Be a Data Magician

An Excel Workshop for Humanists

Joey Stanley
Linguistics Ph.D student, University of Georgia
joeystanley.com
orcid.org/0000-0002-9185-0048

Presented at the UGA Willson Center DigiLab
Friday, January 27, 2017

This workshop will cover some of the basics of Excel, but will then move quickly on to other topics that will be useful for many people. We will cover the following topics: (1) different versions of Excel; (2) the absolute basics; (3) useful stuff like search & replace and sorting & filtering; (4) the awesome power of pivot tables; (5) getting started with functions; (6) lookup tables, which is actually just the application of one fairly simple function that allows for powerful database-like capabilities involving multiple spreadsheets simultaneously; (7) visualizations and how to make some simple graphs and charts; and if there's time, (8) some miscellaneous little tips and tricks I've picked up along the way.

Download this PDF and accompanying datasets from my website at
http://joeystanley.com/excel

or through the UGA DigiLab at
https://digi.uga.edu/news/be-a-data-magician/

(Updated January 27, 2017)

1 Microsoft Office Versions

Before we get started, let's take a step back and figure out all these different versions of Microsoft Office and, by extension, Excel. There are basically three versions of the name brand Excel that are currently being supported by Microsoft: Office 2016, Office 365, and Office Online. Let's take a look at each of these.

1.1 Office 2016

The first is the most prototypical "computer software"—at least from 10 years ago. This is the kind where you can go to OfficeMax and buy the software in a big box and it comes with a CD and a booklet. Except it's 2017 so you just download the software instead.

Office 2016 is a stand-alone software package. It comes with the latest versions of Word, Excel, PowerPoint, and OneNote. It's a one-time purchase of $150, and you get the software forever on one computer. The bad news is it won't update to newer versions.

1.2 Office 365

This is the most common option. With Office 365 you also get stand-alone software downloaded to your computer. It comes with Word, Excel, PowerPoint, and OneNote but also includes Outlook, Publisher, and Access, which updates automatically to never versions. You can download the software on multiple devices (phone, tablet, etc.) and sync documents between them, and you get 1TB of free storage to do that.

The problem is that, it's available by subscription only, and it's about $70 a year. Luckily, we get it for free through UGA, but as soon as your affiliation is over, you lose the software.

1.3 Office Online

The third option is less common and is an online-based version of the software. If you're unaffiliated with a school or business that has Office 365 and you don't want to fork over the cash, this is a decent alternative because it's free. It's based in your OneDrive account though, which might be limited to 15GB of space or something. I've never used this, but I imagine Office Online is limited in some way, in that it probably doesn't contain all the bells and whistles as the stand-alone software.

orcid.org/0000-0002-9185-0048

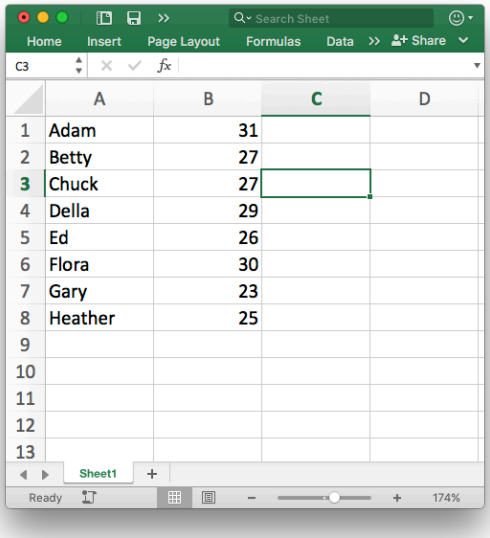
2 Basics

So now let's start getting into Excel. I use Excel 2016 on my Mac, so I can't guarantee things will look the same on your computer, but I imagine things are pretty close. To get started, open up a new Excel spreadsheet.

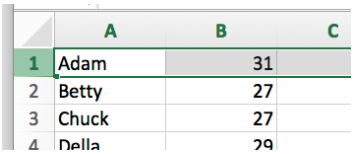
In case you've never used Excel before, this is a spreadsheet software, meaning things are arranged into rows and columns. While there are lots of uses for spreadsheets, what I'm focusing on is Digital Humanities type data. This is typically organized where each row is an observation of some sort, and each column contains information about those observations. In this workshop, I'll be using several data sources, mostly related to linguistics, which the field I'm in, but we'll start out simple and just look a list of some "friends."

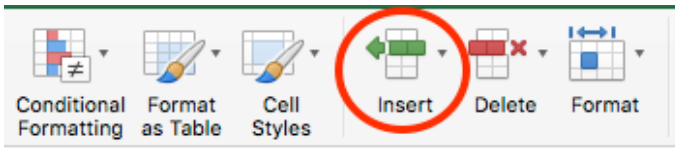
2.1 Entering Text

It's simple to enter text. Click on a cell and start typing. Press tab to move one cell to the right, and press enter to move one cell down. I'll go ahead and enter information going down the first column with the names of me some of my "friends," and another column with their ages.

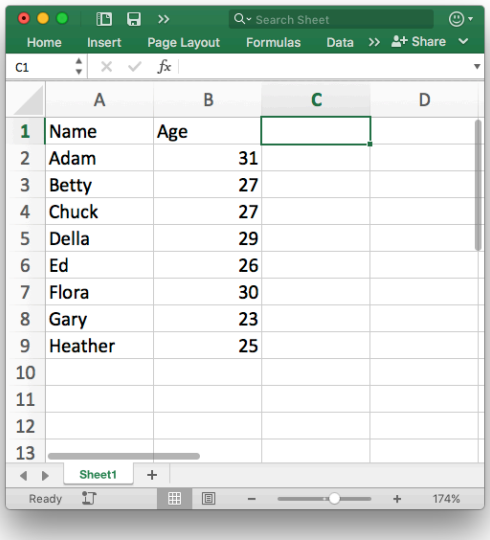


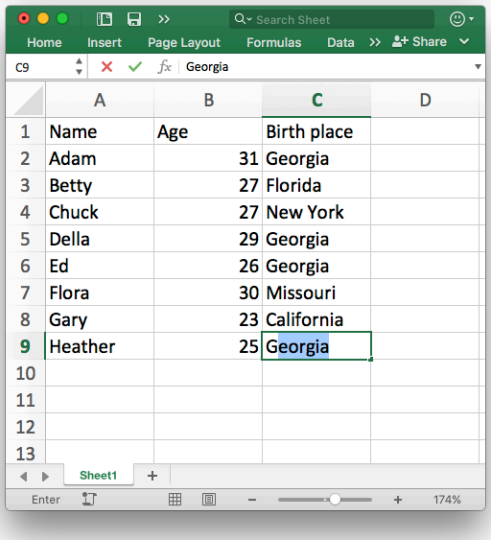
So there you have it. A table of my friends' information. Now, if someone were to look at this, they might not know what these columns mean. Let's add column headers to the table. To do this, we'll need to insert one row above. Highlight the row you want to insert a row over by clicking on the actual number 1 on the far left, and then click on the "Insert" button in Home toolbar.





(Note that if you click on the little triangle/arrow next to the button, you'll get some other options, but the default for the big button is to insert a row above the selected row.) You'll now magically have a new row inserted above, and you can now type the column names.





Column names come in handy no matter the size of your spreadsheet, but they are especially good if you have many columns and you can't remember what they all mean. In fact, unless you have a really good reason not to, I'd recommend using column names on all your spreadsheets.

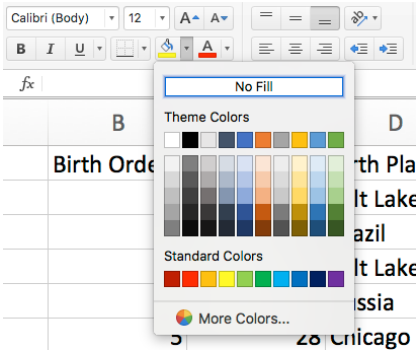
2.2 Auto-Complete

Let's create another column that includes where these people were born. Something Excel does for you is what's called *auto-complete* which is where it tries to help you out and saves you some typing. If you start typing something that has already been seen in the same column, it will fill it in for you, assuming you'll want to type it again. When your spreadsheets get bigger and there are multiple options that match, it'll have many more choices for you to select. You can "approve" auto-complete by hitting tab or enter like normal. If this feature bothers you, you can always turn it off in Excel preferences.

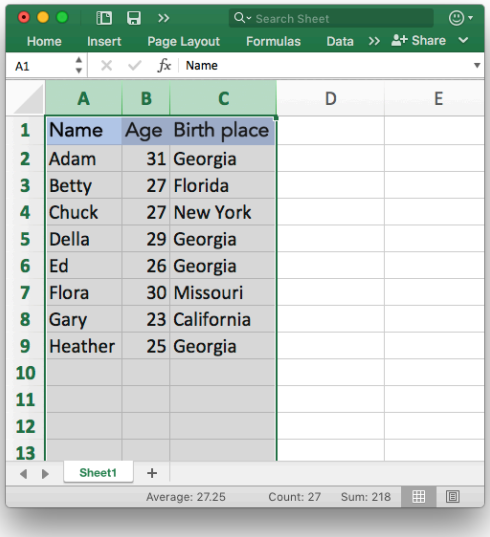
2.3 Formatting

Now that we have some data entered, let's look at the ways you can format it as a whole. In the same toolbar as before, you can see where you can change the typeface, font size (using actual numbers or simply making it bigger or smaller), make things bold, italic, or underlined, add borders to cells, fill the cell with a color, or change the color of the text.

For each of these, there is a dropdown menu that has further options, so if you're not a fan of the super bright yellow fill, you can click the little arrow and change the color to something less fluorescent. Typically, column headers get some slightly fancier formatting, whether it be a bigger size, bold, shaded cells or a different typeface, while the data itself is usually plainer. Go ahead and play around with these formatting options until you get something you like.

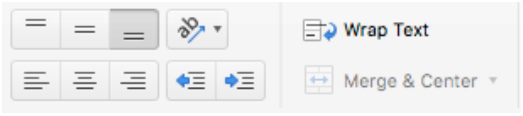


orcid.org/0000-0002-9185-0048

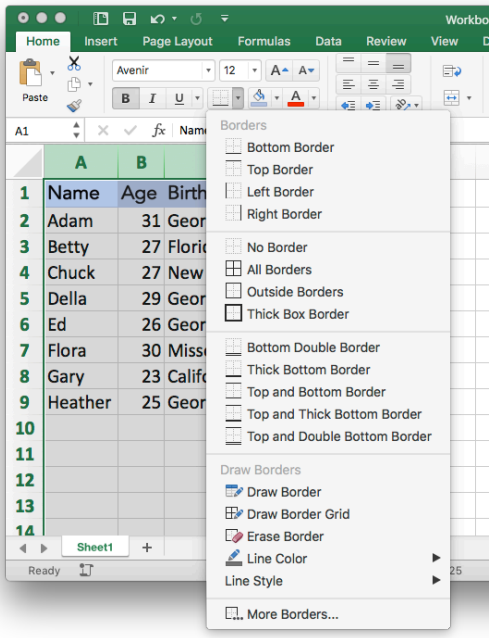


Something you may have noticed if you made things a bigger size, is that now the text is too wide for its column. You can click on the dividing line between the columns (say between B and C), and drag to the right to make the whole column wider[1]. Alternatively, you can double click on that dividing line and it'll make the column just wide enough to see the widest contents of that column, which is handy.

In fact, if you highlight multiple columns—by clicking on the actual letters (A through C rather than the text *in* the columns), you can modify all of them at once. If you want them all the exact same width, you can drag the dividing line between any two columns and they'll all update. If you want them all appropriately sized, you can double click on any dividing line.



You may want to modify how the contents of the cell are aligned within the cell, which can be done with the next portion of the sidebar. Here, you can decide how you want the text aligned vertically (top, middle, bottom) or horizontally (left, center, right). You can even change the direction of the text to something like diagonal or completely vertical, though I rarely used this. If you want to indent within the cell, you can use the buttons to increase or decrease the indent (since the "tab" key is being used for something else now—move to the cell to the right). The "Wrap Text" button makes it so that your text can spill over onto multiple lines if it's too wide for the current cell width. Finally, the "Merge & Center" button is if you want to do some fancy formatting involving merging multuple cells together.



2.4 Borders

You can add some visual structure to your spreadsheet by using borders. In the dropdown menu by the border button, you can see all the various options for kinds of borders you can add. You can add borders to the left, right, top, and bottom, and then

[1] When I take a screenshot, it takes the cursor out of the image, so I can't show you exactly what this looks like.

there are some shortcuts for some other common patterns and styles. I like to put a small border underneath my column headers, so to do this I can highlight the entire first row and add a bottom border. On the very bottom, you can click on "More Borders…" which will open up a new window with lots of other options for the line type and color.

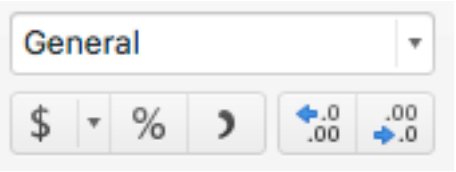
Borders are great if you have a relatively small spreadsheet that will be seen by lots of people. But in this workshop, we're working on learning how to manage your data for a Digital Humanities project, so I wouldn't worry too much about them.

2.5 Number Formats

The last portion of the toolbar that I want to introduce in this first section is the number format section. But this takes a bit of explanation beforehand.

For a given cell, there are two ways of looking at it: there's the way it looks and then there's the underlying contents. I'll call these the "surface" and the "underlying" forms respectively[2]. A lot of the time, what you see is exactly what the underlying contents are. But sometimes you'll want to display the information differently.

Take a number like pi (π = 3.14159…). If you wanted to include this number in your spreadsheet, theoretically the cell would have to be much wider than you'd ever want it to be. You probably just want to round it to the nearest couple decimal points (3.14). With number formats, you can make it so that Excel remembers the full contents of the number, but only displays a portion of it.



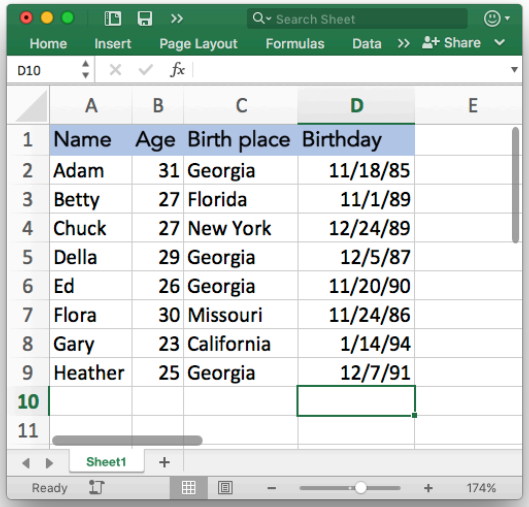
To do that, you'd highlight the cell and hit the button that looks like a comma in the format portion of the toolbar. You can then use the buttons with the little zeros and blue arrows to change how many decimal places you want to display. If you want to display your numbers with a currency symbol, or as a percentage you can do that too. There are a couple reasons for why this is useful. Sometimes you just don't want to look at all the decimal places in your numbers. It can also save you a lot of typing: you don't need to keep putting in the dollar sign or the percent sign for every row.

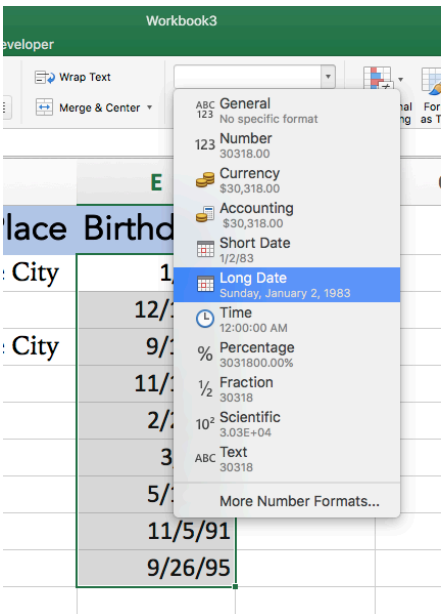
Probably most importantly though, it makes functions work better. If you want to divide "$10" by three, Excel expects numbers when it does math, and the dollar sign throws everything off—that is, if the dollar sign is in the underlying form. But if the cell only contains "10" underlyingly, with the dollar sign showing up superficially, then it knows what to do. Also, after you've divided by three, it's pretty useless to see more than two decimal points after the dollar sign ($3.333333), so you can just display two, though Excel keeps all the others in it's memory. As proof that Excel retains that information, if you multiply that number by 1,000,000, the answer will be $3,333,333.33, whereas if it multiplied it by the surface (rounded) form, the answer would be $3,330,000.00.

[2] Yes, fellow linguists, I'm straight up stealing these terms from generative phonology.

orcid.org/0000-0002-9185-0048

There are many, many more formats available, including the option of creating your own. The only other one I want to bring up today is the date format. To practice, let's add birthdates to the table of friends. In the example, I entered this data with slashes, and that's fine. But what if I decide later on that I want to write out in full? With only eight dates, it wouldn't be unreasonable to do that myself, but if I had hundreds or thousands of rows, this would be a huge pain. Luckily, Excel is smart and knew to interpret a set of three numbers divided by slashes as a date. So if we highlight the cells we want to change, and go to change the format with the dropdown menu and select "Long Date" you can see how they update. Now, under the hood, Excel still knows they're dates, so if you run a function that calculates how long ago something was, it'll still work just fine. We'll get to that later though.





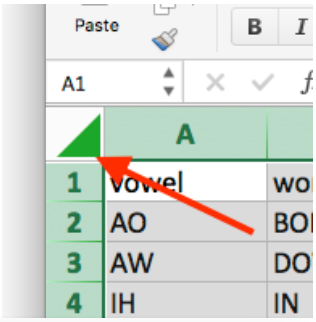
Side note. Sometimes Excel tries too hard and formats things as dates when you don't want it to. If you put anything that looks remotely like a date, it'll interpret it as one. So if you are putting in information about sections from a book or something (like "1-2"), it'll interpret it as "January 2, 2017", which can get annoying. You'll have to go back and tell Excel to interpret it as Text, rather than Date, which is available as the last option in the dropdown menu.

