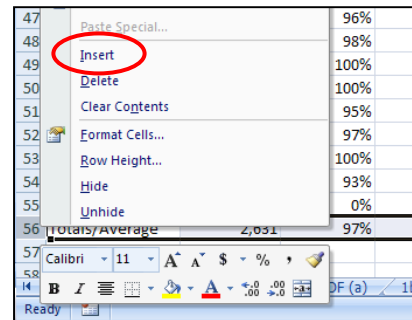
15. Finally, go back to the PDF and copy over the title, the introductory text, and the numbered footnote. (You no longer need the asterisked footnote.) Put these into whichever cells you like at top and at bottom.

   At this point, you should have something very close to the sheet labeled "2. Cleaned-up Data".

16. By the way, here's how to deal with the situation in "1b. Direct from PDF (b)". Click on that sheet. This data is more compacted, with an entire row of the table crunched into one cell.

    It would be possible to use "Text" functions to extract this data into separate cells. However, the formulas would be cumbersome to write. A better initial strategy would be to use **Text to Columns**, an Excel tool designed just for this purpose.

    First, clear the contents of rows 5 through 18, to get rid of the distracting headers. Select column A. Click on the Data ribbon, and choose Text to Columns in the Data Tools area of the ribbon.

    A dialogue box appears. Step 1 of 3 asks

---
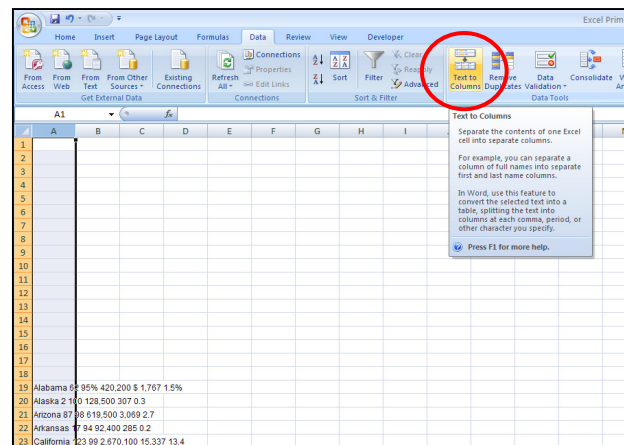
[14] "Trust but verify" is good practice when analyzing data. One of my old bosses used to ask me and my colleagues, "Are you willing to fall on a sword for those numbers?" I was too cowardly to challenge the basic premise of the question, but I appreciated his rigor, and soon I learned to verify the computations of others.

you whether your data is Delimited or Fixed Width. You can answer that: your data is delimited, or separated, by spaces. Excel has already guessed that's the case, so click Next to confirm.

Step 2 of 3 asks you to set what those delimiters are. Make sure that there's a green check next to "Space" in the Delimiters list. Also make sure that there is *not* a green check next to "Comma".

Below, you can scroll the Data Preview up and down to verify what you'll get. If, for instance, you *do* place a check next to the "Comma" delimiter, you'll incorrectly break a number such as "619,500" into "619" and "500".

In Step 3 of 3, you can change the Data Format of each column. Usually, you don't have to do anything here—Excel is almost always smart enough to figure out that text is text, numbers are numbers, etc. So just leave everything in "General" data format and click Finish.

The results are essentially what you have in "1a. Direct from PDF (a)." From that point, you still have a lot of data clean-up to do, but you've avoided a lot of extra work.
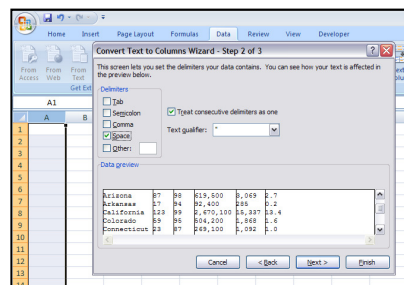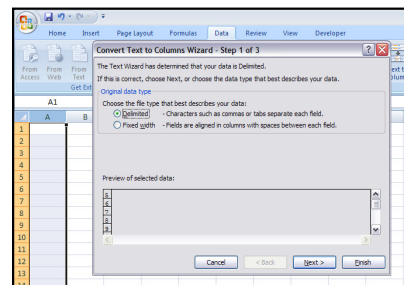
ⓣ *Task:* **Derive new numbers from data.**

Where to work: File – **Excel_Primer-Data_Analysis**
Sheet – **3. Next Starting Point**   ← Do your work on this sheet.
You'll also work with sheet **3a** and **3b**.

Example outcome: Sheet – **4. Analyzed Data**   ← Check your work against this sheet.

⌐ *Process*

Compute new columns.

1. Open the file *Excel_Primer-Data_Analysis* and click on the sheet "3. Next Starting Point". This should be identical to sheet "2. Cleaned-Up Data". Notice that we've zoomed out a little bit, which you can easily control in the View ribbon or at the bottom right of the window, with the same slider that controls the Zoom level in a Word document.

    If sheet 3 has already been messed with, then just make a new copy of sheet 2 and start working from that.

    The task now is to use the given data to figure out "next-step" numbers. Next-step numbers are simple extensions of numbers you are given. For instance, if you know that exactly 85% of the

people in a gymnasium like chocolate ice cream, and you also know that there are 200 people in the gym, then you should *automatically* figure out that 85% of 200, or 85% × 200 = 170 people in the gym like chocolate ice cream. Then you can also figure out that 200 – 170 = 30 people in the gym do *not* like chocolate ice cream.

Next-step numbers are derived from other numbers by simple arithmetic. The two relationships above are very common:

> *Addition*   *Multiplication*
> Part A + Part B = Whole   Percentage × Whole = Part

Study the table in the spreadsheet and ask yourself what next-step numbers you could compute.

Column B gives you Number of Properties, and column C gives you Percent Leased—that is, Percent *of those properties that are leased.* So, if you multiply these numbers (Percentage × Whole), you get the Part: the Number of *Leased* Properties.

2. Center column G and put "Leased Properties" in G5.

In G6, put "=B6*C6" and hit Enter. Change the Number Format to General, if the number accidentally displays as a Percentage. You should see 58.9.

