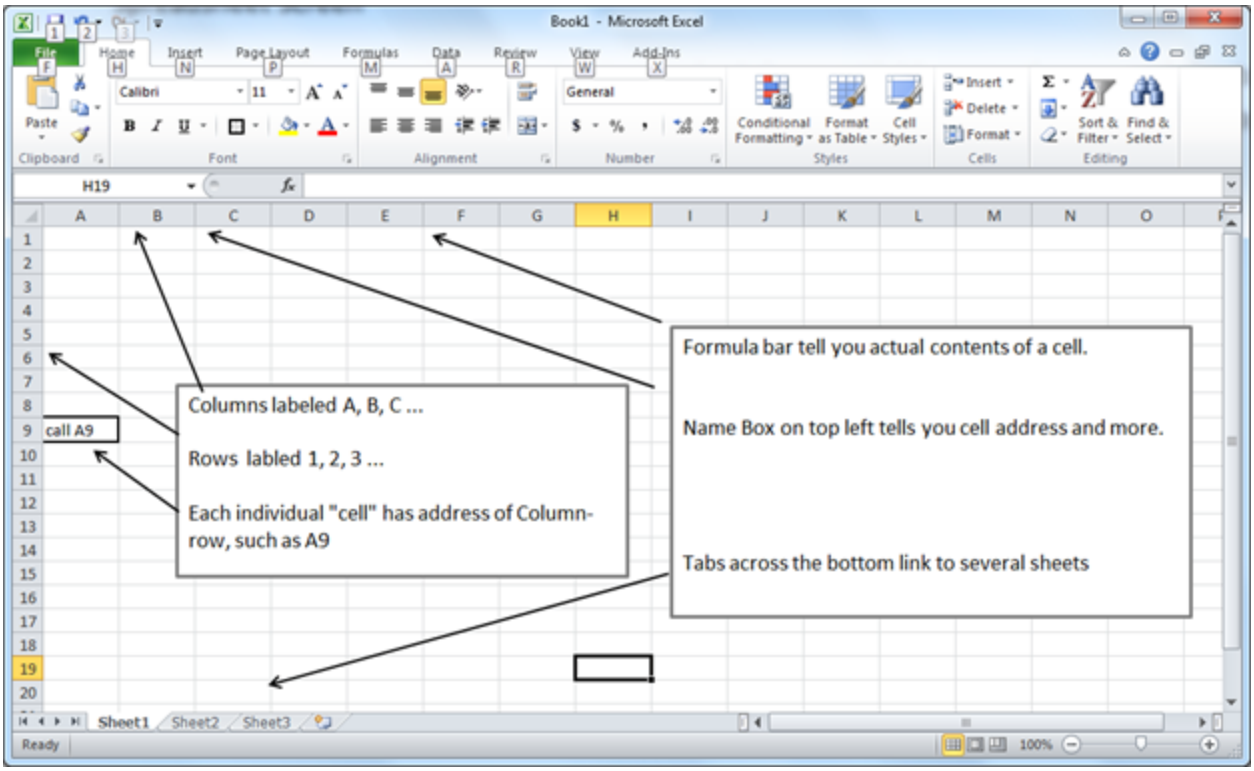
Excel pt 1

Download the spreadsheet called "MLBPayrolls2009_2011

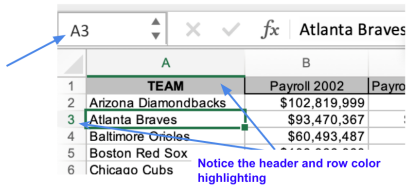
Download MLBPayrolls2009_2011" to follow along with this walkthrough.

This lab is specific to Microsoft Excel (Office 365 version and newer) for Mac, however other spreadsheet software – such as Google Sheets – have the same functionality. Of course, the buttons and specific ways to do things might be different. But I often find answers by doing a web search using the Excel terms (i.e. "how to format cells in Google Sheets). Excel for Mac has all the same functions, but looks slightly different – especially in terms of where buttons and various tools are stored.