3 Useful Stuff

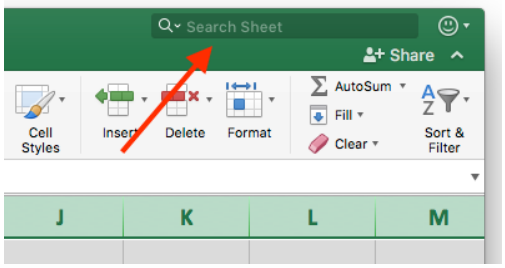
Now that we've finished the basics, let's move on to some useful skills. So far, we've done nothing that will save you time, and you're probably wondering why we even use Excel in the first place (rather than Word or something). Well, from here on out we'll be covering some tricks that'll really help you out in saving time, as well as extracting information from your spreadsheet that would otherwise be super difficult in Word.

I'll switch to a different dataset, something bigger, for illustration. This is a sample of linguistic data from the Linguistic Atlas of the Gulf States,[3] and includes acoustic measurements of one woman's vowel sounds, as well as a bunch of other information. There are over 8000 rows, each with information about what the vowel was, what word it was in, what the previous and following sounds were, and three different acoustic measurements

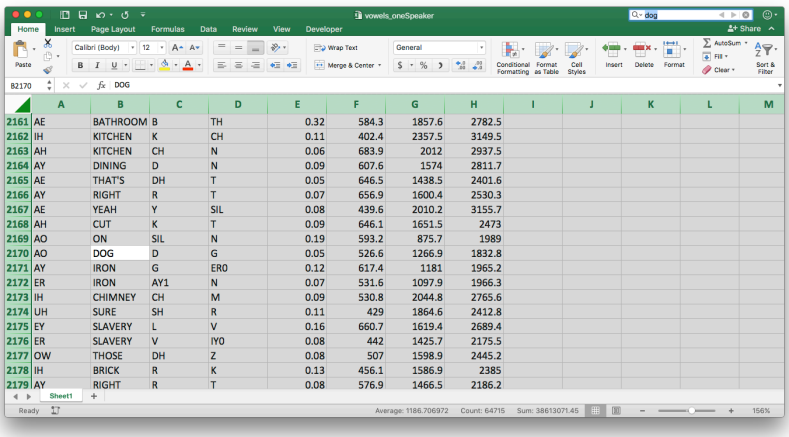
3.1 Search and Replace



The first thing is how to search for things. In a large spreadsheet, you don't want to have to scroll through tons of text to find something specific. What you do first is you can highlight the portion of the spreadsheet you expect to find whatever it is you're looking for. Usually, you have an idea of what column(s) it'll be in, but if it could be anywhere, you can click on the cell in the very top right corner (to the left of "A" and above the "1") to highlight your entire spreadsheet.



Then you go to the search and replace bar at the very top right of Excel (or hit "ctrl+F" for Windows or "command+F" for Macs) and type in what you want to search. Let's say I'm interested in whether this person said the word *dog* in this interview. Sure enough, on row 2107 the word *dog* is there. That's just the first time they said it: if you hit the little rightward-pointing triangle on the search bar, you can see the next match (alternatively, you can use the keyboard shortcut "command+G" for

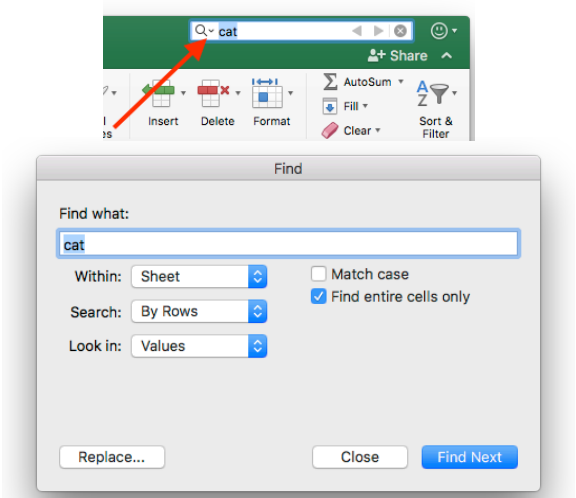


[3] For more information on this and Linguistic Atlas Project in general, go to lap.uga.edu.

Macs and, according to the internet, "shift+F4" for Windows). Now we can see that this person said *dogs* on line 3809 and *dog* on 5104.

Notice that it found the word *dogs* as well as just *dog*. Excel isn't trying to be super smart and knowing that the word *dogs* is the plural form of the word you're looking for. It just happens to be that the letters D + O + G are contained in the word *dogs*.

This is fine, but what it we want to look up the word *cat*. You'll see quickly that the first match is the word *catching*, and then *catch*, and two hits of *education*, and then finally we get to *cats*. Not very helpful for us. What we can do is tell Excel to look cells that contain the search pattern, and nothing else. Click the little dropdown arrow next to the magnifying glass in the search box, and click on the "Advanced Search…" option. That'll open a new window where you can click on the "Find entire cells only" box. Click on that and hit "Find Next" and Excel will tell you there were no matches. I guess this person didn't ever say just the word *cat*. If you change the search pattern to the word *a*, which would be difficult to find using just the regular search function, you can quickly find all the cells.



Sometimes though we want to search for all instances of a word so that we can replace it with something else. For example, in the first column, we have two-letter sequences to identify the 15-ish unique vowel sounds we have in English, since the five vowel letters we have aren't enough.[4] Right now, the sequence "OW" represents the sound in *boat* or *toad*, but what if we want to change it to "OU"? With search and replace, we can do this in seconds.

First, click on column A to highlight the area you want the search and replace to apply to. Then click the little arrow next to the magnifying glass in the search box again and click "Replace…" (Windows shortcut: ctrl+H; Macs: I'm not sure). In the "Find what:" box, type your search pattern (OW) and in the "Replace with:" box type the new letters (OU). You can click on "Find Next" and "Replace" to do them one at a time, if you're not sure you want to replace them all. Or, if you're sure you're right, click "Replace all", and Excel will tell you it replaced 594 instances. Not bad.

Now, in a global search and replace like that, there may be some unintended consequences. In this example, in the very top left cell, the word "vowel" is now "vOUel". Oops. We can always undo it and try again. You can restrict your search pattern so that it's case sensitive (by clicking the "Match case" button in the search and replace window), but I think the better option would have been to click on the "Find entire cells only" button like we did with *cat* above.

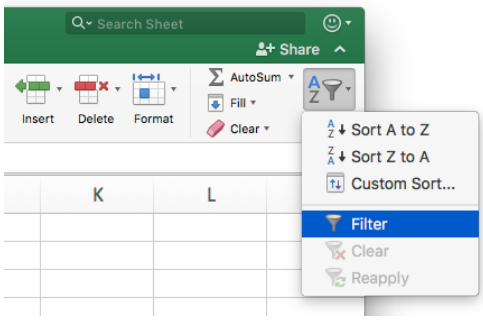
[4] In linguistics, we have special characters for these, but they don't work well in Excel.

Search and replace is a powerful feature and you can save a lot of time with just a couple clicks.

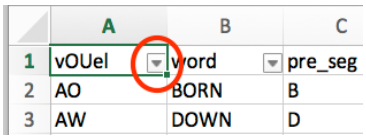
3.2 Sorting and Filtering

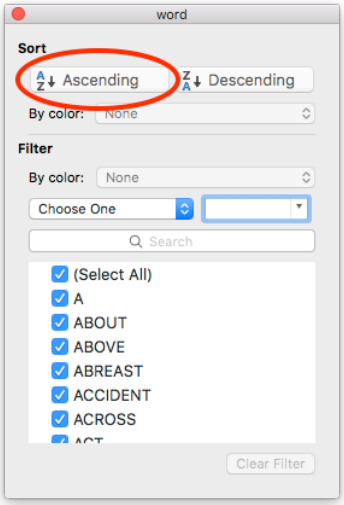
We've seen how to locate certain cells, but it sure made it hard to compare them. If I were interested in comparing the acoustic measurements for every time this person said the word *father*, I'd have a hard time with that just by using the Search function. For this reason, we can sort and filter the spreadsheet.

To get started, on the very right of the toolbar, you'll see a little funnel. Click that and then click the "Filter" button. What this will do is add little boxes with downward-pointing triangles in each cell of your top row. Excel assumes that your top row contains column names. If you click on one of those arrows, you'll now see a lot of sorting and filtering options.



We can sort the table so that the words are in alphabetical order by clicking on the little box in the B column, and then clicking the "Ascending" button. *Voilà*! The table is sorted alphabetically, with all the *a*'s first, all the way down to the two *zero*'s.[5] Now we can scroll down (or use the Find feature) to the instances of *father* and easily compare them since they're all grouped together.





In addition to sorting the window, you can filter it based on whatever criteria you want. For example, if you wanted to see all the instances of *father* and *mother*, you can't use the sort feature to do that. Instead, you can go back to the sort and filter window that pops up when you click on the little arrow at the top of the word column, uncheck the "(Select All)" and then go through and check the words you want. You can use the little search bar to find words quicker.

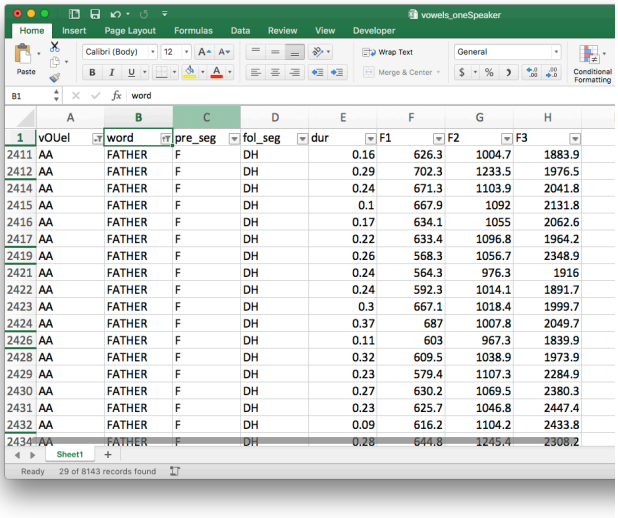
The really useful thing about filtering is that you can do filters on multiple columns at the same time. There are two vowel sounds in *father*, the "ah" and the "er". This spreadsheet contains acoustic measurements, one vowel sound per line, for both vowels, so each instance of *father* is spread over two lines. What if we just want to concentrate on the "ah" sound in *father*? We can do this by first filtering the words so that only *father* and it's various forms

[5] You might see a couple lines of *true* after the *zero*. If a cell contains the word *true*, it will treat it as if it's a true/false data type, so it shows up differently. Admittedly kind of annoying. You can fix this by typing an apostrophe before the word *true*, which tells Excel to treat everything that follows as text rather than anything special.

orcid.org/0000-0002-9185-0048

appear. Close out of that window, and open up the filter for the vowel—or rather, vOUel—column, and uncheck the "ER" box. Perfect. Now we're just looking at just the first vowel in *father*, just 29 of the original 8000+ rows.

So with sorting and filtering, we can quickly find subsets of a much larger spreadsheet, and we can order them the way we want. We can remove any of the filters by undoing the steps we did to get there, or by clicking the "Sort & Filter" button again and unchecking the "Filter" button.



	A	B	C	D	E	F	G	H
1	vOUel	word	pre_seg	fol_seg	dur	F1	F2	F3
2411	AA	FATHER	F	DH	0.16	626.3	1004.7	1883.9
2412	AA	FATHER	F	DH	0.29	702.3	1233.5	1976.5
2414	AA	FATHER	F	DH	0.24	671.3	1103.9	2041.8
2415	AA	FATHER	F	DH	0.1	667.9	1092	2131.8
2416	AA	FATHER	F	DH	0.17	634.1	1055	2062.6
2417	AA	FATHER	F	DH	0.22	633.4	1096.8	1964.2
2419	AA	FATHER	F	DH	0.26	568.3	1056.7	2348.9
2421	AA	FATHER	F	DH	0.24	564.3	976.3	1916
2422	AA	FATHER	F	DH	0.24	592.3	1014.1	1891.7
2423	AA	FATHER	F	DH	0.3	667.1	1018.4	1999.7
2424	AA	FATHER	F	DH	0.37	687	1007.8	2049.7
2426	AA	FATHER	F	DH	0.11	603	967.3	1839.9
2428	AA	FATHER	F	DH	0.32	609.5	1038.9	1973.9
2429	AA	FATHER	F	DH	0.23	579.4	1107.3	2284.9
2430	AA	FATHER	F	DH	0.27	630.2	1069.5	2380.3
2431	AA	FATHER	F	DH	0.23	625.7	1046.8	2447.4
2432	AA	FATHER	F	DH	0.09	616.2	1104.2	2433.8
2434	AA	FATHER	F	DH	0.28	644.8	1245.4	2308.2

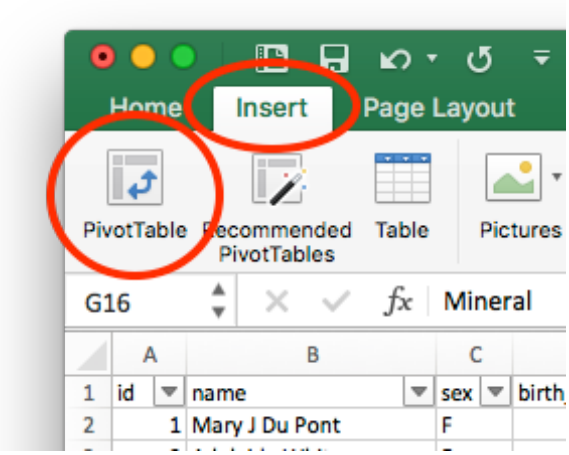
4 Pivot Tables

The next question we might ask ourselves is how many of something do we have in our table? And what is the average acoustic measurement of for each of some sort of category? How many blanks do we have in the data? The answer to all this can be found in a pivot table.

To learn about pivot tables, I'll switch to a larger spreadsheet. This table contains information from the 1930 census for all of the almost 32,000 residents of Cowlitz County in southwest Washington state. I've removed some data to simplify things, and I've added some fabricated data—their height, weight, and favorite color—for illustrative purposes.

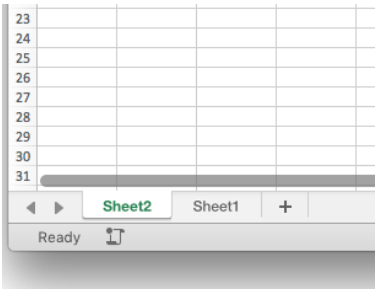
A pivot table is a separate table that can automatically sort, count, total, or average information from another spreadsheet. It has a drag-and-drop interface, allowing you to easily rotate— or pivot—the table in any way you want. Though Excel wasn't the first to implement pivot tables, it did make improvements to it and is probably the most well-known software that incorporates them.

4.1 Pivot Table Basics

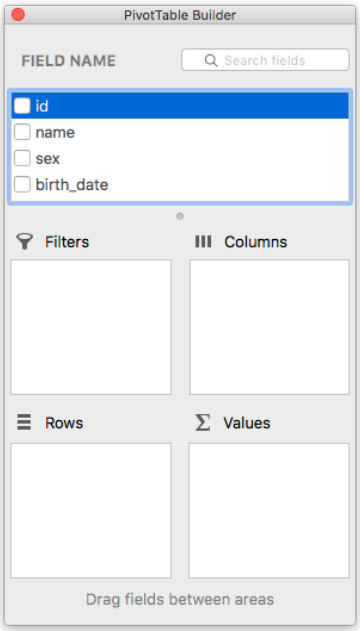


So without further ado, let's make a pivot table. The first step is to highlight the data that you want to summarize, average, count, etc. This is pretty much always going to be your entire table, so go ahead and click the top left corner again and highlight the whole thing. In the "Insert" toolbar, on the far left, click the PivotTable button. You'll see a "Create PivotTable" window pop up, where it'll ask you what data you want to analyze. Since you've highlighted your data already, you should be good. It's always good to put it in a new worksheet, so make sure that button is checked, and click "OK".

The first thing you might notice is that it looks like your data is all gone! Don't worry, Excel just created another spreadsheet for you called "Sheet2". In the bottom left corner, you should see the tab that indicates we're on Sheet2. You can click back and forth between this sheet and Sheet1 (which has all your data). Excel can store multiple spreadsheets within a single file, which can be very useful for things like pivot tables. Some spreadsheets hold data (like our "Sheet1") and others summarize the data (like "Sheet2"). You can have as many of these sheets as you want, so it might be good to right click the name and rename it to something useful.



The next thing you might see is some image of table-like things actually on the spreadsheet. This is just a placeholder until we tell the pivot table what to do.



The important new thing is this "PivotTable Builder" window that has popped up. This is where all the magic happens. There are five windows, and we'll go through each one separately. The first is a list of all the columns in our data with the option to put a check by each one. From this window, we will drag-and-drop columns into one of the other four windows (or simply check the box).

4.2 Adding Rows

The bottom left window is where you'll drag column names if you want the values of that column to be the rows in the pivot table. This makes more sense when we actually try it out. Go ahead and drag-and-drop one of the column names (perhaps "favorite color") to the Rows window and watch what happens.

You see that instantly, a table is created with all the different colors in each row. This is already super useful, because you can what all the unique values are. This also good for quality control, because if there are typos, you'll see them here.

To remove a variable from the pivot table, just drag-and-drop it off the PivotTable Builder window, and, at least on my Mac, it'll *poof!* and disappear. Alternatively you can uncheck the box in the top window.