Hmm. That can't be quite right: Realty Income's business model involves leasing whole properties. So it's got to be a rounding issue—the Percent Leased in column C was only reported to two significant figures[15], and so what's likely to be the truth is that exactly 59 properties are under lease in Alabama, and 59/62 is only approximately 95%. (You can check, using Excel as a glorified calculator, that 59/62 ≈ 0.9516…)

Here's how to correct the calculation: use the function ROUND to round the result to zero decimal places (i.e., to an integer). That will get us much closer to the truth.

Replace the formula in G6 with this one:

=ROUND(B6*C6,0)

The second input (the 0) tells ROUND to round to zero decimal places. Now you see 59 in G6.

Copy the formula down the column to G55 by double-clicking the black square.

---

[15] Remember those from chemistry class? Those were the days, weren't they?

Copy the SUM formula from E57 into G57. You should see 2,543. Center the cell.

3. Now you can also count the Properties *NOT* Leased, which is just Total minus Leased.

   Center column H and put "Non-leased Properties" in H5. Apply Wrap Text to H5.

   In cell H6, put "=B6-G6" and hit Enter. Apply G6's format to H6 with Format Painter. You should have 3 (= 62 – 59) in cell H6.

   Copy H6 down all the way to H55, and copy the SUM formula from G57 into H57. You can see that there are 88 properties that Realty owns but doesn't have under lease.

Scrolling up, you might spot which states those 88 properties are in. An easy way to flag those states is with Conditional Formatting. Highlight H6:H55 and click on Conditional Formatting in the Home ribbon. Choose Color Scales, then the Red-Yellow-Green scale (it's good to flag "bad" numbers as red, and here, red = high).

The red shows you that Georgia is leading the pack with non-leased properties (9), followed by Florida with 7. Those absolute figures are not necessarily obvious from a first glance at columns B and C. In fact, if you apply conditional formatting to column C to flag the states with the lowest *percent* leased, then the colors (and your impression) are immediately skewed by the 0%'s in Hawaii and in Wyoming. Those outlying figures aren't that important, though—Realty doesn't have any properties in Hawaii and only 1 in Wyoming. Fine, it's not great that the Wyoming property isn't rented. But 9 out of 88, or over 10% of all non-leased properties that Realty has, are in Georgia. You could hazard that 10% is a substantially higher proportion than Georgia represents of the US population, without going & looking up the figure. So these non-leased properties do not seem to be evenly distributed across the states by population.

All this is to say that you could make an early recommendation if you had to: before purchasing more properties in Georgia (and in other states with high numbers in column H), Realty should examine *why* those particular properties are not under lease.

Even simple arithmetic operations such as the ones done here can yield surprisingly useful insights.

4. What else would be useful to figure out? So far, the computations have all involved the same unit (the count of properties). You can also figure out ratios of dissimilar units. For instance, you have a count of properties, and you have total square feet. You could figure out the average *square feet per property* in each state, to see how the states vary by typical size of property.

Center column I and put "Square Feet Per Property" in I5. Apply Wrap Text and widen as desired.

In cell I6, put "=D6/B6" and hit Enter. Apply E6's format to I6 with Format Painter. You should have 6,777 (= 420,200/62, roughly) in I6.

Copy I6 down to I55. Zounds! There's an error in I16—because Realty doesn't have any properties in Hawaii, we're trying to divide 0 by 0, and that's a no-no. So the "#DIV/0!" error appears, and if any later formulas depend on this column, the error will cascade.

One solution would be to manually override by typing in 0. That's bad practice. Hardcoded fixes that you "sneak" into columns are very, very easy to miss later on, and if the numbers are updated producing a 0 somewhere else (say Realty sells its one Wyoming property), then you have to patch again.

5. A better solution is to change the formula slightly to catch the error. All you need is an IF function: IF the value in B6 is 0, then put in a 0, otherwise do D6/B6. Put this formula in I6:

=IF(B6=0,0,D6/B6)

Recopy the formula down. Now you get a 0 automatically in Hawaii, and you can reuse these formulas down the road. Try as best you can to future-proof your spreadsheets.

Copy I55 to I57, to get the average square feet per property across all 2,631 properties. The result is 10,406 square feet per property.

Apply Conditional Formatting: Data Bars to the range I6:I55. Unfortunately, the average size of the 2 Alaskan properties (over 64,000 square feet) is so much larger than the average in other states that the other differences are obscured.

You could also argue that it would be better to make the Hawaiian output a blank cell, so that its 0 doesn't count at all. That's an easy edit to the formula—just replace the second 0 with an empty quote, "", but that doesn't fix the Alaskan outlier problem. Moreover, without more info about the business, it's not necessarily obvious what to do with these different average sizes.

6. Let's figure out Revenue Per Square Foot. Center column J and put "Revenue Per Square Foot" in J5. Apply Wrap Text and widen as desired.

In cell J6, put "=E6/D6" and hit Enter. Hmm—you get 0.0042… That seems very small, less than a cent per square foot. Even for just 3 months in an economic downturn, that seems cheap. Don't forget: the revenue is in *thousands* of dollars. Always gut-check the results of your calculations!

Change the formula to "=E6*1000/D6". The result is much more sensible: 4.205… or about $4.21 per square foot. Change the Number Format to Currency, and copy the cell down to J55.

Yikes, we need an error checker again for Hawaii. Edit the first formula again:

=IF(D6=0,0,E6*1000/D6)

Copy down again; now you're fine. Again, you could write it as "=IF(D6=0,"",E6*1000/D6) to leave a blank cell, if you really want.

Copy J55 to J57. The average revenue per square foot is $4.18.

Apply Conditional Formatting: Data Bars (green this time) to the range J6:J55. You can see that Realty is really getting a lot of revenue per s.f. in Delaware ($13.00), far above any other states. It's probably not because Delaware was the first state to ratify the Constitution, back in 1787.

You'd want to find out what sort of properties Realty has in Delaware and in Vermont ($10.24), and what market conditions allow Realty to extract such higher rents per square foot there. Some of those conditions are probably unique to these small states, but there may be lessons applicable in other locations. Realty undoubtedly knows all this, but you'd better have the conversation with them so that *you* know the reason for these differences.