First, a few things about spreadsheets:

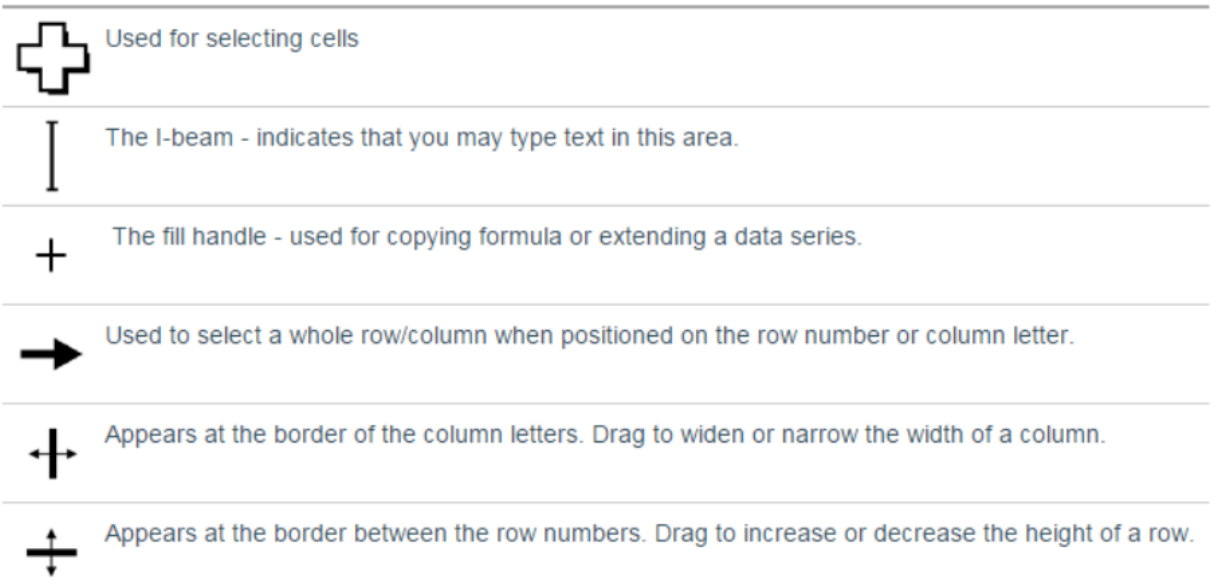


Each sheet is made up of columns (labeled by letters) and rows (labeled by numbers). Each cell is identified by a "cell address", consisting of the letter and number. For example, in this picture, the value "Atlanta Braves" is stored in cell A3. Note that when the cursor is on a cell, the cell is outlined in black. Note the small black square in the corner – we'll be using that in a couple minutes. Also note that the letter and number are highlighted in green and the cell address is displayed in the upper left corner of the sheet.



If you've ever played the board game, "Battleship", you'll find this similar.

You'll also notice that your cursor will change as you move it around. Most of the time, it will look like a big fat white plus sign. But if you hover over the top of the black box (around "Atlanta Braves") you'll get a tool that allows you to move that cell and if you hover in the lower right corner – where that little black square is – you'll get a thin black cross. That thin black cross is used for copying formulas down or across the page. We'll be using that a lot.



TERMS:

- **Columns** contain "categories" of data and are <u>vertical</u>

- **Rows** contain "individual records" and are horizontal
- A **cell address** consists of a letter followed by a number, i.e. "B4"
- The **active cell** is the cell with the darker border. Click on a cell to make it active.
- Each Excel file is described as a **workbook**, which can contain multiple **worksheets**.
- The **formula bar**, which is located in the menu bar, is similar to the address bar in a web browser. Here you can view and edit formulas that you've created in the worksheet.

Excel's original purpose was for accountants, so it is really useful for doing math on small or large datasets. But you'll find that Excel can do a lot of things – more than you'll probably ever need – including summarizing large datasets. This tipsheet is going to cover some of the main mathematical functions you will likely use. After learning these, the next thing you should learn is how to use Excel's Pivot Tables, which allows you to summarize your data (i.e. who is the largest donor in the governor's race? What's the average pay for each position in Major League Baseball? How many times has each restaurant been inspected?)