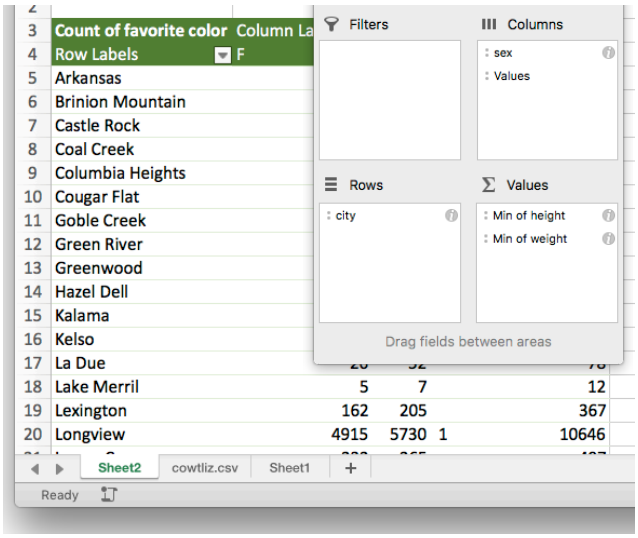
Try adding "birth_state" as the rows instead, and you can see how this is already super useful to see all the states people In Cowlitz County were from: something that would have hard to do otherwise.

4.3 Columns and Values

Back to the PivotTable Builder. (If you accidentally closed out of the window, you can click on the "PivotTable Analyze" tab and press the "Field List" button on the far right.) Let's put "city" in for the rows of our pivot table. But now, drag "sex" to the Column window. You'll now see the structure of a spreadsheet, complete with row and column headers. For the sexes in the census data, we have Female, Male, and Unknown. The pivot table also keeps track of any blank data, and includes that as a separate column as well.

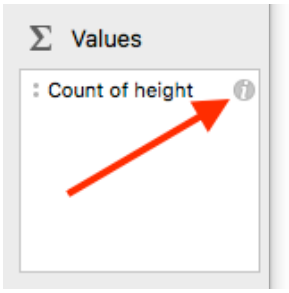
Let's do one more step and drag the "sex" column, from the top window, to the bottom right Values window (so the same "sex" column wo;; be in two of the bottom four panels). Boom. Instantly, you've got a pivot table, complete with tons of useful information. Based on the rows and columns in the pivot table, Excel automatically finds how many people in the original spreadsheet are in each cell of the new pivot table. It also does a grand total column for the rows on

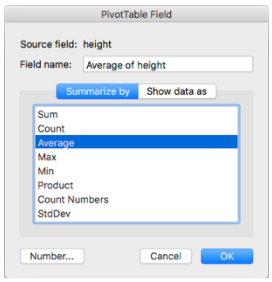
the right and, if you scroll down to the bottom, for the columns as well. This is already a lot of information that you just couldn't feasibly gather from 32,000 rows by yourself. And we've only just scratched the surface!



4.4 Averages

For counting things, it actually doesn't matter what variable you put in the "Values" window. In fact, you can put any variable in there and it'll work just the same. However, for numerical data, like height and weight, there are additional, powerful options.





Go ahead and change the pivot table so that the "height" column is in the Values window. The pivot table shouldn't change for now. But click on the little information button in the Values window. This will take you to a new window where you can change how the pivot table summarizes your data. Right now, it's simply counting the number of people that *have* height. Which is everyone. Instead, you can change it to display the *average* height, which is a lot more useful. Change that and then hit "OK".

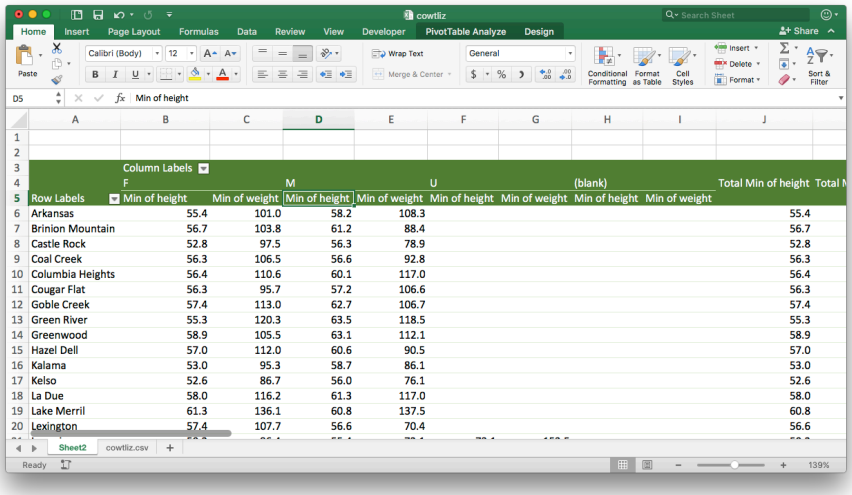
Instantly, the pivot table updates. Now we have the average height for all the residents, divided up by what city they live in and their sex, with the grand total column showing the average overall. Admittedly, this is a little information-overload, because of all the decimal places, so we can just highlight all the numbers and, back on the Home toolbar, round those numbers to fewer decimal places like we did earlier.

Since basically everyone averages out to the same height, this is not too terribly interesting. What we can do instead is see the shortest and tallest man and woman in each city. Go back to the PivotTable Field window and switch it from Average to Max. Looks like there are several very tall people in this county! Change it to Min, and you can see the height of the shortest people in each city.

4.5 Adding More Variables

What if you want to display more information at once? No problem: just drag another column in the Values pane in the PivotTable Builder window. Let's do "weight", but change it so we continue to see the minimum values.

orcid.org/0000-0002-9185-0048



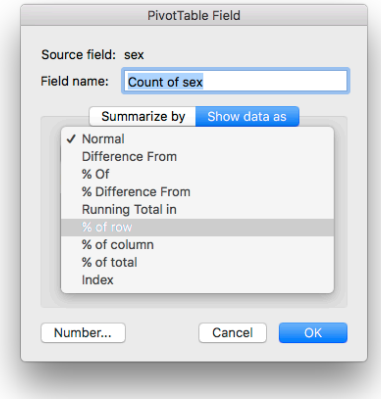
Row Labels	F Min of height	F Min of weight	M Min of height	M Min of weight	U Min of height	U Min of weight	(blank) Min of height	(blank) Min of weight	Total Min of height
Arkansas	55.4	101.0	58.2	108.3					55.4
Brinion Mountain	56.7	103.8	61.2	88.4					56.7
Castle Rock	52.8	97.5	56.3	78.9					52.8
Coal Creek	56.3	106.5	56.6	92.8					56.3
Columbia Heights	56.4	110.6	60.1	117.0					56.4
Cougar Flat	56.3	95.7	57.2	106.6					56.3
Goble Creek	57.4	113.0	62.7	106.7					57.4
Green River	55.3	120.3	63.5	118.5					55.3
Greenwood	58.9	105.5	63.1	112.1					58.9
Hazel Dell	57.0	112.0	60.6	90.5					57.0
Kalama	53.0	95.3	58.7	86.1					53.0
Kelso	52.6	86.7	56.0	76.1					52.6
La Due	58.0	116.2	61.3	117.0					58.0
Lake Merril	61.3	136.1	60.8	137.5					60.8
Lexington	57.4	107.7	56.6	70.4					56.6

Now we see the table get wider because it has more information. The rows have stayed the same, but now there are column and subcolumns. Columns B and C are for the females, D and E are for the males, F and G for the unknowns, and H and I for the blanks. The left column of each sex is the minimum height for that sex in that city. The right column has the minimum weight. It takes a second to understand what you're seeing here, but when it clicks, you'll understand that you're looking at a lot of information all at once. If you want, you can even view two different columns based on the same data, such as the minimum and maximum height. Honestly, I don't keep multiple columns going at once very often because it's a little too hard to read, but the option is there.

4.6 Percentages

Before we get to the last of the four panes in the PivotTable Builder column, I want to show a few more tricks. Let's go back and build the first table we did earlier, with cities for the rows and sexes for the columns. First, we can sort and filter our data right within the pivot table, by going to the little downward triangle next to "Column Labels". Since there are no blanks in the data, we can remove those, and you can choose to remove the "unknown" column as well since there are very few in this data (and are likely data collection errors anyway). If we uncheck those boxes, the table looks a bit cleaner.

Count of sex	Column Labels		
Row Labels	F	M	U
Arkansas	204.0	236.0	
Brinion Mountain	104.0	112.0	
Castle Rock	863.0	920.0	
Coal Creek	157.0	178.0	
Columbia Heights	69.0	73.0	
Cougar Flat	280.0	319.0	
Goble Creek	45.0	63.0	
Green River	11.0	10.0	

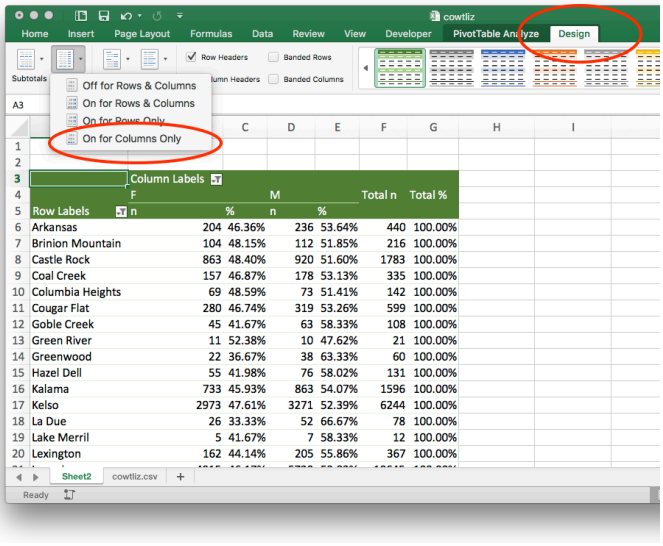


Now, let's say we want to see the percentage of males and females in each city, rather than the raw count. Because right now it's hard to tell whether one city has a disproportionate amount of one or the other. Go back to the PivotTable Builder and click the options for the "sex" values again. Instead of the "Summarize by" option, click the "Show data as" option, and click on "% of

row". When you view your pivot table now, you can see that it has updated as percentages, where each row equals 100%. Now it becomes very clear which cities are evenly split and which are not.

If you want to be really ambitious, try to put the raw count *and* the percentage in the same pivot table. Do this similar to how we did the minimum and maximum height on the same table. You can even rename the columns to something more useful.

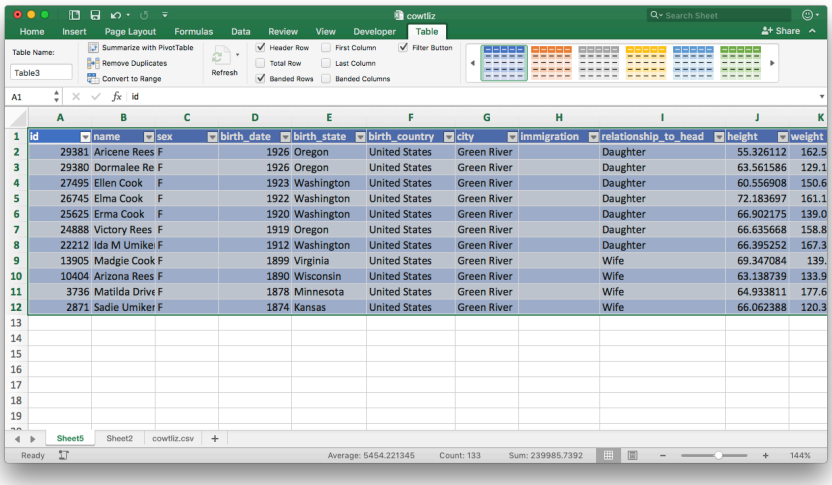
Sometimes, as useful as it can be, you want to remove the "Total" column. When we're displaying percentages per row like this, it's useless to have a column of nothing but 100%. To remove the column, go to the Design toolbar, to the left click on the "Grand Totals" button, and keep them on for columns only. Unfortunately, I don't know of a way to keep just the Grant Total column for the raw count, and turn it off for the percentages, which would be useful. Oh well.



4.7 Adding Filters

Awesome. We now have a table that is super useful if you need to know some general information about this county. But now that we see this information, what if we want to examine the data more closely. What if we wanted to see who the 11 women were in Green River? We could go back to the data and apply two filters like we did before. But with pivot tables, all you need to do is just double-click on the portion of the data you want to examine and it'll show it for you as a separate spreadsheet. Try this by double-clicking the 11 women in Green River.

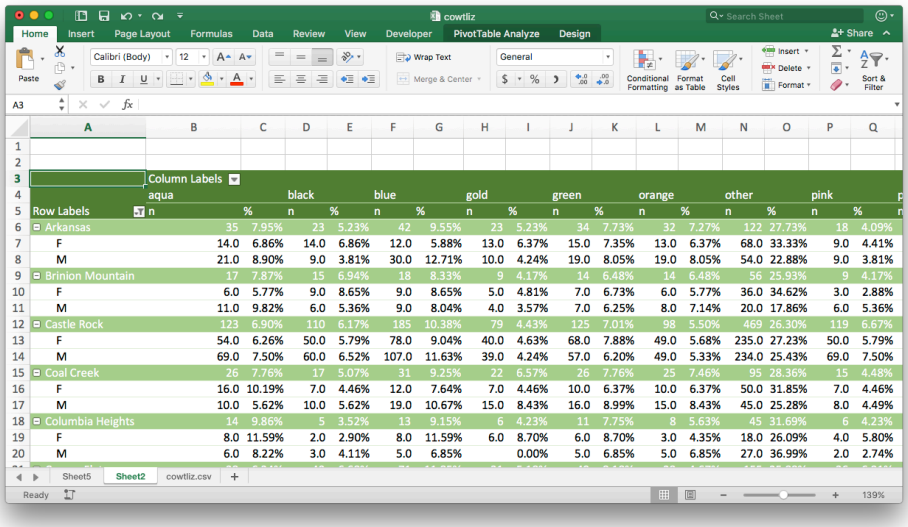
You're instantly taken to a new sheet that has just those 11 women in Green River. Cool! But, the one problem I see with this is that this table isn't linked to the original spreadsheet. So



orcid.org/0000-0002-9185-0048

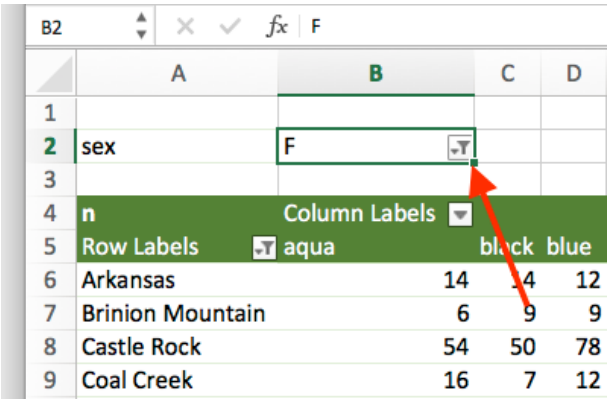
if you want to merely examine the data, this is perfect, but if you want to modify anything, you won't be able to do it through the pivot table and you'll need to apply the filters to the main spreadsheet.

Now let's do multiple row variables just like how we did multiple column variables. Let's say we want to see people's favorite color (the raw count and percentage), split up men and women in each city. We could potentially add a third column variable, but that gets a little cumbersome. Instead, we can add a second variable to the row. To do this, go to the PivotTable Builder window and drag the "sex" variable from the Columns pane to the Rows pane, underneath the "city" variable (order matters here). You'll see the table update. Now, drag the "favorite color" variable to the Columns pane, *above* the values variable. Now we have a pretty big table

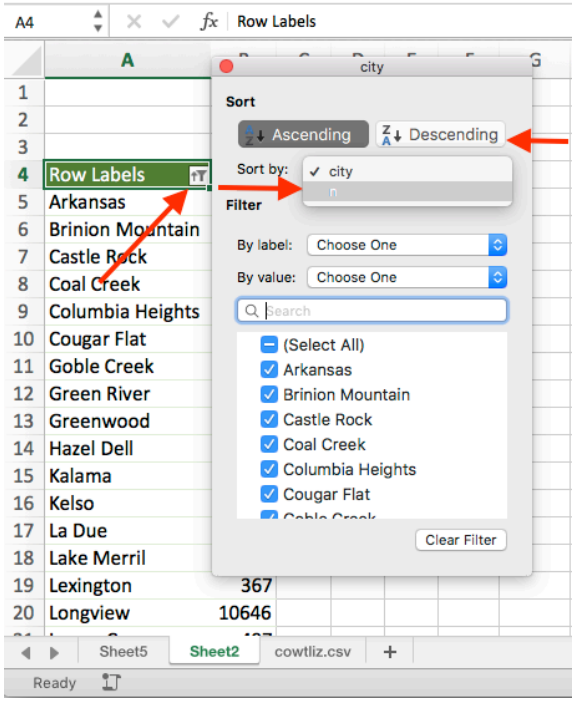


that has now has a lot of information. We can see how many people have a particular favorite color, and what proportion of people do. This is split up by sex, but there's also (the light green) subtotals for each city, in addition to Grand Totals for each color (down at the bottom). If this was something more useful than fabricated favorite color data, this would be incredible to look through.

If this is a little too overwhelming, we can insert a local data filter, using that last top left pane in the PivotTable Builder window. Go ahead and drag "sex" from the Rows pane to the Filters pane. And let's get rid of the percentage columns to calm things down a bit. Now we have a table of how many people indicated that a particular color was their favorite color in each city. But now we have a local data filter. Up in the top left, we can now filter the data based on sex. This



filter works like any other filter that we've seen so far in Excel[6]. You can check the boxes you'd like to include and uncheck others. It may also be useful to add multiple filters to the pivot table, for example if you wanted to see the distribution of favorite colors in each city by the female heads of households. You can do that by simply adding more variables to the Filters pane in the PivotTable Builder window.



4.8 Sorting

Finally, one more trick. Let's say you're not satisfied with the order of how things appear in the pivot table. For example, if you build a pivot table to see the population of each city (by keeping city as the row, no columns, and any variable as the Value), it'll sort them alphabetically. What if we want to order them from biggest to smallest? When you create the pivot table, in the "Row Labels" cell, click on the sort and filter button. Under the "Sort by:" menu, click "n" or whatever the other option is. Then click Descending. Now you'll see that Longview has the most people with 10,646 residents, followed by Kelso, Woodland, Castle Rock, and Kalama.

Pivot tables are a beast. You can sort, count, total, and average a lot of information really quickly. Our main spreadsheet has tens of thousands of rows, but in just a few clicks we were able to extract a lot of useful information. I'll be the first to admit that it takes a lot of trial and error at first to build a good pivot table. But soon it clicks, and you'll feel like a boss.