Leased Properties	Non-leased Properties	Square Feet Per Property	Revenue Per Square Foot
59	3	6,777	$4.21
2	0	64,250	$2.39
85	2	7,121	$4.95
16	1	5,435	$3.08
122	1	21,708	$5.74
56	3	8,546	$3.70
20	3	11,700	$4.06
17	0	1,959	$13.00
177	7	10,223	$4.12
134	9	8,748	$3.92
0	0	0	$0.00
11	1	6,725	$3.97
99	2	13,227	$4.52
78	3	9,864	$4.52

You can make other arithmetic computations from what you've got (e.g., Revenue per Property), but let's go get some outside data that you might have already thought about: the population of each state. This will give you a way to compare differently sized states.

7. Click on the worksheet "3a. Pop Data". We've already pulled data from the US Census at census.gov and put it in this sheet. (Fortunately, you can download US Census data in the form of Excel workbooks, so you have far fewer data clean-up problems. The biggest issue with the Census is that there's so much available, you can get a little lost looking for what you need.)

The population of each state as of April 1, 2010, is listed, as well as the geographical Census "division" that the state belongs to.

You can just copy this data over. Both sheets are sorted by State (more on Sorts later), but you should still verify that you don't miscopy. Here's how to create an error-checker in this circumstance.

Select the range A7:C57. That is, all three columns (*including* the State column) and all rows, including the headers.

Hit Copy, then switch back to "3. Next Starting Point" and Paste starting at K5. Widen the columns as necessary to see all the data.

Square Feet Per Property	Revenue Per Square Foot	State	Population	Division
6,777	$4.21	Alabama	4,779,736	East South Central
64,250	$2.39	Alaska	710,231	Pacific
7,121	$4.95	Arizona	6,392,017	Mountain
5,435	$3.08	Arkansas	2,915,918	West South Central
21,708	$5.74	California	37,253,956	Pacific
8,546	$3.70	Colorado	5,029,196	Mountain

Now you have *two* columns labeled State. You can create a temporary column to check whether these match.

Put "Temp Match" in N5. In N6, put the following formula:

=IF(A6=K6,"",1)

This formula says "If the two states in my row are the same, give a blank cell; otherwise, if they're *not* the same, put in a 1."

ulation	Division	Temp Match
79,736	East South Central	
0,231	Pacific	
92,017	Mountain	
15,918	West South Central	
253,956	Pacific	
29,196	Mountain	

Cell N6 should look blank, quite properly! Copy it down to N55.

878,102	Mid-Atlantic	
35,483	South Atlantic	
72,591	West North Central	
536,504	East North Central	

Count: 50

Notice the "Count: 50" at the bottom. You might have noticed before that when you highlight a range, different "on-the-fly" computations show up there. "Count: 50" tells you that 50 cells with *something* in them are highlighted. The fact that there's *no sum* tells you that all the states match. (Try highlighting another range with numbers in the cells, and you'll quickly see that a sum appears on the fly, along with an average.)

Such an error-catcher is really useful when you're dealing with thousands of rows.

Delete column N (Temp Match), because it was only there for this reason. Likewise, delete column K (States), which is redundant. Copy cell D57 into the new cell K57, to get the total population of the US (308,143,815).

8. In the new, blank column M, we'll put Properties per Capita—or better, Properties per 100,000 Population. This way, you aren't dealing with tiny decimals, because the number of properties in a state is much, much smaller than the population.

Center column M and put "Properties per 100K Pop" in M5. Apply Wrap Text and widen as desired.

In cell M6, put "=B6*100000/K6" and hit Enter. The 100000 is there so that the result is per 100,000 people (and inside a formula, Excel doesn't want any comma separators inside of numbers). Reduce the number of decimal places to 2. Notice that we don't need to put in a "divide by zero" error catcher, because every state has a non-zero population (and always reasonably will).

Copy the formula down to M55, and also copy it to M57. The average is 0.85 properties per 100,000 people.

Highlight M6:M55, and choose Conditional Formatting. For a change of pace, choose Icon Sets, specifically 5 Quarters, as shown at right.

This particular set of icons shows fully filled circles for the highest values, helping you spot at a glance where Realty has many properties (relative to the population) and where it has only a few properties.
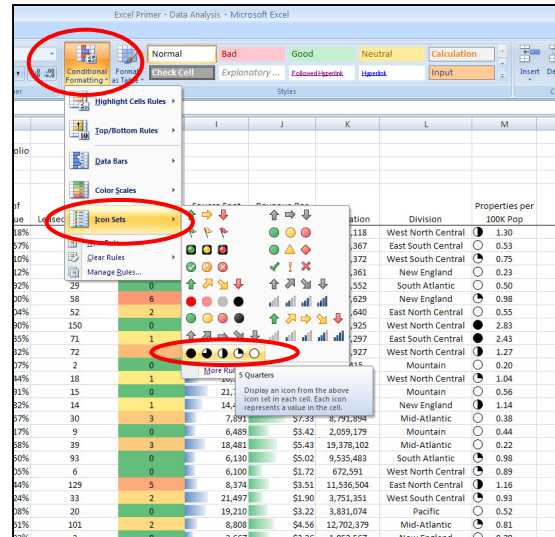
Interpretation is not automatic. Should Realty pursue opportunities in states where it has relatively few properties per capita? Maybe, if market conditions are similar to other states with higher penetration. Maybe *not*, if market conditions are different. After all, Realty may have already built a higher concentration of properties in states where the market is favorable. For instance, Delaware with its $13.00 per square foot rental rate has a high density of properties (1.89 per 100,000 population).

Compute statistics.

9. You've already computed averages and totals in row 57. However, you can refine one number you might have forgotten about: the average Percent Leased.

   This should *not* be the average of cells C6:C55, of course. Rather, you need to take the total Leased Properties (2,543, in cell G57) and divide by the bigger total Number of Properties (2,631, in cell B57).

   Put "=G57/B57" in C57. There's no evident change: the percent rounds to 97%, the value you had before. If you increase the number of decimal places, however, you can squeeze another reasonable digit out (96.7%, say), if that were important.

   The median is a good number to contrast to the arithmetic mean (which is properly computed as Total over Total, as shown above). When the mean is above the median, you probably have a positive "skew" (a few big numbers and a lot of small numbers), and vice versa. The median function is nicely called MEDIAN.

Put "Median" into A58. Put "=MEDIAN(B6:B55)" into B58, then copy that cell across to M58.

L58 gives you an error, since column L contains text. Delete L58. To make the formats comparable to the cells above, highlight B57:M57, select Format Painter, then highlight B58:M58. Excel figures out which cell formats go where.

55	Wyoming	1	0%	5,400	0	0
56						
57	Total/average	2,631	96.7%	27,376,900	114,566	100
58	Median	34	97.5%	335,000	1,248	1
59	Footnote (1):					
60	Includes rental revenue for all properties owned by Realty Income at March 21, 2012, including revenue f

The other function worth introducing is CORREL. This gives you the correlation coefficient between *two* columns of numbers. Do they tend to *rise* together (positive correlation), *fall* together (negative correlation), or is there no such pattern?

10. Let's see how Revenue Per Square Foot is correlated with Properties Per 100K Pop. If this is strongly positive, then Realty would seem to be concentrating its properties (relative to population) where it's already successfully getting high rents (on a per square foot basis). Conversely, if the correlation coefficient is strongly negative, then the high rents happen where Realty *doesn't* have many properties relative to population.

Put the following formula into M2:

=CORREL(J6:J55,M6:M55)

The number comes back as 0.288… That's not huge, but it's not tiny either (the

fx	=CORREL(J6:J55,M6:M55)

	C	D
	rth certain state-by-state information r	
	s in thousands):	
	Percent	Approximate Leasa
per of erties	Leased	Square F
2	95%	420,
2	100%	128,
7	98%	619,
7	94%	92,
2	99%	2 670

	J	K	L	M
				0.288225115
	Revenue Per			Properties Per
	Square Foot	Population	Division	100K Pop
7	$4.21	4,779,736	East South Central	1.30
0	$2.39	710,231	Pacific	0.28
1	$4.95	6,392,017	Mountain	1.36
5	$3.08	2,915,918	West South Central	0.58
2	$5.74	37.253.956	Pacific	0.32

maximum is 1, and that never happens in the real world). There seems to be some mild push on Realty's part to concentrate properties where it can demand high rents. That said, there may be states in which Realty can get high rents but hasn't penetrated very far, for whatever reason, and also states in which Realty gets low rents but has many properties per capita.

(By the way, what's *not* reflected in these numbers is the cost of capital—how much money Realty *invested* per square foot to own these properties, and how "expensive" that money was for Realty to get from investors. After all, a low rental rate per square foot on property X could be more profitable than a high rental rate on property Y, because X was far cheaper to purchase than Y. To do this analysis right, you'd want to know these costs—but Realty would not release those numbers publicly unless forced to!)

Use lookup functions.

11. What are "lookup" functions? They help you join tables together. For instance, the population data could have been "looked up" rather than copied & pasted, if we used the State columns in each table as the lookup key.

An example will illustrate. Right now, in column K you have "Division"—the 9 divisions that the US census divides the 50 states into. Each of those 9 divisions falls within one of 4 regions: Midwest, Northeast, South, and West.[16]

On sheet "3b. Lookup Divisions", there's a little table showing which region each division falls into. Since you already have each state's division in the big state-by-state table, can't you just tell Excel to put the corresponding *region* into another column?

Sure you can. That's what lookup functions do.

Go back to sheet "3. Next Starting Point" and put "Region" into N5. In cell N6, start the following formula:

=VLOOKUP(L6,

The first input is L6, the "lookup value" as Excel calls it. That's the "key"

Division	Properties Per 100K Pop	Region
East South Central	1.30	=VLOOKUP(L6
Pacific	0.28	VLOOKUP(**lookup value**, table_array, col_index_num, [range_lookup])
Mountain	1.36	
West South Central	0.58	

that exists *both* in this table (the destination) and in the other table (the source).

The second input is the source table. This is the first time you're creating a formula that spans different sheets. Make sure that your cursor is at the end of the formula as you've written it so far (after the comma), then click on the tab taking you to "3b. Lookup Divisions".

Notice that the formula gets updated like so:

=VLOOKUP(L6, '3b. Lookup Divisions'!

The part before the exclamation mark tells Excel what *other* sheet to look for a cell reference on. We haven't needed this up till now, because every cell reference in a formula referred to a cell on the same sheet.

Select A5:B13. Make the reference absolute on both ends of the range by hitting function key F4 (or typing in $ everywhere as shown below). The formula should now look like this:

=VLOOKUP(L6,'3b. Lookup Divisions'!$A$5:$B$13

Alternatively, you could name the source table range something such as "DivisionLookup"—then the reference is automatically absolute, and it's not so cryptic.

At this point, type in the last bit of input: ,2)

The comma separates off the last input. The 2 tells Excel to "return" (bring back) the value from the *second* column of the source table. The final function should look like this:

=VLOOKUP(L6,'3b. Lookup Divisions'!$A$5:$B$13,2)

---

[16] Not exactly the four points of the compass, but they serve.

Hit Enter. Cell N6 should now display "South", the Region corresponding to the East South Central Division.

Division	Properties Per 100K Pop		Region
East South Central	◗	1.30	South
Pacific	○	0.28	West
Mountain	◗	1.36	West
West South Central	◕	0.58	South
Pacific	○	0.33	West
Mountain	◗	1.17	West
New England	◕	0.64	Northeast
South Atlantic	●	1.89	South
South Atlantic	◕	0.98	South
South Atlantic	◗	1.48	South
Pacific	○	0.00	West

Copy N6 down to N55. Each formula takes the division in its row and uses it in the lookup table to find the region & bring it back to the big table.

VLOOKUP is super-powerful, though it does have a couple of disadvantages: the source table has to have the lookup key in the first column, and that column has to be sorted in ascending order. If that's not the case, VLOOKUP screws up. That said, you can almost always keep your lookup table fixed, so just set it up right & leave it untouched. (This is a reason to put it on a separate sheet, untrammeled by further activity.) VLOOKUP is probably the only lookup function you need.

If you're curious, here's how to build a lookup that *doesn't* require the lookup table to be sorted in ascending order:

> =INDEX(return_column, MATCH(lookup_value, lookup_column, 0), 1)

> Return_column: absolute address of the column you want to return values from (in this case, the column in the lookup table that contains Region).

> Lookup_value: L6, the single cell you're using as the key. Always use a relative address here.