[6] Note that you can accomplish essentially the same thing by adding a way-cooler-sounding "Slicer" to the table, in the PivotTable Analyze toolbar.

orcid.org/0000-0002-9185-0048

5 Functions

So far, we've done an incredible amount of work in Excel, and we've been able to extract an insane amount of information from our data. And we haven't even typed a single function! In this section, I'll give you a primer to Excel functions and show some functions with practical applications on how to use them. Also…

You've been warned.

But honestly, it's not too bad.

We can already start to play around with functions and return values without getting too crazy by doing some (*gasp!*) basic math. For example, you can type `=2+2` and it'll return 4, `=10-2` will be 8, `=4*6` will be 24, and `=5/2` will be 2.5. You can even use parentheses to embed things: `=2*(4+5)` will be 18.

5.1 Function Basics

If you've never seen functions before, their syntax takes a little bit of getting used to. If you ever used a scientific calculator in a math class, there are a lot of similarities. At their core, a function looks something like this:

```
=functionName(argument1, argument2, etc.)
```
[7]

What we see here are basically four parts. First, they always start with an equals sign ("=")[8]. That's how Excel knows you're starting a function. Then there's the name, and there are tons of those, though I've only ever used maybe a dozen or two ever. These are not case sensitive, so `function` and `FUNCTION` are the same. There is an opening and closing pair of parentheses, and then there's some number of "arguments" inside those parentheses. Assuming all the syntax is correct, the function will "return" some value.

5.2 Arguments

Arguments are relatively straightforward. They're basically the information that a function needs in order to work. I'm a linguist, so let's compare this to language. The function is like a

[7] I'll use a monospace font for the functions themselves.

[8] Unless you're embedding one function into another. But we'll get to that.

verb or some sort of action, and the arguments are like the thing acted upon or the direct object. Let's play up this metaphor.

So for example, for some verbs like *sleep*, *sit*, *stand*, or *arrive*, there's nothing being acted upon. Similarly, there are some functions that don't take any arguments. Typing `=rand()`, will give you a random number (with a lot of decimal places) between 0 and 1. Typing `=today()` will give you today's date. Typing `=now()` will give you today's date and the time right now. Typing `=pi()` will give you 3.1415…

Some verbs, typically violent ones like *hit*, *slap*, and *tackle* require some sort of object. Something needs to be hit, slapped, and tackled. Similarly, there are functions that require exactly one argument. So the function `=sqrt()` requires some number (like 4) so that `=sqrt(4)` will return the number 2. Or `=abs(-2)` will return the value 2.

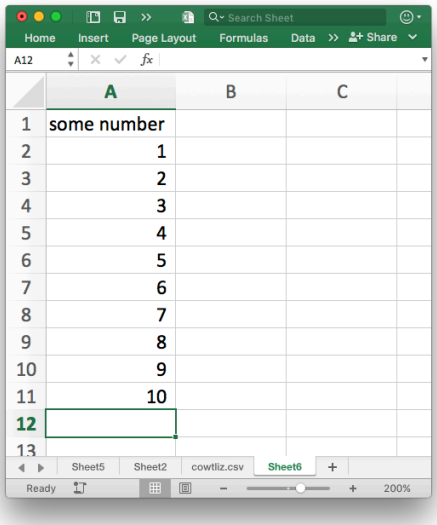
Then there are verbs that can take two arguments: when you *give* or *send*, it's always something$_1$ to somone$_2$. One function that take two arguments is `=round()`, which takes as it's first argument some number that you want to round, and—separated by a comma—another number saying how many decimal places you want to round to. So `=round(0.98765, 3)` will return 0.988. We'll see more of this later, but the concatenate function can string two words (or text of any length) together: `=CONCAT("Digi", "Lab")` will return the word "DigiLab" all as one word. Remember to always use commas to separate arguments.

Some functions require even more arguments, so this is where the verb metaphor breaks down. Also, a lot of functions have optional arguments. We'll see some of these later in this section, but I wanted to give you a heads up.

5.3 Referencing Cells

So, what we've seen so far is good, but you really unlock the power of Excel functions when you start to reference other cells within a function. To illustrate this, let's continue doing basic math. Create a new spreadsheet with the numbers 1–10 down one column.

Now let's create a column called "times 5" where it's everything in the first column multiplied by 5. In the cell B2, start typing a function using the = sign. Then, instead of typing the number 1, just click on the A2 cell. Then go and finish the function. It should look like this `=A2*5`. Now when you hit enter, you'll see that it returns 5. What it's doing is it's looking in the A2 cell and finding what the value is, and using that in the function. Pretty cool. Go ahead and update your table by typing in the appropriate function for the rest of column B.



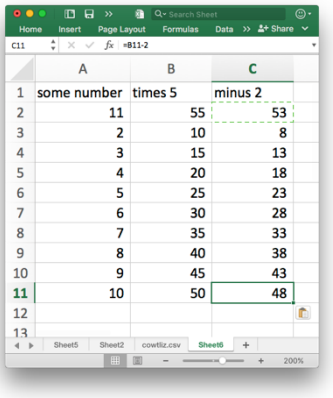
That was kind of a pain though, wasn't it? With 10 cells it's fine, but what if you had a hundred or a thousand or tens of thousands of cells? Luckily, Excel

orcid.org/0000-0002-9185-0048

has made it super easy to copy and paste funtions. In fact, if you literally just copy and paste the function from B2 onto the cells B2–B11, it'll update automatically. How does that work and why doesn't it just paste "5" for all the cells? When you copy and paste a function, it keeps track of the relative position of the source and destination. So if you copied from B2 and pasted it *down one cell* onto B3, the cell that is referenced in the function also gets shifted *down one cell* so that the function in B3 appropriately looks at A3 to do the math. There is a way to override this, called anchoring, and you can see how this is done in section 8.1.

Why reference cells in the first place? Why not just type the numbers in? The reason is because the functions update if the cells they refer to change. Try it out: change the value in A2 from 1 to 11. You should see the 5 in B2 change to 55.

Something else to keep in mind about function is that you can use the return values from a function in one cell in the function of another cell. To illustrate this, create a third column and call it "minus 2". In C2, type the function `=B2-2`. Now, copy and paste that function for the entire column C. See how C2 returns the value (55–2=) 53. It's totally fine to use cells that contain functions as arguments for cells with other functions.



	A	B	C
1	some number	times 5	minus 2
2	11	55	53
3	2	10	8
4	3	15	13
5	4	20	18
6	5	25	23
7	6	30	28
8	7	35	33
9	8	40	38
10	9	45	43
11	10	50	48
12			
13			

In fact, what if we wanted to expand our table so that it's a hundred rows tall? We could sit there and type 11, 12, 13… all the way to 100, but there's got to be a faster way. Can you think of a solution?

Something I do *a lot* is just the formula `+1`. So in the cell A12, write the formula `=A11+1`. It returns the number 11. You can copy and paste the function from A12 to A13, A14, A15… all the way down to A101. Just copy the cell, highlight the whole range, and paste. Boom. You have a sequential list of numbers, each cell referencing the value from the cell above it. Since they're all linked, if you change the value in A11 from 10 to 100, they'll all update. Pretty neat. In section 8.6 you can read why having such a column is super useful.

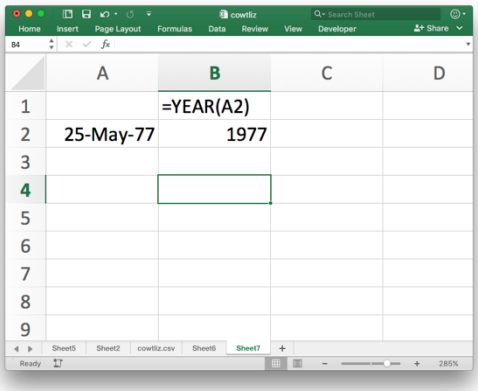
If you see a number on your spreadsheet and you'd like to see what the function is under the hood, if you select the cell, the function will appear in the bar above, as seen in the previous screenshot.

Now that we've covered the basics, I want to dive into some of the more common functions, grouped by what kind of data they need. We'll start with date functions.

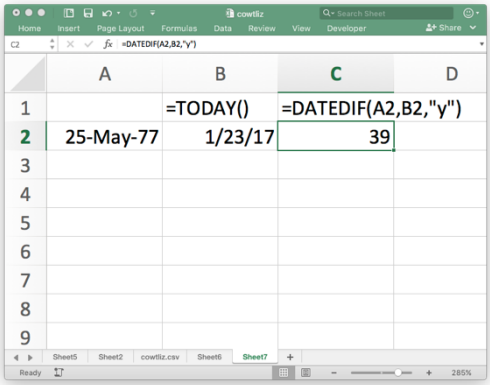
5.4 Date Functions

We've talked about dates in the past. Excel, sometimes unnecessarily so, interprets date-looking things as dates. There's a bunch of functions that are good for dates, which may come in handy for your research. In this and the next couple sections, I'll be briefly explaining a handful of functions, just to show you what is possible in Excel. There are tons and tons more, but this is just to expose you to some of the more common ones.

First, sometimes it's useful to extract just a portion of the date. If you have something like "Wednesday, May 25, 1977", sometimes you'd like to perform a function on just the year. Well, we can extract the year using the `=year()` function, with the date as the argument. You can do similar things by extracting the month, day, or weekday using the `=month()`, `=day()`, or `=weekday()` functions in the same way.



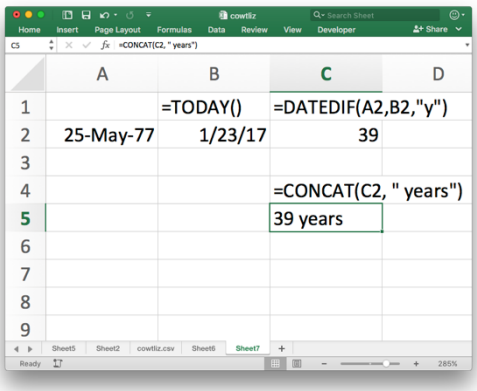
Sometimes you want to know how long ago something was. In cases like these, we can use the `=datedif()` function, which has three arguments: the first date, the second date, and how you want your return value: "y" for years, "m" for months, or "d" for days. This is good for calculating someone or something's age.



5.5 String Functions

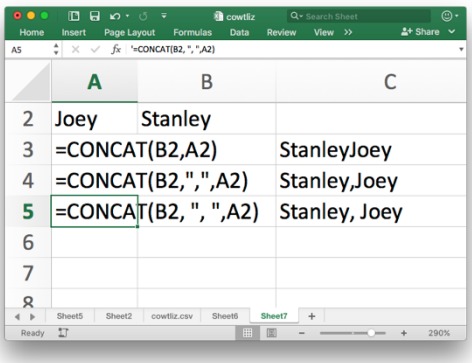
In the computer world, things that use letters and are text-like are called "strings". Let's take a look at some string functions that might come in handy.

The first is one that I showed above, the concatenate function. As I pointed out, `=CONCAT("Digi", "Lab")` will return the word "DigiLab" all as one word. Let's take the previous example, 39 years, and turn it into the string: "39 years". If we use the function `=CONCAT(C2, "years")`, we get "39years" with no space though, so we need to be sure to add the space in there before "years".



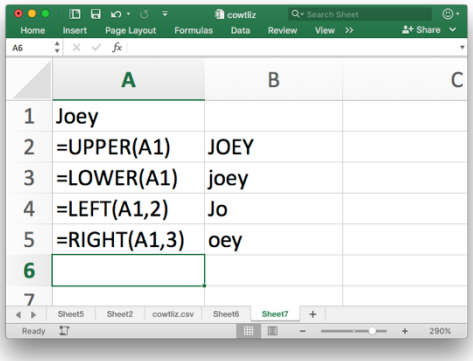
What if we have a table with hundreds of names, where first names are in column A and last names are in column B? What we would like though is a new column, with the template "Last, First". How can we use `=CONCAT()` to fix this problem?

We can start by simply concatinating the last name and the first name: `=CONCAT(B2, A2)`. But that'll give us "LastFirst". We'll need to add a comma. So we can just add it as a third argument, between the two cells: `=CONCAT(B2, ",", A2)`. This will give us "Last,First", which is good, but we



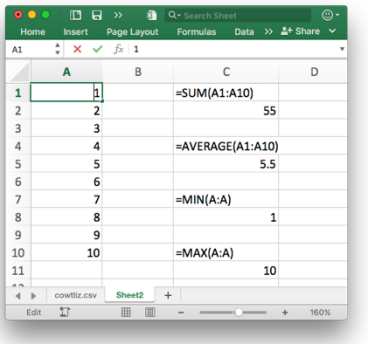
orcid.org/0000-0002-9185-0048

forgot that pesky space again. Finally, we can settle on the last one: `=CONCAT(B2, “, ”, A2)`, which gives the correct output of “Last, First”.



Some other useful functions involve changing the case, which you can simply do with `=upper()` and `=lower()`. You can also extract the left and rightmost number of characters, with a second argument saying how many characters you want to extract, using the `=left()` and `=right()` functions.

5.6 Referencing ranges



So we’ve seen functions that can take information from individual cells, but there are actually cells that can make calculations on and return information about many cells at once. To do this, you simply select a range of cells instead of a just a single cell.

Starting with math (*gasp!*) functions that we’ve seen before, functions like `=sum()`, `=average()`, `=min()`, and `=max()` perform a calculation based on a range of cells. The way to select a range is to simply click and drag to highlight the entire range. Alterantively, if you want to perform a function on an entire column, you can click on the column letter. Note that the syntax for ranges is the top cell, a colon, and the bottom cell. You can even do functions across multiple columns, in which case the syntax would be the top left cell, a colon, and the bottom right cell.

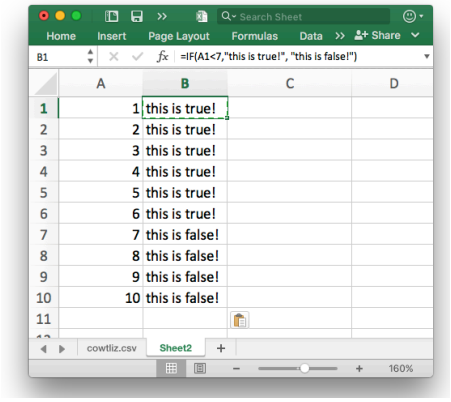
5.7 Conditionals

Another powerful way to draw information from your data is to use conditionals. These are functions that, rather than blindly perform some calculation, will do one thing depending on the contents of the cell.

The most basic conditional is the `=if()` function, which takes three arguments: the true-false statement you’d like to test, the return value if it’s true, and the return value if it’s false. So the function

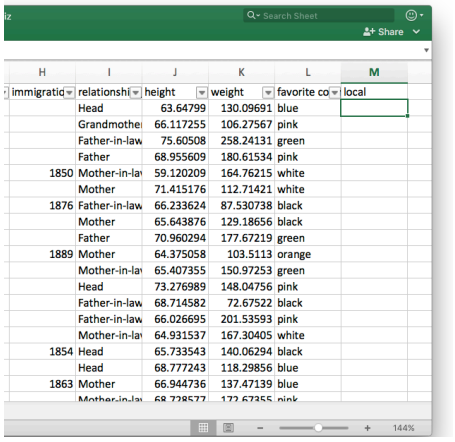
```
=if(5<7, “this is true!”, “this is false”).
```

would check to see if the 5 is indeed less than 7. Since it’s true, the return value for this cell would be the string “this is true!”. Note that if you want to return a string, you have to put it in quotes, while numbers are fine on their own.



Conditionals are boring though unless you actually reference other cells, so let's see how this works. Let's set up a table with two columns, the first being the numbers 1–10, and the second being some sort of conditional function. We can use the exact same function as in the example above, testing whether each cell is less than 7 (but substituting the 5 for the correct cell), and then returning the same strings. If we apply the functions correctly, we should see a table like this.

If we go back to the Cowlitz County census data, we can see how this might be useful. the column "birth_state" has the state of where each person was from. What if we want to just know whether people were from Washington or not? In other words, let's see if we can create a new column that indicates whether people are local.



Go ahead and create a new column, simply by typing the new column name at the right of your spreadsheet. Let's call it "local". In the cell just below it, let's start the function. To identify whether they're from the state of Washington, we need to see if the value in the column "birth_state" is equal to the string "Washington". So that part of the function would look something like `E2="Washington"`. What do we want to do if they are local? Let's return the string "local", and if they're not local, the string "not local". So the entire function would look like this:

```
=IF(E2="Washington", "local", "not local")
```

If we copy and paste this function for all 32,000 rows, we'll have a new, very useful column that we can then use in pivot tables.

There are other conditional functions that may come in handy in your research, like `=isblank()`, and `=countif()` (we'll get to both of these later). You can also use `=and()` if you want multiple conditions to match for something to apply (they have to be from Washington *and* be born after 1900) or the `=or()` function, which will return a match if *any* of some number of conditions apply (like they can be from Washington *or* Oregon). There is lots of help available for all these functions online and I encourage you to look them up and to also see what others you can find.

5.8 Embedding Functions

In this last section on functions, I wanted to show you that you can embed functions into one another just as you can use parentheses to embed numbers in basic math. The only difference

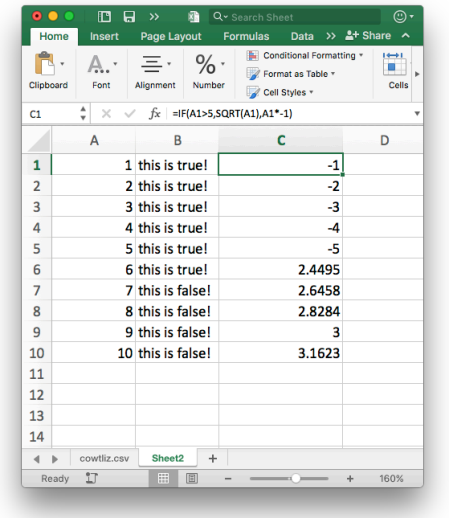
orcid.org/0000-0002-9185-0048

between embedded functions compared to regular ones is that you only need the equals sign before the outermost function.