> Lookup_column: absolute address of the column you're matching the key to. In this case, the column in the lookup table that contains Division).

ⓣ *Task:* **Sort, filter, graph, and pivot.**

Where to work: File – **Excel_Primer-Data_Analysis**
                     Sheet – **5. Final Starting Point**   ← Do your work on this sheet.
                     You'll also work with sheet **5a**.

Example outcome: Sheet – **6. Final Data**   ← Check your work against this sheet.

⌐ *Process*

Sort & filter data.

1. Make sure you're on the "5. Final Starting Point" tab.

   The Sort & Filter tools require that you have a clean data table. What does that mean? It means that your data should all be in a single rectangular *island*—one block of cells, with the data all touching (you can have occasional blank cells, but no "rivers of emptiness" cutting the island in

two), and an empty channel all around. Nothing else should touch the island. You need at least one blank row or column separating the data island from other information in the spreadsheet.

This is why we inserted a blank row between the Wyoming row and the Total/Average row. This is also why above the column headers in row 5, row 4 is blank.

In addition, your data table should be organized just as we've done—with **records** in rows and **fields** in columns. That is, each row contains a single *record* about something (in this case, the properties in each state). Meanwhile, if you look at any column in the table, it always contains the same kind of information about each record: the Percent Leased, for instance, or the Division. That's what a field is. Row 5 contains the headers corresponding to each field.

If you have single record in each row and a single field in each column, forming a rectangular block of data isolated from anything else on your spreadsheet by at least one blank row or column—then you're ready for Sort & Filter.

That's exactly what we have on this sheet in the range A5:N55, so we're ready to go. You just need to recognize what Excel wants, for when you build your own data tables.

2. Click on cell J5 (the header "Revenue Per Square Foot"), then click on the Data tab at the top of the page to bring up the Data ribbon.

In the Sort & Filter area, click the **Sort Z to A** button.

The table re-sorts in **descending** order of Revenue Per Square Foot. Delaware ($13.00) moves into the first position in row 6, followed by Vermont ($10.24), New Jersey ($7.33), and so on down the line.

Of course, you can sort in **ascending** order by clicking the **Sort A to Z** button. You can sort by any column.

You can click anywhere in the relevant column to tell Excel which column you want to sort by.

What if you want to sort by *more* than one column at once? For instance, maybe you want to group the states by Region at the highest level, then group by Division, and finally within a division, you want Population in ascending order.

There are two easy methods. Pick either one.

a. Sort by one column at a time. Which column do you start with? Work from the lowest level to the highest. Sort first by Population in A-Z order. Then re-sort by Division in A-Z order. Observe that within each Division, Population is increasing. When you re-sort by a column that contains *repeated values*, as Division does, then previous sorts break the ties.

    Finally, re-sort by Region in A-Z order. Region is now the **primary** sort. Division is the **secondary** sort. And if you really felt like naming it, Population is the **tertiary**, or third-level, sort.

b. Alternatively, you can do a three-level sort all at once. Click anywhere in the range A5:N55 (so Excel recognizes the data table). Next, click on the big Sort button.

    A dialogue box is opened. In the "Column… Sort by" drop-down menu, choose "Region" for the highest level sort. Notice that the drop-down contains all your column headers.
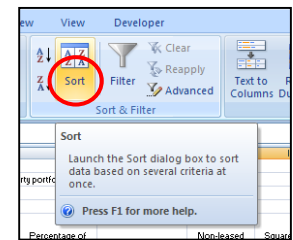
    Click "Add Level" and choose "Division" for the second level. Click "Add Level" once more and choose "Population" for the third level. In all levels, leave Sort On = Values, and leave Order = A to Z, or Smallest to Largest. Click OK.
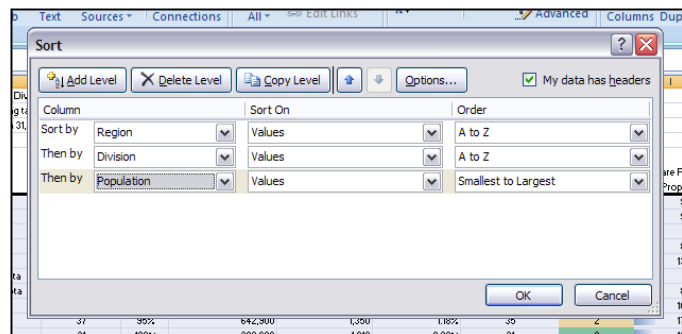
The table is sorted as shown above.

Sorts are simple but powerful. In the early stages of exploring a new data set, don't overlook this humble tool.

3. **Filter** is Sort's cousin. It's most useful when you have subgroups created by repeated values, such as you have in Region and Division.

What if you want to look at all the New England states together? One option is to sort by Division. That groups all the New England states together, but you can be distracted by all the other data.

Filter lets you just look at the New England states, temporarily hiding all the other rows.

Re-sort the table by State, in ascending alphabetical order.

Click anywhere in the data table and click on Filter, the button next to Sort.

Each of the column headers gets a little floating arrowhead. That arrowhead opens up a personal drop-down menu containing all the possible values in that column.

Click on the arrowhead in cell L5, the "Division" header.

By default, all the possible Divisions are checked. By checking and unchecking the options, you can show and hide the various Divisions.

Uncheck the "(Select All)" option to clear all the checks, then put a check back next to New England. Click OK.
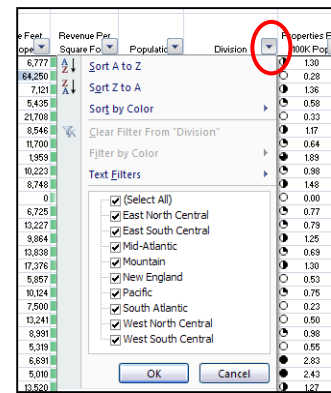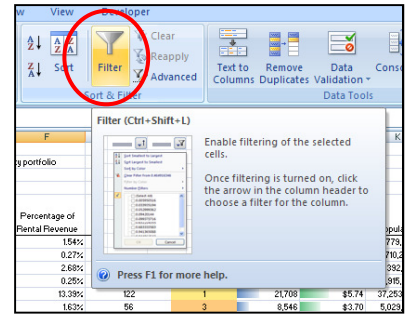
Only the six New England states are shown. You can tell that the Filter is active in the Division column because a little *funnel* (get it? It's like a filter) appears in the arrowhead button.

You can have an active filter in more than one column. Just click on the arrowhead drop-downs in other columns and check or uncheck options to further filter down the data table.

To get rid of the Filter, click on the Filter button in the ribbon to deactivate it.

Filter is worth its weight in gold if you have a giant data table. Filter lets you see little slices at a time, so you can wrap your head around bite-sized chunks. In fact, if you don't have natural groupings (such as Division and Region), it can be worth creating additional derived columns to "bucket up" continuous data, just as the states are bucketed up into Divisions and Regions.

For instance, you might decide that meaningful buckets for "Revenue Per Square Foot" are "$0 to $3", "$3.01 to $4", "$4.01 to $5", "$5.01 to $7", and "$7.01 up". That's fine. Just name a new column (something like "Rev Per SF Groups"), then put the groups in there (you could use a VLOOKUP with a little table, or you could use a nested IF formula: IF the number is less than

$3.01, then give me "$0 to $3"; otherwise, IF the number is less than $4.01, then give me "$3.01 to $4"; etc.).

Bucketing up data might sound like a pain, but if you do so in a meaningful way, then you can apply the real power of Filters (and of Pivot Tables, which we'll get to last) to those buckets.

Graph data.

4. You've already put some visuals into the sheet with Conditional Formatting. Excel can make many more graphs for you, such as bar charts, column charts, pies, etc.

    Click on the Insert tab at the top of the window, bringing up the Insert ribbon. You can see the zoo of graph possibilities in the ribbon.
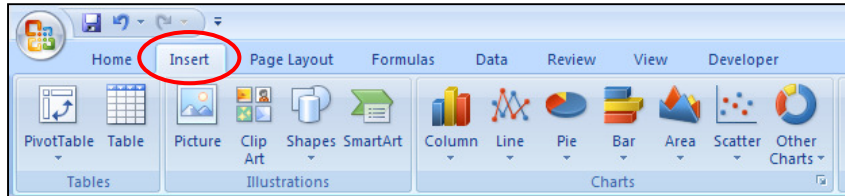
    You can go honestly crazy with all the possible things you can graph. The one example we'll work through now is a **scatterplot**, which puts two columns of numeric data on an x-y plane, so you can really see how the two columns are related.

    Earlier, you computed the correlation coefficient between column J and column M—namely, Revenue Per Square Foot and Properties Per 100K Pop. But that value of roughly 0.29 only provides a very rough guide to the relationship between these two columns. A scatterplot can give a more accurate view.

5. Click on a cell to the right of the data somewhere and then click on the **Scatter** button on the Insert ribbon. In the drop-down menu, choose the first option: Scatter with only Markers.

    An empty graph window appears in the middle of the sheet, floating over the cells. When you put a graph in a sheet, it doesn't live in a cell—it floats on top of them. (You can also create a graph as a completely separate "page" in the workbook.)
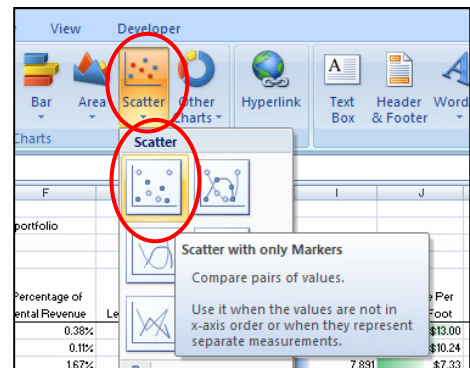
    Notice that three new ribbons have appeared, all having to do with Chart Tools: Design, Layout, and Format. Excel has activated the Design ribbon for you.
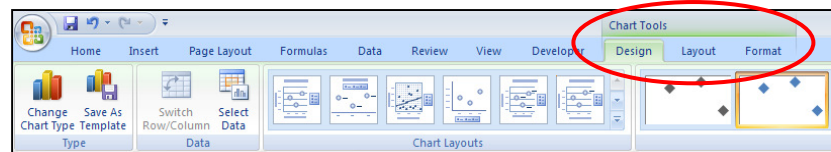
Grab the empty graph window by one of its edges and drag it over to the right, so it's not covering up any cells in your data table. Resize the graph window by grabbing a corner (say, the bottom right) and stretching.
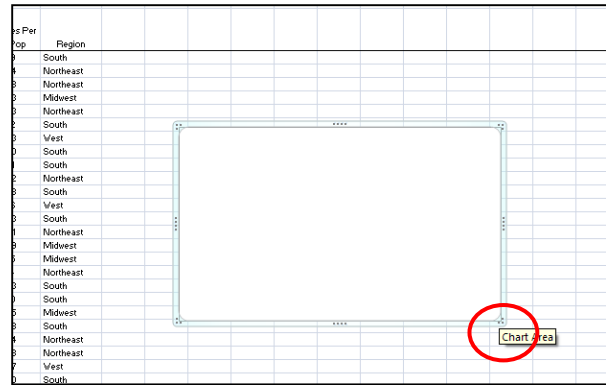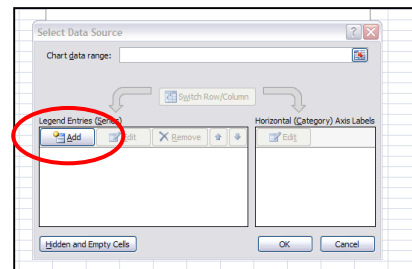
Where's the data? If we had selected it first, it'd show up, but in the case of scatters, it can be easier to add the data after you create the window. In particular, because our data is in non-adjacent columns, we want to be careful.

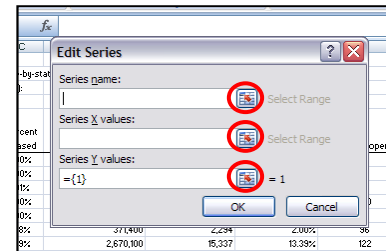In the Design ribbon, click on the Select Data button.
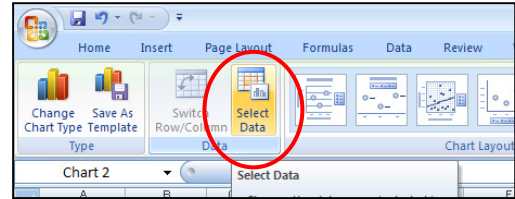
A pop-up window appears: Select Data Source. Click on the Add button under Legend Entries (Series).