To show a useless example, let's go back to our table of numbers from 1 to 10 and create a function that will see if a number is greater than 5. If it is, it'll return the square root, otherwise, it'll return the negative. To do this, we'll need to embed functions into the `=if()` function. Here's what you'll need to do:

```
=IF(A1>5, SQRT(A1), A1*-1)
```

This a little hard to interpret, but the function works. Notice that I have to make reference to the cell multiple times, which is totally fine.

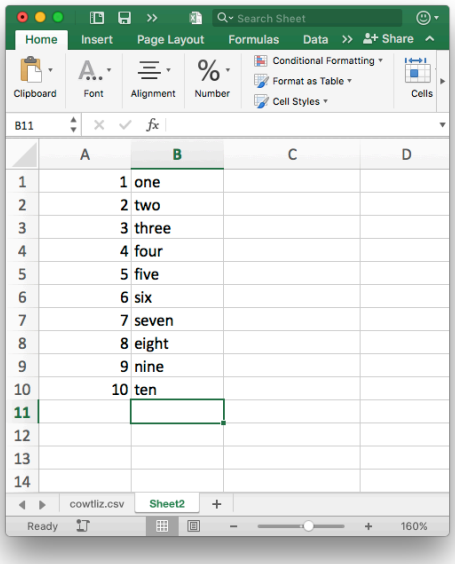


Functions are fantastic and they can save you a lot of time, especially if you need to create new columns based on information from other columns within the same spreadsheet. In the final section, I want to really dive into a particular function, `=lookup()`, which I use all the time in my research.

6 Lookup Tables

In this section we'll cover the topic of lookup tables with the help of the `=lookup()` function[9]. Lookup tables are separate from your main spreadsheet, and are stored in another sheet or potentially in another file entirely. We'll see how to write a function in one spreadsheet that accesses another spreadsheet, we'll see how you can dynamically combine information from multiple spreadsheets into one, and how to incorporate pivot tables into these functions and feed their summarized data back into the main spreadsheet.

6.1 Anatomy of the lookup table



Before we get too crazy, let's just examine what a lookup table might look like. You can think of it as a dictionary. In one column, you have some list of unique values called the *lookup vector*. In this example, I have the numbers 1–10. In a second column, called the *result vector*, you have some other values that are paired with the information in the first column in some way. In this example, I have the words spelled out. The number would be like the dictionary headword and the numbers spelled out would be like the dictionary definition.

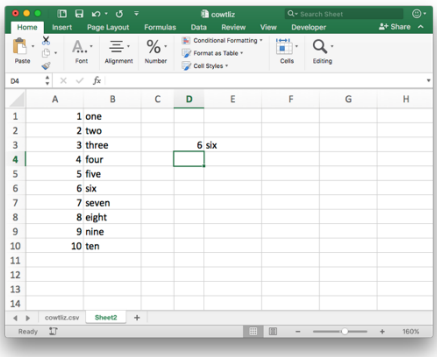
A couple of details that you should be aware of. First, you don't need column headers for lookup tables. Second, this table is stored in a separate sheet within the file, so you'll need to save it as an Excel file (as opposed to a .csv or .txt) if you want it to remain together with your main table. Finally, this behaves like a dictionary in many ways. The lookup vector must be in alphabetical order and contain unique values. However, the result vector does *not* need to be in alphabetical order or be unique in any way. Also, there must be the exact same number of values in the two vectors. Excel will yell at you if there's a violation of any of these dictionary-like properties.

6.2 The =lookup() function

Now let's look at what this function needs. It takes three arguments: the value you want to look up (a single cell), the lookup vector (a range of cells), and the result vector (also a range of cells). To see it in action, we need a value to lookup. For illustrative purposes, I'll type the number "6" in cell D3, Then in E3, I'll add this function:

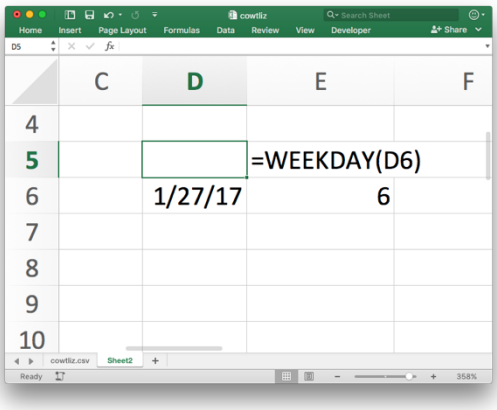
[9] If you've used lookup tables in the past, you might recognize the `=vlookup()` and `hlookup()`. From what I can tell, these are depreciated and are both being replaced by plain `=lookup()`, which can do the task of both.

```
=LOOKUP(D3,A1:A10,B1:B10)
```

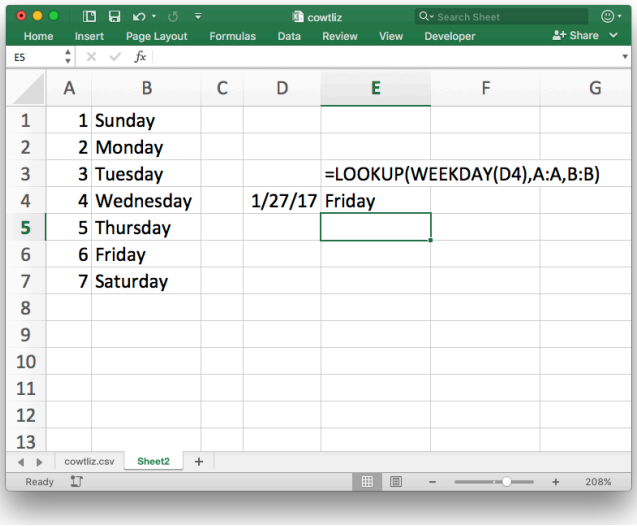


I almost say this out loud when I do the function: "lookup *this* cell in *this* range and return something from *this* list". In this case, lookup D3 in A1:A10, and return something from B1:B10. When I hit enter and let it perform the function, I see that it returns the string "six".

So what's a practical application of this? I already know how to spell the number 6. Remember when we looked at the date functions and I mentioned one called `=weekday()`? This function takes in some date value and returns what day of the week it is. The problem is it returns them as a number, from 1–7 with 1 being Sunday and 7 being Saturday. What I want it to do is return the value "Friday" instead of "6" since I'm not a computer.



Well, we can use a lookup function for this. Let's modify the one we currently have to include the days of the week instead of how to spell the numbers. Now, we need to *embed* the `=weekday()` function inside the `=lookup()` function. In other words, find out what day of the week it is (as a number), and then take that number and send it through the lookup table. We get the desired result.

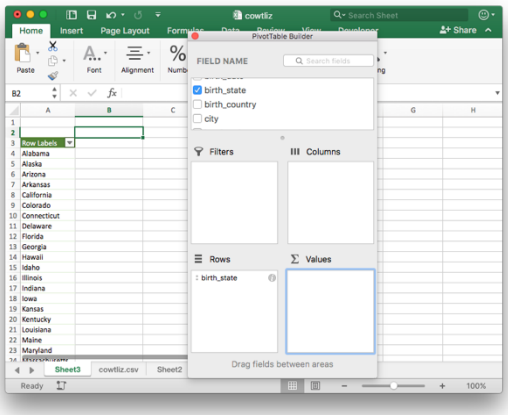


So this works just fine with one cell. Now let's see how we can use it for many more. Unfortunately, Excel apparently thinks that time starts at January 1, 1900, so even if we fabricated birthdays for people, we can't see what day people were born, at least for those 30 and older. But you can see that by copying and pasting the lookup function as a new column, you'd be able to get that information from everyone. Let's try a different application of a lookup table: consolidation of variables.

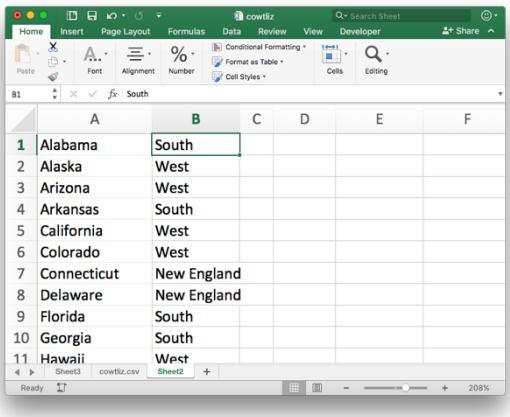
6.3 Grouping Categorical Variables

Sometimes, we have some sort of categorical variable that has just too many unique values. For example, there were people born in all 50 states were living in Cowlitz County in 1930. Sometimes, we don't need to know exactly what state they're from, but maybe just the region,

like New England or the South. With the help of lookup tables, we can add a new column to the main spreadsheet that indicates what region someone was born in, based on the state.

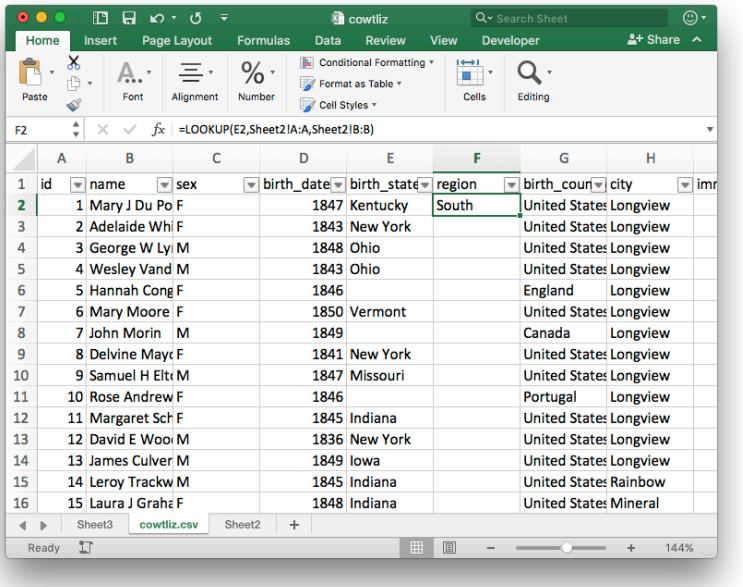


So the goal here is to create a lookup table that has all 50 states in alphabetical order as the lookup vector, and their associated regions as the result vector. Now, before you go off typing all 50 states, let me show you a shortcut. Create a pivot table with nothing but the states as the rows. Boom. Just like that you've got yourself a lookup array. This is even better than just typing them in because you can be sure that you're including everything that is in your table. Copy that column of the pivot table, and paste it onto a new sheet. This is now your lookup table, and is just another one of the many useful applications of the pivot table. Unfortunately, I can't think of a decent shortcut to put in the region, so that one you're going to have to do yourself. Here's my table with completely subjective categorization into five regions: New England, South, North, Midwest, and West.



	A	B
1	Alabama	South
2	Alaska	West
3	Arizona	West
4	Arkansas	South
5	California	West
6	Colorado	West
7	Connecticut	New England
8	Delaware	New England
9	Florida	South
10	Georgia	South
11	Hawaii	West

Now if we go back to the main spreadsheet, let's create a new column to the right of the "birth_state" and call it "region". Now for this column we'll be writing a function that will reference cells on the *other* sheet. It's better to watch me do it live, but essentially when you go to put in the function, you just type what you can, and then click to the lookup table and highlight what you need to, and the function will take care of itself. My keystrokes and mouse clicks were these: click on the first cell in the region column, type "`=lookup(`", click on the top row of the "birth_state" column (Kentucky, E2), type a comma (since multiple arguments in a function are separated by commas), click down to the lookup table (called "Sheet2" in mine), click on column A, type a comma, click on column B, type ")" to finish off the function, hit enter.



	A	B	C	D	E	F	G	H
1	id	name	sex	birth_date	birth_state	region	birth_coun	city
2	1	Mary J Du Po	F	1847	Kentucky	South	United States	Longview
3	2	Adelaide Whi	F	1843	New York		United States	Longview
4	3	George W Ly	M	1848	Ohio		United States	Longview
5	4	Wesley Vand	M	1843	Ohio		United States	Longview
6	5	Hannah Cong	F	1846			England	Longview
7	6	Mary Moore	F	1850	Vermont		United States	Longview
8	7	John Morin	M	1849			Canada	Longview
9	8	Delvine Mayo	F	1841	New York		United States	Longview
10	9	Samuel H Elt	M	1847	Missouri		United States	Longview
11	10	Rose Andrew	F	1846			Portugal	Longview
12	11	Margaret Sch	F	1845	Indiana		United States	Longview
13	12	David E Woo	M	1836	New York		United States	Longview
14	13	James Culver	M	1849	Iowa		United States	Longview
15	14	Leroy Trackw	M	1845	Indiana		United States	Rainbow
16	15	Laura J Graha	F	1848	Indiana		United States	Mineral

orcid.org/0000-0002-9185-0048

I now have a cell that tells me that Kentucky is in the South, according to my lookup table. With some copying and pasting, I can paste this function to all 32,000 rows and, like magic, I have a new column with region of birth. You can now use this column in pivot tables and whatever else you need. Five regions is a lot more manageable than 50 state names.

6.4 Handling Empty Cells

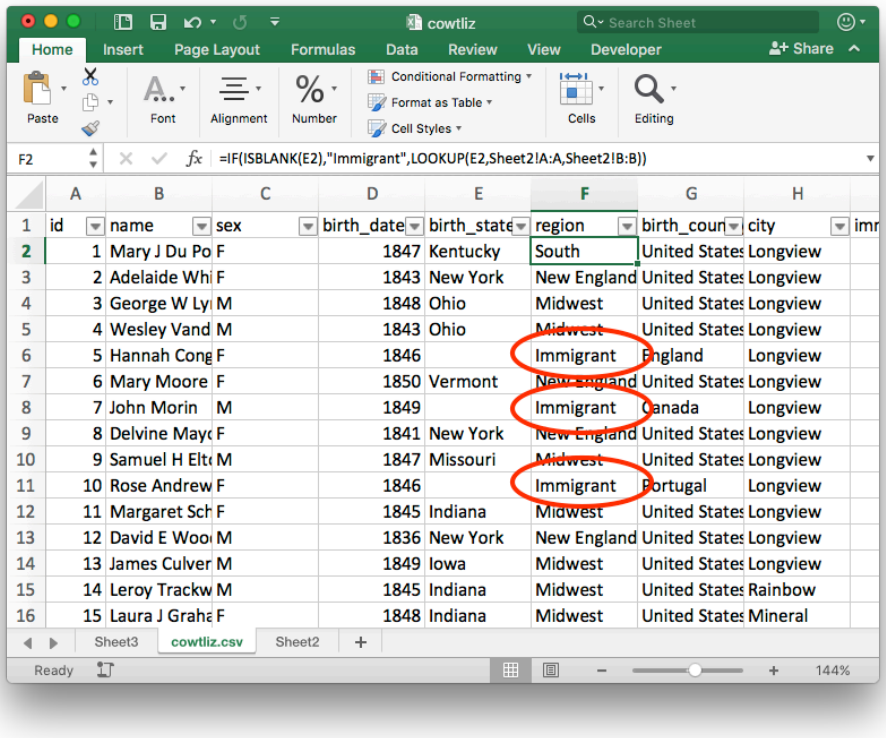
Now, if you scrolled back to the top of our new and improved table, you'll see some errors in the "region" column. Namely, if there is no state listed in the "birth_state" column, meaning the person was born in a different country, it returns an annoying "#N/A". We could go through and delete these, but there's a better way to fix that.

Let's make use of the `=isblank()` function I mentioned in passing in section 5.7. This takes a single argument, a cell reference, and returns "TRUE" if it's blank, and "FALSE" if it's not. What we want to do is modify the function in the "region" column so that if there is no state listed, it will return as "Immigrant" instead of a state name or an error message.

So, what we'll actually do is embed the lookup table into an `=if()` function. In prose, what we want to tell Excel to do is this: look at this cell; if it's blank, return the string "Immigrant"; otherwise, lookup that same cell in the lookup table. The full function looks like this:

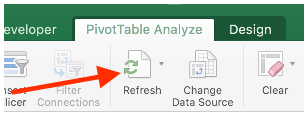
```
=IF(ISBLANK(E2), "Immgrant", LOOKUP(E2, Sheet2!A:A, Sheet2!B:B))
```

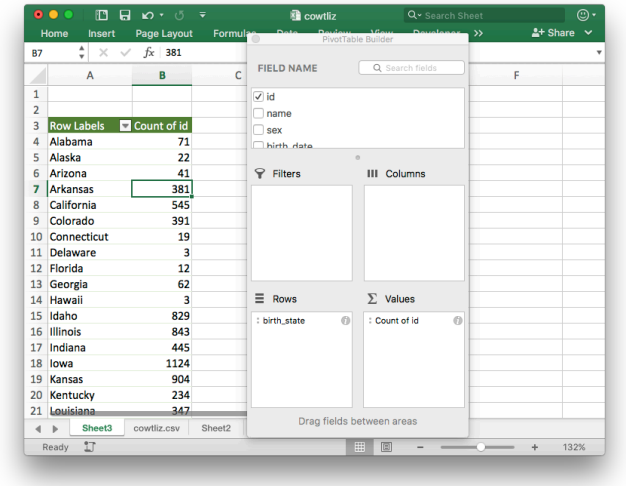
Again, this is long and a little hard to parse, but if look at it one piece at a time, you'll find that it makes sense. Copy this updated function to all the cells in the "region" column, and you'll have yourself an error free column.



6.5 Pivot tables again??

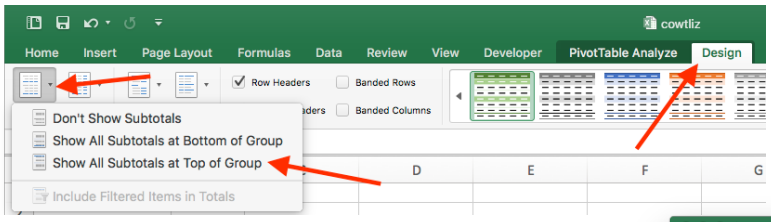
Let's see how useful this is by setting up a pivot table with this new information. Luckily we already have a pivot table set up: the one we used to copy and paste the state names. But if you go to add region to the pivot table, you'll see it's not there. The reason is that the pivot table, unlike functions, are static. This pivot table is based on an older version of the main spreadsheet. No worries though. In the PivotTable Analyze toolbar, we can hit Refresh, and everything will update.