Yet another pop-up window appears: Edit Series.

This window gives you the power to choose the Series Name, the Series X values, and the Series Y values.

By clicking on each of the little blue buttons at the right of the entry boxes, you can click-input the references to the cells you want.

We should put the Properties number along the x-axis and the Revenue number along the y-axis. The Properties number is arguably the business *input* (it's more under Realty Income's control), so it should go on the x-axis. In contrast, the Revenue number is more of an *output* or outcome, so it goes on the y-axis.

Put the following references into the entry boxes:

Series Name: ='5. Final Starting Point'!$J$5

(Rather than type that, you can just click on the blue button, then click on J5, then click on the altered blue button to confirm the entry.)
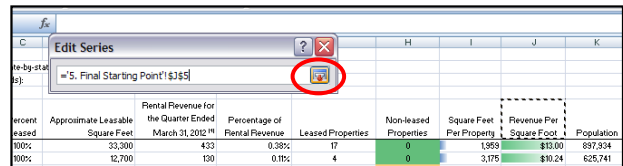
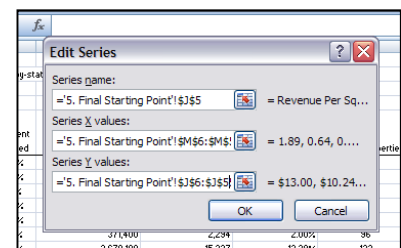Series X values: ='5. Final Starting Point'!$M$6:$M$55

This is the "Properties Per 100K Pop" column (with no header).

Series Y values: ='5. Final Starting Point'!$J$6:$J$55

This is the "Revenue Per Square Foot" column. Click OK once, and

then again a second time to finish.

You get the basic scatterplot you want. You can do lots of cosmetics here (e.g., delete the redundant legend, add axis titles, change the gridlines, and so on) by choosing various options on the three Chart Tools ribbons, but explore those on your own.

For now, just consider the scatterplot itself. What does it tell you?
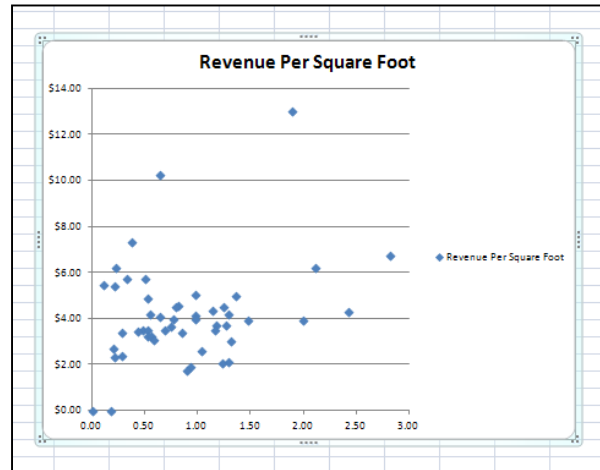
One observation is that the positive correlation you computed earlier (0.29) is probably driven a lot by the outliers on the right side of the graph, such as Delaware and Minnesota with above average numbers on both axes.[17] In fact, if you ignore the outliers, you might even see a negative trend in the main constellation of points.

If you were really digging into this data, you'd want to understand *why* the outliers are the way they are. You'd also want to highlight the big revenue states (e.g., California, Texas). In fact, a type of graph that you could use for *selected* representative states (not all 50) is a **bubble chart**, which is kind of a scatterplot on steroids, with the points each pumped up to various sizes corresponding somehow to importance (e.g., overall revenue from that state). A bubble chart with all 50 states would be overwhelming, but pared down to 8 or 10 key states, it could illustrate key differences between markets.

Create Pivot Tables.

Last but not least for this primer, Excel enables you to create **pivot tables** (a.k.a. pivots), which let you slice and dice your data table in various fun ways. Moreover, pivots are quick to create and mess around with. In fact, they're arguably more suited for data *exploration* than for anything else.

Pivots are best explained with an example.

6. Click anywhere on the data table (within A5:N55). Bring up the Insert ribbon at the top of the window.

Click on the PivotTable button. It's a bit odd that this button is here and not on the Data ribbon, but ours is not to reason why. Choose the Insert PivotTable option.

---
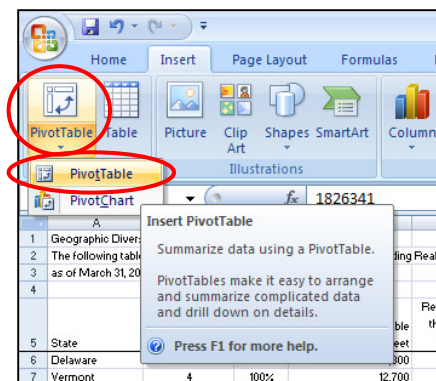
[17] Annoyingly, Excel for some reason does not provide a turnkey, native way to label each point with the proper state's name. On the Microsoft website, the advice given is to open up the hood of Excel and dig into programming a non-trivial "macro." Some third-party programmers have released shareware to accomplish the goal. Lacking these add-ons, you can just manually keep track of the outlier states.

A dialogue box titled *Create PivotTable* pops up, while the data table gets highlighted with the moving dotted border (the "unfinished business" border).

Leave the default options. The reference to the table should be correct (as long as you selected just one cell somewhere in the table, the reference expands to the whole "data island"). Also, you almost always want to put pivots on new worksheets, as Excel prompts you. Click OK.
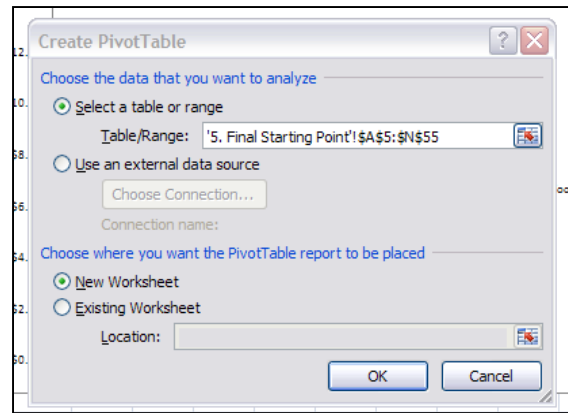
A new sheet appears. Rename it "5a. Pivot" and move it to after the sheet you've been working on. You can compare to "5a. Pivot Sample" as well.

The sheet has two key areas. On the left, you have the blank pivot table. Unlike the floating graph window, pivot tables actually reside in cells. However, pivots have special properties; like genies, they sometimes escape their containers. This is why it's often good to sequester pivots in solitary, so they don't bump up against each other.

On the right side of the sheet, past the sheet itself, you have a brand-new window called the PivotTable Field List. This is the control center for the pivot table you're building. As you do stuff on the Field List, your pivot over on the left will grow, change shape and generally cavort.

The top part of the Field List is the list of all the fields in the original data table, back on the other sheet. By dragging and dropping these field names in and out of the bottom part of the Field List, you'll alter the pivot table.

The bottom part has four panes. Focus on the bottom right, labeled Values. That's where you're going to Count, or Sum, or take Averages, etc.

7. Drag the State field down and drop it in the Values pane.

    Three things happen:

    a. At the top of the Field List, the State field is still there—and it gets a green check next to it. That means you're using this field somehow in your pivot.
    b. In the Values pane where you just dropped the State field, it shows up as "Count of State." This is because the State field contains only text (the state names), so the only math a pivot can do to that is Count. Don't worry—that's plenty useful in and of itself, as we'll see.
    c. Over on the left, the big white concrete pad that was the blank pivot table has turned into a tiny seed of data (50) and one label: Count of State.

    That's a *long* way to go to find out there are 50 states. Rest easy—more revelations await.
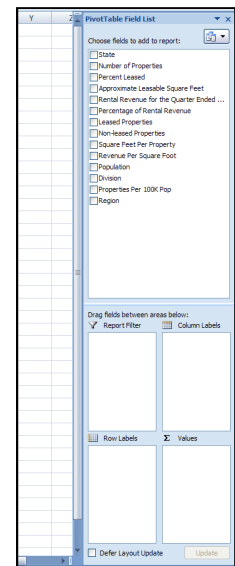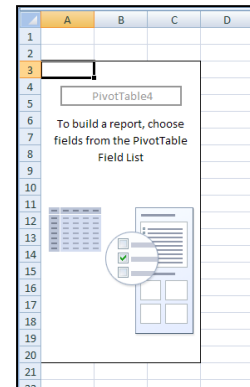
8. Drag the Region field into the top right pane, labeled Column Labels. Notice again the three effects:

    a. In the top half of the Field List, the Region field gets a green check.
    b. Down below, the Region field shows up in the Column Labels pane.
    c. On the left, the actual pivot table has ballooned to the right.

This tells you that there are 12 Midwest states, 9 in the Northeast, etc.Okay, it's cool that the table expanded on its own like that, but the results aren't wow yet. Let's keep going.[18]

	A	B	C	D	E	F	G
1							
2							
3		Column Labels					
4		Midwest	Northeast	South	West	Grand Total	
5	Count of State	12	9	16	13	50	
6							
7							
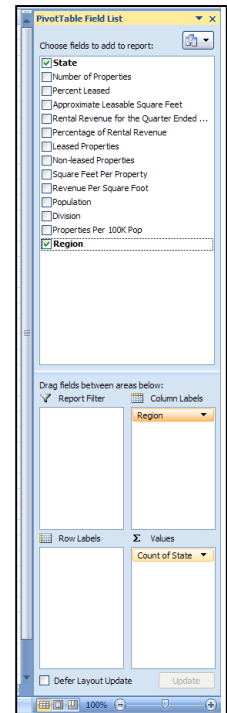
9. Drag the Percent Leased field into the bottom left pane, labeled Row Labels.

Now that's wow. Not the Field List effects (the green check, the fact that Percent Leased is now in Row Labels), but rather the pivot table itself. The table has enlarged downward, and now it's telling you how many states in the Midwest, for instance, have a 95% Percent Leased rate (3). It's almost a scatterplot, or a bubble chart, but with more comprehensibility. The counts sum up both ways, and in fact in the lower right corner, you still have your trusty data seed: 50 states.

At a glance, you can see that the Northeast has a leasing issue relative to the rest of the country—5 states in the Northeast have lower Percent Leased figures than normal. (The two 0%'s in the West are Hawaii, which actually shouldn't count in at all, and Wyoming with its one unrented property.)

	Count of State	Column Labels				
	Row Labels	Midwest	Northeast	South	West	Grand Total
5	0%				2	2
6	87%		1			1
7	91%		2			2
8	92%				1	1
9	93%	1	2			3
10	94%			3		3
11	95%	3		2	1	6
12	96%	3		3		6
13	97%				1	1
14	98%	1	1	2	1	5
15	99%			1	1	2
16	100%	4	3	5	6	18
17	Grand Total	12	9	16	13	50
18						

You can easily flip the pivot by using other fields for rows and columns. You can apply a filter in the fourth pane we didn't use. You can do sums and averages in the Values pane, not just counts. A word to the wise, however: avoid taking averages of ratios. If you want to figure out the average Revenue Per Square Foot in different regions or divisions, you should *never* stick "Revenue Per Square Foot" in the Values pane and tell Excel to take the average. Rather, you have to separately compute the *total* Revenue and the *total* Square Feet in each division or region (the pivot table can do that). Then compute the overall average as a ratio.

Play around with the Pivot Table feature to get a sense of what it can do. Which is a lot.

---

[18] By the way, if you ever click away from the pivot table cells for whatever reason, the Field List vanishes. Not to worry: magic has not left the world. Simply click somewhere on the pivot table, and just as if you'd stepped inside a ring of toadstools, again the Field List shimmers to life.

Conclusion

You've covered a great deal of material at this point. Don't worry if you haven't completely mastered everything. The goal is exposure and awareness, not mastery.

Thanks for sticking with me to the end of this primer. It's not the end, nor is it the beginning of the end. It's just the end of the *beginning*, as is said at commencements.

With that valediction, welcome to Excel, and happy number-crunching!

Chris Ryan
Manhattan Prep