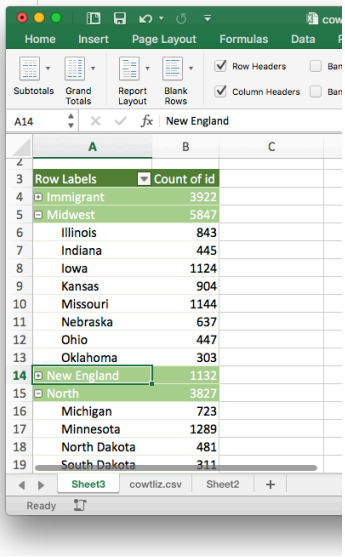


Now, let's see how many people were born in each state. We can add a column like "name" or "id" to the Values pane, and see the count that way. This is good and all, but we can group these states by region now that we have this fancy new column.

In the PivotTable Builder window, add the "region" column to the Rows pane, above the "birth_state" column. You'll see your table instantly updates exactly how we want. Note that we can collapse or expand the regions as well. If it's not done already, we can see how many people were born in each region by going to the Design toolbar, clicking Subtotals on the far left, and clicking "Show all subtotals at the top of group" option.



Now we have a pivot table that shows us a lot of useful information that was not available to us before. This took a bit of work: first creating a bare-bones pivot table with state names, creating a lookup table, then putting in the `=lookup()` function, account for blanks in the data, updating the pivot table, and then organizing the pivot table how we want. A lot of work, likely involving new skills. But dang useful if you ask me.

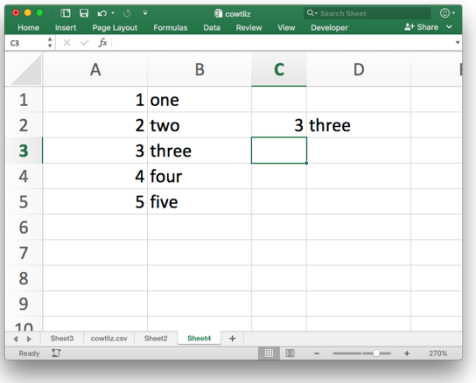


6.6 Quality checking

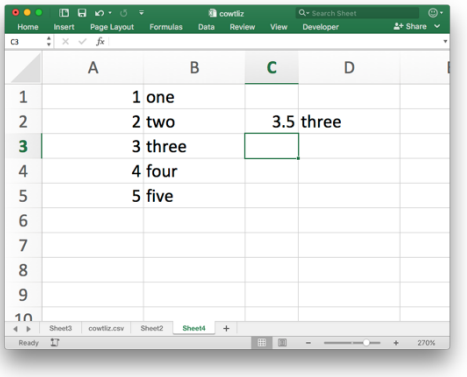
There is one major problem that the `=lookup()` function has: if it encounters some value that isn't in the lookup array, it looks at the closest match and goes with that instead. To illustrate, let's make that same basic lookup table we had from before, where

orcid.org/0000-0002-9185-0048

we lookup how to spell numbers. Right now the cell C3 is the reference cell to the lookup function in D3. It works great if there's a number between 1 and 5 in there. But what happens if you put something else, like 1.5? In this screenshot, you can see that it still returns a value, instead of an error like we'd expect.



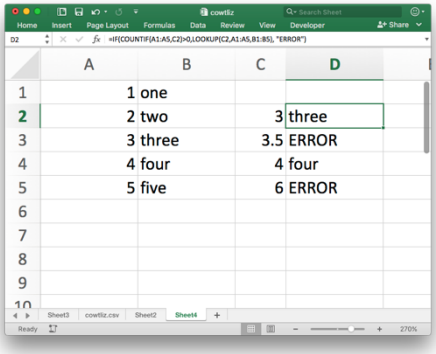
The reason for this is related to why the values are in alphabetical/numerical order. When Excel has the value it needs to lookup (3.5 in this case), it goes through the lookup vector and finds where the value would be (between 3 and 4). At this point, for some reason, instead of throwing an error message at you, it guesses that 3 is close enough and goes with it.



How do we fix this? We can use the `=countif()` function, another one I alluded to earlier in section 5.7. This function takes two arguments: a vector of values, and a single cell reference to look up. What it does is it counts the number of times that particular value is contained in that vector. If you wanted to count how many native Washingtonians there were in the Cowlitz census data, you could do it that way. For our purposes now, if it's not there at all, it'll return a value of 0; otherwise it'll be 1 or some number greater than zero.

Well, let's use this to our advantage. Before doing the lookup, let's first check if the value is even in the lookup table. If it is, great, carry on with the lookup; if not, let's say return the string "ERROR". We can do this with the following function:

```
=IF(COUNTIF(A1:A5,C2)>0, LOOKUP(C2, A1:A5, B1:B5), "ERROR")
```



Now if we apply this to our simply lookup table, we can see that it works great.

Now if want to be really ambitious, we can combine both the quality checking we just did with the check for blanks that we did earlier into one mega function. But it turns out that we actually don't need to because the `=countif()` returns the same error message if the reference cell is a blank. Pretty cool, huh?

7 Visualizations

To be perfectly honest with you, I never use Excel for visualizing my data. I find it clunky to use, not very pretty, and extremely limiting. I bring this up not to deter you from using Excel for visualizations, but rather to explain why I don't know too much about them. In this handout, I'll go over the basics on how to get some quick and dirty visualizations started, for the purpose of helping you see what is going on in your data. I would not recommend using these in presentations or publications, but that is just my opinion. We'll continue working with the Cowlitz County census data for this section.

7.1 The right visualization

Take a step back and think about why we're trying to visualize our data. What's the goal? If you want to include some sexy chart or graph for the sake of breaking up text, that is not a good reason. Some visualizations are good because they tell a story. Others are just meant just so we can process lots and lots of data all at once, without having to scan through thousands of rows in a spreadsheet.

Something that is important to keep in mind is that not all visualizations are appropriate for all types of data, and not all data lends itself well to all visualizations. For example, if you want to make a pie chart and have 500 different wedges, it gets overwhelming for the viewer to process that much information, and you fail at your goal of making something easier to comprehend. So it's important to know what kinds of data are required for particular kinds of visualizations.

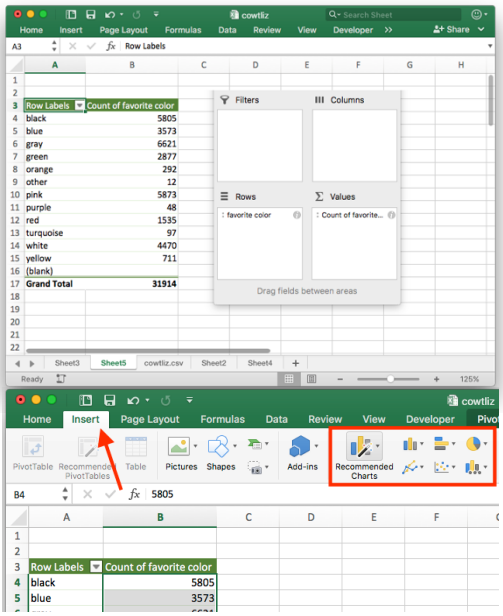
Bar charts, for example, usually take some category (favorite color) and some number, usually a count. So a bar chart would be perfect to compare how many people like certain colors. A scatterplot would be great to compare people's heights and weights because you need two number-like things to make a scatterplot. Plotting things on a map is essentially making a scatterplot with latitude and longitude as the numerical variables and then overlaying a map. Just as it's pointless to try to include a map in your paper if you don't have coordinates or geographic information, it's pointless to make any other sort of visualization if you don't have the right kind of data.

7.2 Oh geez, not pivot tables again!

To get started though, we need to prepare our data for visualizations. Unfortunately, Excel can't really work with your full spreadsheet and make a meaningful visualization. In other words, you can't just highlight a column in your spreadsheet and hit the "make a cool graph" button. You'll need to summarize your data first. What do I mean by this?

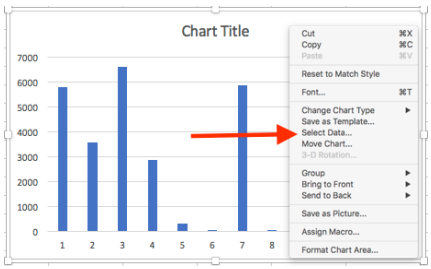
Let's say you want to make a simple chart showing people's favorite colors. Right now we have 32,000 lines of favorite colors, but they're not summarized in anyway, so Excel doesn't

know what to do with them. We need to create a table that just has the 10 or 15 color names in one column and how many times people chose that as their favorite color in another column. Yes, we need to make a pivot table.



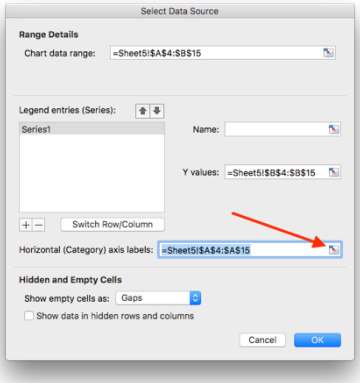
Let's go ahead and do this. Create a new pivot table, which will show up on yet another sheet of your Excel file. Put the favorite color in one column, and the count of favorite color in another column. We've managed to collapse 32,000 rows into just a dozen. Not bad. Excel can work with this now.

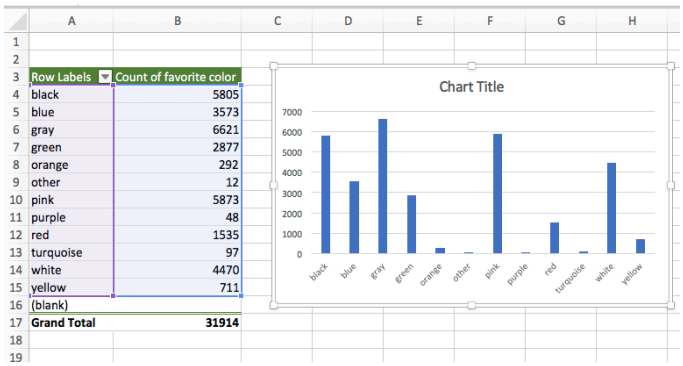
Highlight the twelve numbers that contain the counts of favorite colors, from 5805 in row 4 to 711 in row 15. In the Insert toolbar, there are several options for visualizations. You can choose to make a pie chart, bar graph, line chart, scatterplot, etc. But honestly, I just click on the Recommended Charts button and look through the options Excel guesses. It usually does a pretty good job. The vertical bar chart looks good, so let's go with that.



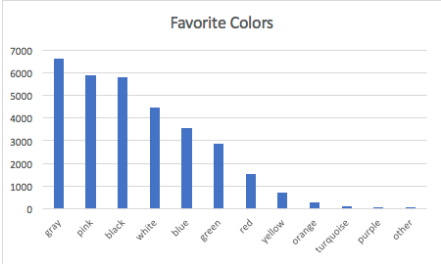
Immediately, we get a bare bones bar chart. The data is displayed correctly, in that the heights of the bars to correspond to the values you've highlighted, but there's nothing else there. Most importantly, there's no way to tell which bar represents which color.

To add labels across the bottom (the *x*-axis), right click the horizontal access and click on "Select Data…". This will take you to a Select Data Source window. In the Horizontal (Category) axis labels area, click the little grid icon, and select the data you want to act as the labels. In this case, it's the first column of our spreadsheet. Click "OK" when you're done, and you should see the chart update, now with the data in the pivot table selected in blue, and the labels in purple.



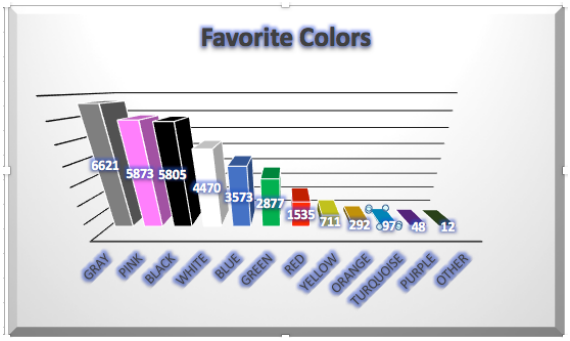


There are a couple of things we can do to make this chart better. First off, we can add a title by double-clicking on the existing title and just typing a new name. It might also be good to sort the data by frequency. We'll have to do this sorting in the pivot table, but we've seen how to do that: click the little sort and filter button at the top of the pivot table, and sort descending by count of favorite color. Now we have a half decent graph.



7.3 Customizing

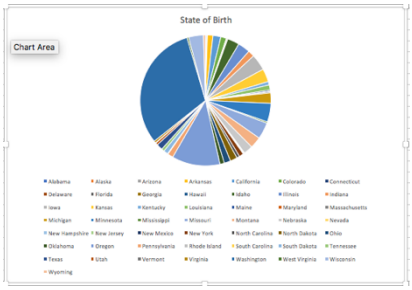
What you see in the graphs are just the default settings. It's possible to change whatever you want about the graph. Colors, sizes, and all sorts of other things. I won't cover them here because there is just too much to do, and it's largely up to you to decide what looks the best.



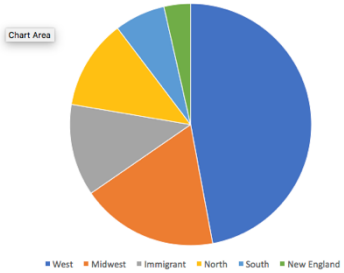
A word of caution. It's possible to make super fancy 3D graphs with shadows and glowing fonts and all the bells and whistles. There is absolutely no need for any of that. Let the data speak for itself. Don't let the format of the visualization eclipse the information it's trying to portray. I usually keep my graphs as simple as possible, removing any extraneous lines, angles, or text where possible to allow for maximal clarity.

With that said, the places to look for customizable options are in the Chart Design toolbar (mostly for general layout options), particularly on the far left under the Add Chart Element button, and also in the Format toolbar for more fine-tuned details. You can always right click on any portion of the chart (the axes, the bars, etc.) and see what options show up there.

7.4 Pie charts

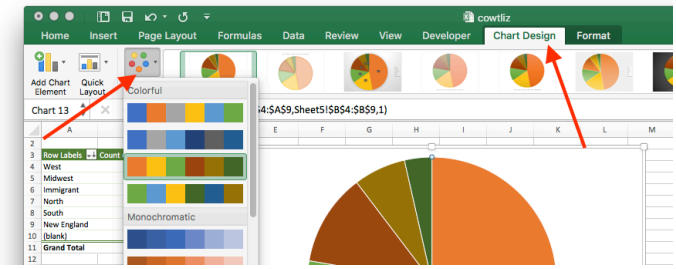


We'll start with pie charts. This is good for counting some category that doesn't have too many unique values. So a pie chart of what state people were born in would be useless since there would be too many slices and too many colors. However, if we do a pie chart on what *region* people were born in, it looks a lot more interesting.



If you're not satisfied with the colors, you can change them. You can click on the whole chart, and in the Chart Design toolbar, click on Change Colors, and pick a theme you like better. If those are still no good, you can click on individual wedges and change the colors manually that way. (You can do this same thing with bar charts too.)

orcid.org/0000-0002-9185-0048

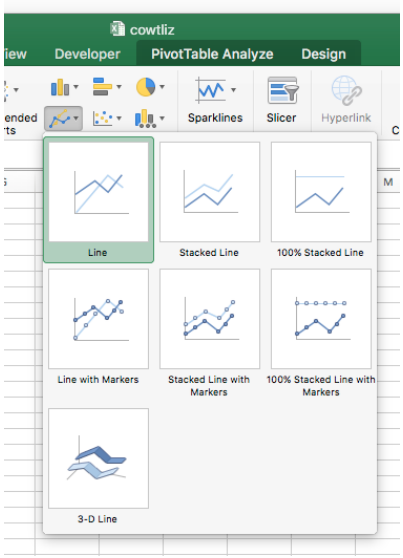


Personally, I'm not a huge fan of pie charts. I've heard that humans aren't as good as judging angles as well as we like to think we are, so the information isn't being conveyed as clear as possible. Whatever you can do in a pie chart, you can also do in a bar chart, and differences in height are easier to see than differences in angles. And I also feel like it comes off as amateurish and like something you'd see in an elementary school science fair. But that's just me.

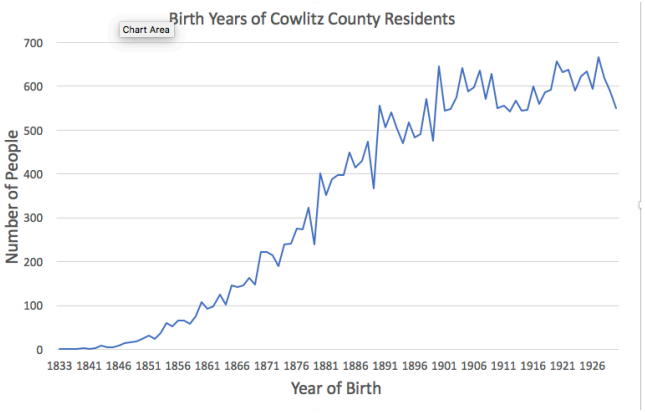
7.5 Line Charts

So pie charts are for information with only a few categories, and bar charts can handle more. But these are for categories that have no meaningful order. Putting red before blue in a bar chart is no more intuitive than blue before red. What if there's some variable that has a meaningful order, like people's year of birth? One solution would make a line chart. Similar to a bar plot, we have some category or label along the bottom, with how far up the chart representing more instances of that value.

To make a line chart, we'll create our pivot table, with the "birth_date" column as the row, and the count of "birth_date" as the value. We want them in the default order (1833–1930) so no need to change the order. Just like the other charts, we'll highlight the data we want to display (the counts) and choose the line chart in the Insert toolbar. You'll probably need to make the chart bigger and change the axis labels like we did with the bar plot.



For this one, it might be helpful to add axis labels, just so people know what the numbers mean. To do that, click on the chart, go to the Chart Design toolbar, and under Add Chart Element, click on axes titles and pick either horizontal or vertical (and then repeat with the other one). You'll then see some default titles and you can then change them to something more meaningful. If you right click on the years and go to Format Axis…, you can change it so the years display horizontal instead of vertical. I haven't found a good way to display certain years (1840, 1860, etc.) though, which is kind of annoying.



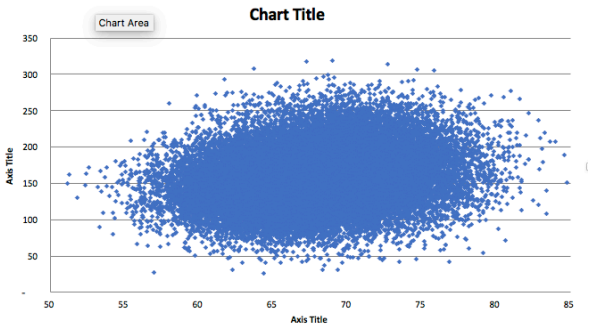
Bar charts are pretty nice, but again, you need some sort of variable that has some

meaningful order. Doing this same thing for favorite colors wouldn't be as meaningful or intuitive as a bar chart.

7.6 Scatter Plot

The last chart I'd like to cover is the scatter plot. This is perfect for two variables that are numbers. In this case we'll take the height and weight. Unlike the other charts, which required summarizing the data, scatterplots work with the main spreadsheet, so we can work from there.

To make a scatterplot, simply highlight the height and weight columns, and pick a scatterplot in the chart selection. This plot is good, but I would like to color code them based on the sex of the person. Unfortunately, this is a nightmare in Excel and there really is no easy way of doing it, which is pretty lame.



7.7 Other Visualizations

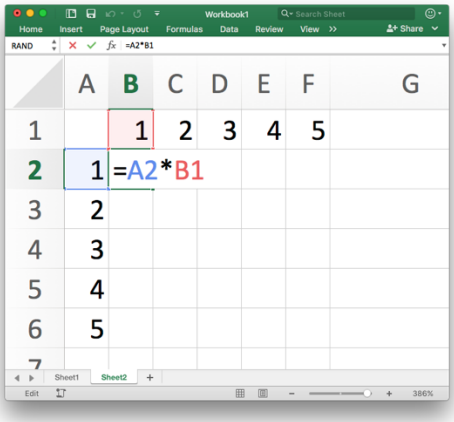
There are a bunch of other possibilities with visualizations in Excel. You can do fancier things like a surface or radar charts, slightly different versions of each of the ones we've done so far (stacked bar plots, secondary pie plots to show the smaller slices, bubble plots, etc), as well as overlaying multiple plots on top of each other. I'll leave those up to you, but they admittedly are pretty cool, but kinda flashy, so still not publication quality in my opinion.

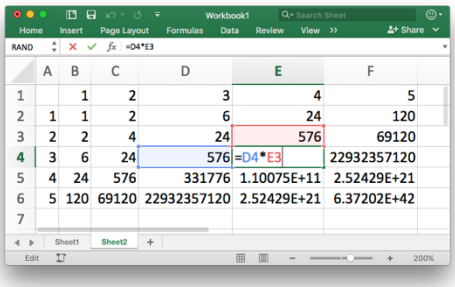
orcid.org/0000-0002-9185-0048

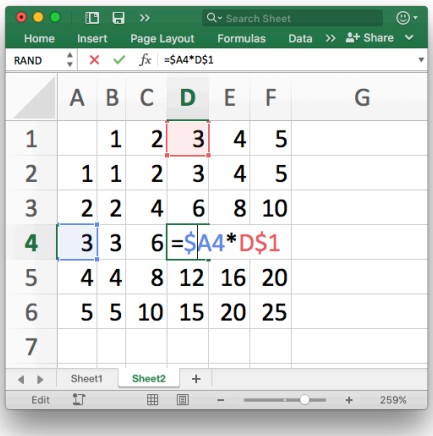
8 Bonus Features

In this chapter, we'll look at random little tips and tricks I've picked up along the way that didn't quite fit in well in previous sections. Some are more useful than others, but I do use these regularly.

8.1 Anchoring







Remember how when you copy and paste functions from one cell to another the references automatically change as well? What if you wanted to *not* do that? For the sake of demonstration, let's say you wanted to make a multiplication table. We've got numbers down the first column and across the first row. What would be the quickest way to fill in the rest of the table?

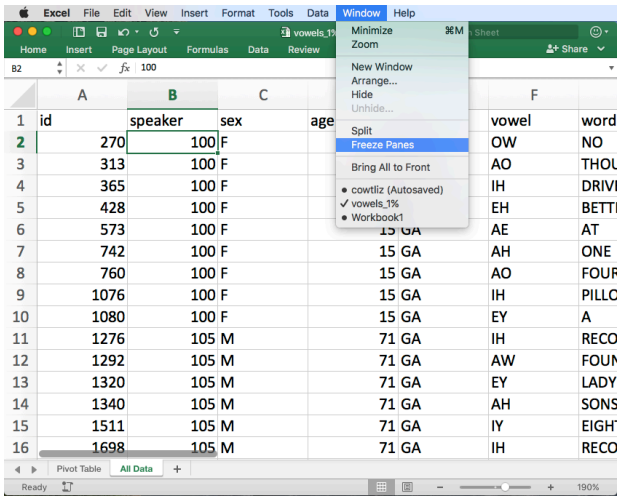
First, we start off with the function in B2: multiply the value in A2 by the value in B1. But if we copy and paste that function to all the other cells, that's definitely not what we wanted, since the E4 would be multiplying the cells D4 by E3 instead of A4 and E1.

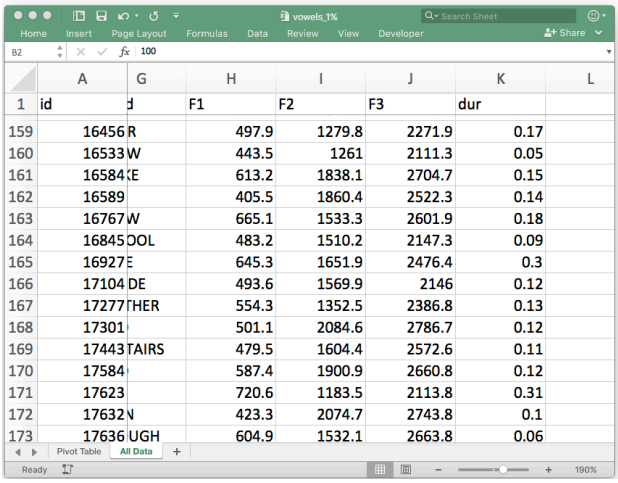
The solution to this is to anchor the cell references. The way to do this is to put a dollar sign before the column letter if you want to anchor the column, the row number if you want to anchor the row, or both.

Going back to the first function, `=A2*B1`. If we copy and paste this to other column on the same row, we still want the column A in the function, so we want to anchor column A. So the function should be `=$A2*B1`. But then when we copy and paste it onto other rows, we still want it to refer to the first row. So, we can update the function to `=$A2*B$1`, which will anchor that row. Now, if we copy and paste this to the whole multiplication table, you'll see that it works out perfectly.

Anchoring is super useful if you want to refer to the column or row names within the table functions. You can also do this for ranges too. It's a useful too to have.

8.2 Freezing and Splitting





Sometimes you get a table so big that you can't see all the rows or columns at once. If you only have a few columns or if all the columns are different enough from each other, then there's probably no confusion as to what column you're looking at when you're a couple hundred rows down the table. But if you have a lot of similar-looking columns, like those acoustic measurements in the vowel data, sometimes it can get easy to lose track of the order columns are in. For this reason, you can freeze certain rows/columns or split the window. These are two slightly different ways of allowing you to see two discontinuous portions of the table at once. I'll use the vowel measurements spreadsheet as an example.

We'll start with freezing. This spreadsheet is really long, and the last couple columns are all a bunch of numbers. Let's freeze the top row so I can always see the column names and also the first row so I can always see the ID numbers. To do this, put your curse *under* the row you want to freeze and *to the right of* the column you want to freeze. Up in the main Excel menu bar, there's a Window menu. Click on Freeze Panes. What this does is it locks that row and column. Now if you scroll down and across, you'll still see column A and row 1.

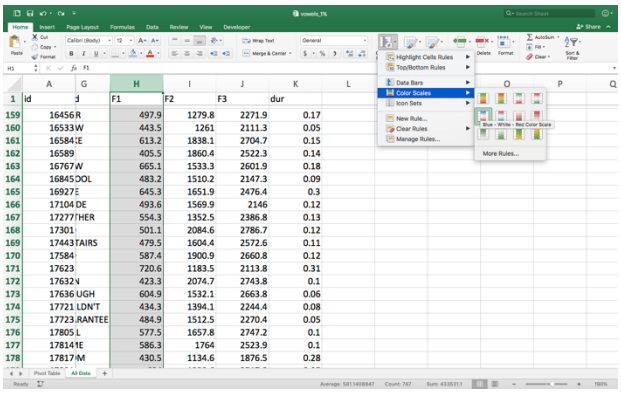
Splitting the table allows for viewing independent portions of the table at once, so is more flexible than freezing. You can do this the exact same way as freezing, except you click the "Split" option instead.

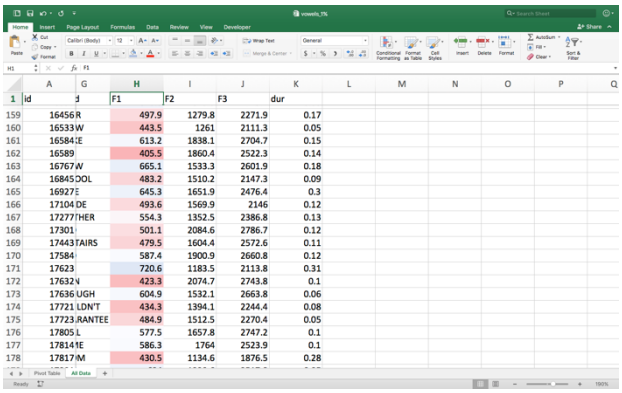
8.3 Conditional Formatting

We haven't talked too much about formatting here, but sometimes it can be handy. You can highlight certain columns or rows to make them stand out, especially if you have a lot of them. But did you know you can do what's called *conditional formatting* where a cell with be formatted in a certain way depending on its contents?

Here's an example. Image you have a bunch of number-like data (like the acoustic measurements in the vowels data). Yes, you can look at all the numbers to see the larger and smaller values, but that takes too much work. What if we could highlight all the low values red and all the high values blue? That would make them stand out, but where's the cut-off point? It would

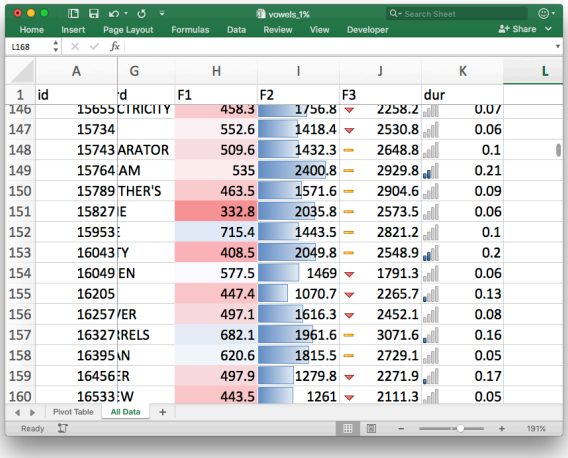
orcid.org/0000-0002-9185-0048





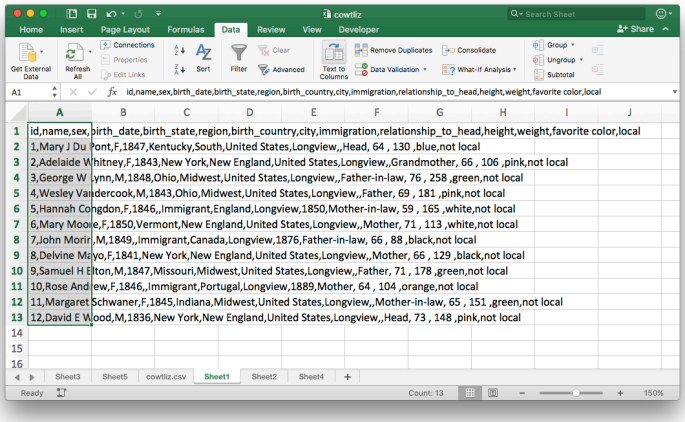
be really sweet if we could do a gradient shading where the high values *gradually* go from blue to white (for middle of the road) to red. That would be nice.

You can literally do this in three clicks. Okay, you ready? Highlight the column you want to see this done to, click on Conditional Formatting in the Home toolbar, and under Color Scales, pick the one that looks good to you. *Viola*! Just like that, now you instantly see which values are higher and lower than others.



There are some other options that are useful too. Instead of colors, you can add data bars, which fill the cell in from left to right, proportional to how high the number is. It turns it into what looks like a very tall, horizontal bar graph. Icon sets give you more discrete options if the gradience isn't for you. You can choose from an assortment of red-yellow-green icons with 3–5 levels, as well as some other ways of visualizing how big a number is.

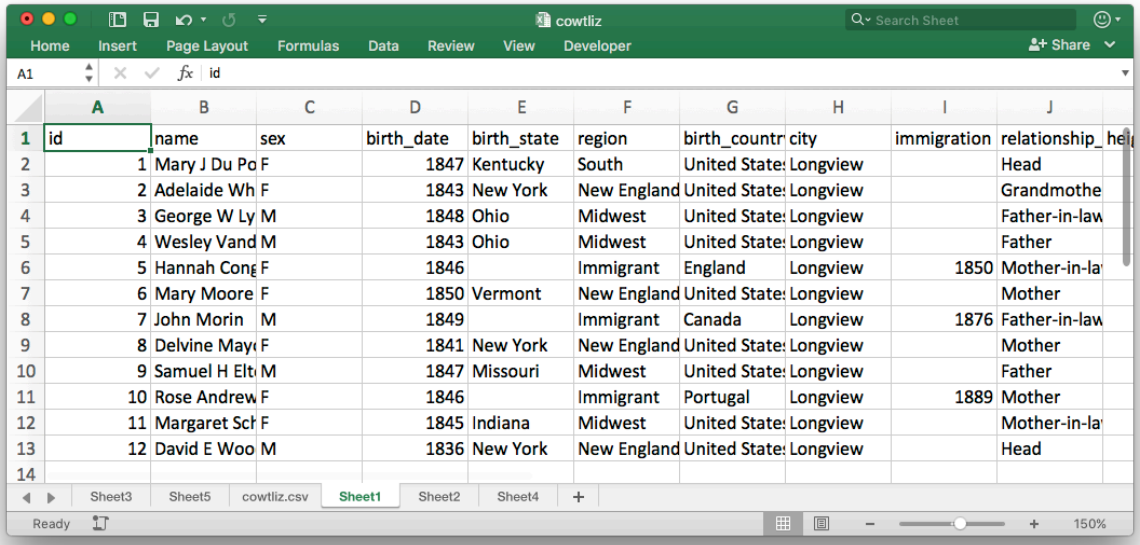
I use conditional formatting all the time. Like, pretty much every time I have a number-like column. It actually starts to look a little psychedelic after a while, but whatever. It helps me see in an instant how high or low the values are for that column. If you use the gradient ones, the extreme values stand out easily (super dark blue or super dark red), which alerts me to go back and check to see if those numbers are indeed correct.



8.4 Text-to-Columns

Sometimes when you copy and paste data in from another source, it's not formatted to be a table. I've taken some of the Cowlitz data and turned it into something I've seen before when I download a dataset or get it from some other source. When I copy and paste this into Excel, it treats the rows as one cell. What you can do it go to the Data toolbar and click Text-to-Columns.

This will take you to a series of windows (a "Wizard" in Microsoft Office terms—how appropriate for this workshop). The first asks whether the data is "delimited" meaning there's some sort of character like a tab, comma, or semicolon between each cell, or if it's Fixed width, which is only used if you have data only looks like columns in a `fixed width font`, like data that's generated through scripts on the command line. This data is delimited, so make sure that option is selected and click Next. It'll then ask what the delimiter is. In this case, it's a comma. When you check that you'll see the preview update. I would normally finish now, but you can go on and do some additional stuff like making sure the columns are formatted correctly. When you click finish, you'll see a perfectly formatted table. It's like magic.



	A	B	C	D	E	F	G	H	I	J
1	id	name	sex	birth_date	birth_state	region	birth_countr	city	immigration	relationship_
2	1	Mary J Du Po	F	1847	Kentucky	South	United State	Longview		Head
3	2	Adelaide Wh	F	1843	New York	New England	United State	Longview		Grandmothe
4	3	George W Ly	M	1848	Ohio	Midwest	United State	Longview		Father-in-law
5	4	Wesley Vand	M	1843	Ohio	Midwest	United State	Longview		Father
6	5	Hannah Cong	F	1846		Immigrant	England	Longview	1850	Mother-in-la
7	6	Mary Moore	F	1850	Vermont	New England	United State	Longview		Mother
8	7	John Morin	M	1849		Immigrant	Canada	Longview	1876	Father-in-law
9	8	Delvine May	F	1841	New York	New England	United State	Longview		Mother
10	9	Samuel H Elt	M	1847	Missouri	Midwest	United State	Longview		Father
11	10	Rose Andrew	F	1846		Immigrant	Portugal	Longview	1889	Mother
12	11	Margaret Sch	F	1845	Indiana	Midwest	United State	Longview		Mother-in-la
13	12	David E Woo	M	1836	New York	New England	United State	Longview		Head
14										

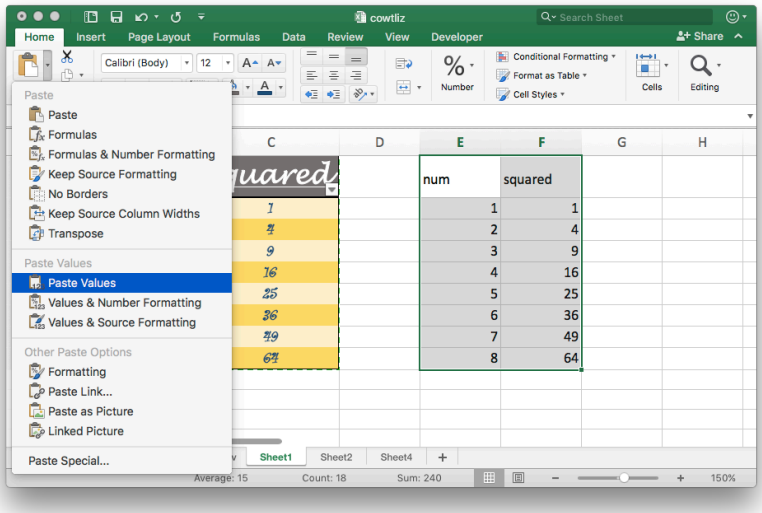
This text-to-columns feature is handy for spreadsheet-internal modifications as well. If you have a column with people's names that is formatted as "Last, First" but you want to split it up into a last name column and a first name column (the opposite of what we did with the `=concat()` function earlier), you can use text-to-columns, splitting on the comma, to do that.

Something to be aware of is that Excel can't distinguish between commas that are meant to separate columns and actual commas. So if you have some data that you're importing where one of the cells has sentences in it (some sort of corpus), there will likely be commas and Excel will split those sentences over multiple columns. Something to look out for.

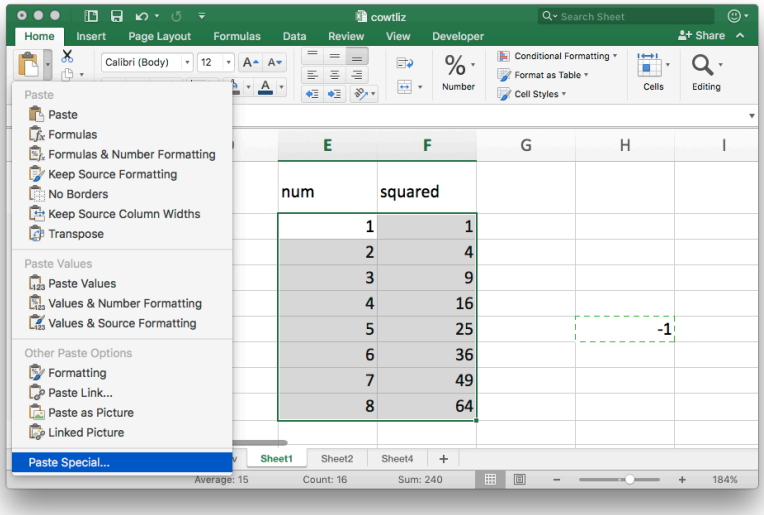
orcid.org/0000-0002-9185-0048

8.5 Paste Special

Copy and pasting stuff is great. But sometimes it doesn't work right. You'll often copy over formatting that you don't want, which can be annoying. Also, it turns out there's a lot of hidden shortcuts in copying and pasting that are super handy.

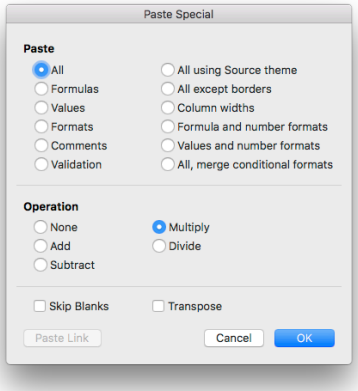


First, it's good to know how to copy and paste *just the contents* of something. So if you have some table that's been formatted in some way (font, colors, size, bold, italics, underline, borders, etc.), if you copy and paste it elsewhere, it'll keep all that junk. We don't want that.



To keep just the values, copy and table like normal and select where you want to paste it. Now, go up to Paste, and after clicking the little downwards arrow, click "Paste Values." You'll now have bare bones table, with no formatting at all.

You can, of course, keep some formatting. You keep everything but the borders, or keep the column widths, or the number formatting. You can even paste it onto an already-formatted portion of the spreadsheet and it'll make it blend right in.

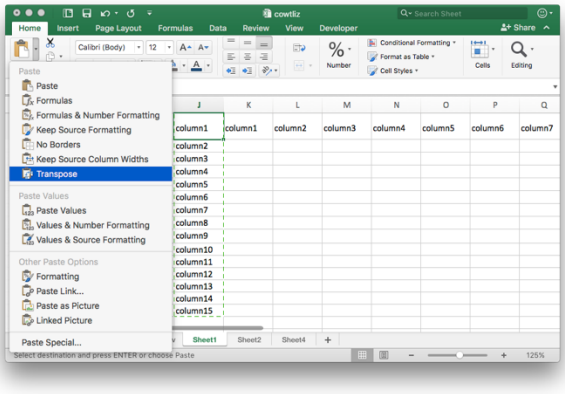


There are some other cool things you can do with paste special. For example, I'm not sure why you'd need to do this, but what if you wanted to turn all your numbers negative? You certainly wouldn't want to re-type the entire table. You could use a bunch of functions and make a parallel table. Or you could use paste special.

In any cell go ahead and type –1. Now, copy it, and highlight the cells you'd like to turn negative. At the very bottom of the Paste Special dropdown menu, there's a "Paste Special…" option that'll open a window with further options.

On the operations portion of the menu, click Multiply and then hit OK. What this does it it'll multiply all the cells you highlighted by the –1 that you copied. As you can see, you can also add, subtract, and divide as well. I don't use this too often, but it's good to know about.

Finally, another useful thing to know about is the transpose option. Let's say you want to turn all your rows into columns and vice versa. Or let's say you've got a long list of words and you want to turn them into columns? For small numbers, this is fine to do by hand, but what if you've got a list of 200? It'll be hard and annoying to type these all by hand.



For that reason, we have the transpose option. Simply highlight the portion you want to flip, and under the Paste Special dropdown, click Transpose. Perfect. Now all the columns have become rows and the rows have become columns, as if everything was rotated along the diagonal.

8.6 Overwriting functions

One very useful application of pasting the Values only is that it overwrites any functions. What this means is if you copy a cell that has a function like `=2+2`, and paste just the Value on top of the same cell, it'll overwrite the function and the cell will only contain 4, nothing else.

Now why in the world would you want to do that? I thought we learned that functions are awesome and will save the future of humanity with all the automatic updates and stuff. Well, there are at least two reasons why I overwrite them.

The first is to speed things up. I recently helped a student that, after a useful application of Transposing a list of 200 rows to make 200 columns, anchoring cell references, and lookup tables for his 40,000 rows of data, had several hundred thousand functions in a single spreadsheet. Economists, statisticians, and mathematicians will think that the formulas we used were lightweight, but for humanities research, they were pretty complex. This caused a significant strain on the student's computer, and it took a solid 15 seconds for Excel to calculate all these functions. So what we did is copied the entire table and pasted just the values on top of themselves. This turned the hundreds of thousands of battery-draining CPU-straining functions into "merely" a couple hundred thousand words. Much less work for the computer, everything runs faster, and there's less risk of Excel freezing and you losing all your work.

Another useful application of overwriting functions is for what are called ID columns. You'll notice in a lot of my datasets that the first row is just a numerical count from 1 to however many rows I have. I do this out of habit so that each row is uniquely identifiable (a relic of my brief stint in using Microsoft Access for databasing). But there is a practical application too.

Let's say you want to sort your data alphabetically or numerically in some way. You then add a filter or two. Re-sort. Remove the filters. Sort again a new way. By now the original order

orcid.org/0000-0002-9185-0048

of your data has been lost, and there's no way to get it back. If the original order meant nothing, then no harm done, but sometimes it's nice to return the table back to its original state. For this reason I include an ID column. No matter how I sort the table, I can return it back to its original order by sorting by that first column.

So what does this have to do with overwriting functions? Well, to make the ID column, I do the same trick where I have each cell equal to the cell above it +1. Once I've done that, I highlight the entire column and paste just the values to overwrite the functions. Now I have a perfect ID column.

8.7 Conclusion

I hope this workshop has been helpful. Excel is a lot more than a place to store numbers. As you can see, you can format things, find and replace, sort and filter, summarize them with pivot tables, add functions to make calculations on your tables, add lookup tables to save you typing, and visualize your data. There are so many little tips and tricks you can do to save yourself some time hidden within the menus and toolbars and I hope this workshop has exposed you to some new and useful tools to help you with your research.

Excel 2016 Shortcuts

Find shortcuts for previous Excel versions at at www.wallstreetprep.com/excel-shortcuts

Edit	Windows	Mac
Copy	Ctrl + C	Ctrl + C
Paste	Ctrl + V	Ctrl + V
Undo	Ctrl + Z	Ctrl + Z
Redo	Ctrl + Y	Ctrl + Y

File		
Open	Ctrl + O	Ctrl + O
New	Ctrl + N	Ctrl + N
Print	Ctrl + P	Ctrl + P
Save	Ctrl + S	Ctrl + S
Save as	F12	⌘ + ⇧ + S
Go to next workbook	Ctrl + Tab	Ctrl + ~
Close file	Ctrl + F4	Ctrl + W

Formatting		
Open Format Cells dialog box	Ctrl + 1	⌘ + 1
Bold	Ctrl + B	Ctrl + B
Italic	Ctrl + I	Ctrl + I
Underline	Ctrl + U	⌘ + U
Number format	Ctrl + ⇧ + !	Ctrl + ⇧ + !
Percent format	Ctrl + ⇧ + %	Ctrl + ⇧ + %
Date format	Ctrl + ⇧ + #	Ctrl + ⇧ + #
Increase font size	Alt H FG	⌘ + ⇧ + >
Decrease font size	Alt H FK	⌘ + ⇧ + <
Insert/edit comment	⇧ + F2	⇧ + F2
Increase decimal	Alt H 0	
Decrease decimal	Alt H 9	
Increase indent	Alt H 6	Ctrl + M
Decrease indent	Alt H 5	⌘ + ⇧ + M
Clear cell data	Delete	Delete
Clear cell formats	Alt H E F	Ctrl + Opt + V V
Clear cell comments	Alt H E M	Ctrl + Opt + V F
Clear all	Alt H E A	Ctrl + Opt + V C

Borders		
Outline border	Ctrl + ⇧ + &	Ctrl + ⇧ + &
Remove border	Ctrl + ⇧ + -	Ctrl + ⇧ + -
Left border	Alt H B L	⌘ + Opt + ←
Right border	Alt H B R	⌘ + Opt + →
Top border	Alt H B T	⌘ + Opt + ↑
Bottom border	Alt H B O	⌘ + Opt + ↓

Paste Special		
Paste special formats	Ctrl + Alt + V T	Ctrl + ⌘ + V T
Paste special values	Ctrl + Alt + V V	Ctrl + ⌘ + V V
Paste special formulas	Ctrl + Alt + V F	Ctrl + ⌘ + V F
Paste special comments	Ctrl + Alt + V C	Ctrl + ⌘ + V C

Ribbon	Windows	Mac
Show ribbon accelerator keys	Alt	
Show/hide ribbon	Ctrl + F1	⌘ + Opt + R

Getting around a worksheet		
Move from cell to cell	Arrows	Arrows
Go to end of contiguous range	Ctrl + Arrows	⌘ + Arrows
Move one screen up	PgUp	Fn + ↑
Move one screen down	PgDn	Fn + ↓
Move one screen left	Alt + PgUp	Fn + Opt + ↑
Move one screen right	Alt + PgDn	Fn + Opt + ↓
Go to cell A1	Ctrl + Home	Fn + Ctrl + ←
Go to beginning of row	Home	Fn + ←
Go to last cell in worksheet	Ctrl + End	Fn + Ctrl + →
Open the Go To dialog box	F5	F5

Selecting data in a worksheet		
Select a cell range	⇧ + Arrows	⇧ + Arrows
Highlight a contiguous range	Ctrl + ⇧ + Arrows	Ctrl + ⇧ + Arrows
Extend selection up a screen	PgUp	Fn + ⇧ + ↑
Extend selection down a screen	PgDn	Fn + ⇧ + ↓
Extend selection left a screen	Alt + ⇧ + PgUp	Fn + ⇧ + ⌘ + ↑
Extend selection right a screen	Alt + ⇧ + PgDn	Fn + ⇧ + ⌘ + ↓
Select all	Ctrl + A	⌘ + A

Data editing		
Fill down from cell above	Ctrl + D	Ctrl + D
Fill right from cell left	Ctrl + R	Ctrl + R
Find and replace	Ctrl + F	Ctrl + F
Show all constants	F5 Alt + S O	
Highlight cells with comments	F5 Alt S C	

Data editing when inside cell		
Edit the active cell (Edit mode)	F2	F2
While editing cell, allow use of arrow keys to create reference	F2	F2
Confirm change and leave cell	Enter	Return
Cancel cell entry and leave cell	Esc	Esc
Insert line break within cell	Alt + Enter	Opt + Enter
Highlight within a cell	⇧ + ← or →	⇧ + ← or →
Highlight contiguous items	Ctrl + ⇧ + ← or →	Ctrl + ⇧ + ← or →
Jump to beginning of cell	Home	Fn + ←
Jump to end of cell	End	Fn + →
Delete character to left	Backspace	Delete
Delete character to right	Delete	Fn + Delete
Accept AutoComplete suggestion	Tab	Tab
Reference a cell from another worksheet	Ctrl + PgUp Arrows Ctrl + PgDn Arrows	Ctrl + Fn + ↑ Arrows Ctrl + Fn + ↓ Arrows

Excel 2016 Shortcuts

Find shortcuts for previous Excel versions at at www.wallstreetprep.com/excel-shortcuts

Calculations

	Windows	Mac
Start a formula	=	=
Insert autosum formula	Alt + =	Ctrl + ⇧ + T
Recalculate all worksheets	F9	F9
Anchor cells (A1), toggle anchors (edit mode)	F4	F4
Insert a function	⇧ + F3	⇧ + F3
Enter array formula (edit mode)	⇧ + Ctrl + Enter	⇧ + Ctrl + Enter

Auditing formulas

Inspect cell values (edit mode)	F9	F9
Switch to formula view	Ctrl + ~	Ctrl + ~
Select direct precedents	Ctrl + [	Ctrl + [
Select direct dependents	Ctrl +]	Ctrl +]
Trace immediate precedents	Alt M P	
Trace immediate depedents	Alt M D	
Remove tracing arrows	Alt M A A	
Go to last cell	F5 + Enter	F5 + Enter

Excel Utilities

Recalculate all worksheets	F9	F9
Open Excel Options dialog box	Alt F O	Ctrl + ,
Accessing data validation	Alt A V V	
Get inside a drop-down list	Alt ↑ or ↓	Opt + ↑ or ↓
Insert data table	Alt A W T	
Sort a data range	Alt A S S	⇧ + Ctrl + R
Autofilter selection	Alt A T	
Insert a pivot table	Alt N V	
Insert a chart	Alt N R	
Zoom	Alt W Q	Ctrl + Mouse scroll
Name a cell or cell range	Ctrl + F3	Ctrl + L

Rows and Columns

	Windows	Mac
Select column	Ctrl + Space	Ctrl + Space
Select row	⇧ + Space	⇧ + Space
Delete row(s)/column(s)	Ctrl + -	Ctrl + -
Add row(s)/column(s)	Ctrl + ⇧ + +	Ctrl + ⇧ + +
Set column width	Alt H O W	
Autofit column width	Alt H O I	
Fit to specific row height	Alt H O H	
Group rows/columns	Alt + ⇧ + →	Opt + ⇧ + →
Ungroup rows/columns	Alt + ⇧ + ←	Opt + ⇧ + ←

Navigating across worksheets and panes

Jump to next worksheet	Ctrl + PgDn	Fn + Ctrl + ↓
Jump to previous worksheet	Ctrl + PgUp	Fn + Ctrl + ↑
Change worksheet name	Alt H O R	
Rearrange tab order	Alt H O M	
Freeze pane	Alt W F F	
Split screen	Alt W S	
Toggle from tab, ribbon, task pane, status bar	F6	
Close help (and other panes)	Ctrl + Space + C	

Moving inside Excel forms (format dialog, page setup, etc.)

Move to next control	Tab	Tab
Move from tab to tab	Ctrl + Tab	Ctrl + Tab
Move to previous control	⇧ + Tab	⇧ + Tab
Move within a list	Arrows	Arrows
Activate control	Alt + Underlined Ltr	
Toggle checkboxes	Spacebar	Spacebar
Close a dialog	Esc	Esc
Apply change	Enter	Enter

Optimal Excel settings (PC and Mac)

1. Calculation options

Open Excel settings/preferences (Alt T O on Windows, Ctrl + , on Mac). Under "Calculation options," (under the "Formulas" tab in Windows), chose "Automatic except for data tables" and click on "Enable iterative calculation."

2. Disable Autocomplete

Open Excel settings/preferences. Click off "Enable Auto-Complete for cell values. In Windows, this can be found under Options > Advanced > Editing Options.

3. Disable Error Checking

Open Excel settings/preferences. Click off "background error checking." (Found under the "Formulas" tab in Windows.)

Disabling conflicting Mac OS shortcuts

Enable Ctrl + Arrows by disabling Mission Control settings

1. Go to System Preferences > Keyboard.
2. Go to “Keyboard shortcuts” tab.
3. Click “Mission Control” in the left window.
4. Expand the “Mission Control” tab in the right window and click off “Move left a space” and “Move right a space”

Enable Ctrl+Spacebar for highlighting columns by disabling Spotlight Search

1. System Preferences > Keyboard.
2. Go to “Keyboard shortcuts”s” tab.
3. Click “Spotlight” in the left window.
4. Disable “Show Spotlight Search.”

A Note on Mac function keys

By default, Mac function keys control system settings and Mission Control. To use function keys for shortcuts, you'll need to hold down the "fn" key before you press F2, F3, etc. You can change this in **System Preferences > Keyboard** by checking "Use all F1, F2, etc. keys as standard function keys." You can now use the function keys without pressing "fn."