Excel's original purpose was for accountants, so it is really useful for doing math on small or large datasets. But you'll find that Excel can do a lot of things – more than you'll probably ever need – including summarizing large datasets. This tipsheet is going to cover some of the main mathematical functions you will likely use. After learning these, the next thing you should learn is how to use Excel's Pivot Tables, which allows you to summarize your data (i.e. who is the largest donor in the governor's race? What's the average pay for each position in Major League Baseball? How many times has each restaurant been inspected?)

F8 | 73558500

	A	B	C	D	E	F	G
1	**TEAM**	Payroll 2002	Payroll 2003	Payroll 2004	Payroll 2005	**Payroll 2009**	**Payroll 2010**
2	Arizona Diamondbacks	$102,819,999	$80,657,000	$69,780,750	$62,329,166	$73,516,666	$60,718,166
3	Atlanta Braves	$93,470,367	$106,243,667	$90,182,500	$86,457,302	$96,726,166	$84,423,666
4	Baltimore Orioles	$60,493,487	$73,877,500	$51,623,333	$73,914,333	$67,101,666	$81,612,500
5	Boston Red Sox	$108,366,060	$99,946,500	$127,298,500	$123,505,125	$121,745,999	$162,447,333
6	Chicago Cubs	$75,690,833	$79,868,333	$90,560,000	$87,032,933	$134,809,000	$146,609,000
7	Chicago White Sox	$57,052,833	$51,010,000	$65,212,500	$75,178,000	$96,068,500	$105,530,000
8	Cincinnati Reds	$45,050,390	$59,355,667	$46,615,250	$61,892,583	$73,558,500	$71,761,542
9	Cleveland Indians	$78,909,449	$48,584,834	$34,319,300	$41,502,500	$81,579,166	$61,203,966
10	Colorado Rockies	$56,851,043	$67,179,667	$65,445,167	$48,155,000	$75,201,000	$84,227,000
11	Detroit Tigers	$55,048,000	$49,168,000	$46,832,000	$69,092,000	$115,085,145	$122,864,928
12	Florida Marlins	$41,979,917	$49,050,000	$42,143,042	$60,408,834	$36,834,000	$57,029,719
13	Houston Astros	$63,448,417	$71,040,000	$75,397,000	$76,779,000	$102,996,414	$92,355,500
14	Kansas City Royals	$47,257,000	$40,518,000	$47,609,000	$36,881,000	$70,519,333	$71,405,210
15	Los Angeles Angels	$61,721,667	$79,031,667	$100,534,667	$97,725,322	$113,709,000	$104,963,866
16	Los Angeles Dodgers	$94,850,953	$105,872,620	$92,902,001	$83,039,000	$100,414,592	$95,358,016

Any time you have a list of numbers that require some sort of calculation like percent change from one year to the next, percent of total, average or just a simple total, it's easier to put the figures into Excel and let the computer do the tough stuff for you. Excel is also extremely helpful when comparing a list of numbers to find out which one is the biggest, which is the smallest, etc.

This tipsheet will use total Major League Baseball payrolls to show how to calculate percent change from year to year, percent of total, average payroll and how to sort the list to compare the teams.

Here's a sample of what the data looks like when we start out.
Before doing any work on your data, here are a few other tips to keep in mind:

1. Make sure all your columns are labeled and that the column label only takes up one row. I frequently get datasets from government agencies that have two rows worth of labels – you'll have trouble sorting this type of spreadsheet, so you'll want to change the header/label row before you start working. Note that in the picture above, my column labels are highlighted in grey (I did this by highlighting the cells I wanted to change to grey and right-mouse click to choose "Format Cells", then changed the "Fill" color to grey. You don't need to color them like this...I just find it's useful to have the labels jump out at me.
2. Make sure that any stray information that came with your dataset – such as a source tag at the bottom or notes or titles or anything above the data table – are separated

3. Make sure your data table doesn't have any empty rows or empty columns. All the data should be neatly set up in consecutive columns and rows. In this case, we have one row for each team and each row has the team name on it. Then there are columns for the total each team spent on payroll in 2009, 2010 and 2011. This is a really important point because government agencies that send you data in Excel (or post it on their website) are horrible about following this. I often get datasets with blank columns jammed in the middle – seemingly to just make it look better.

Shortcuts and other useful tips

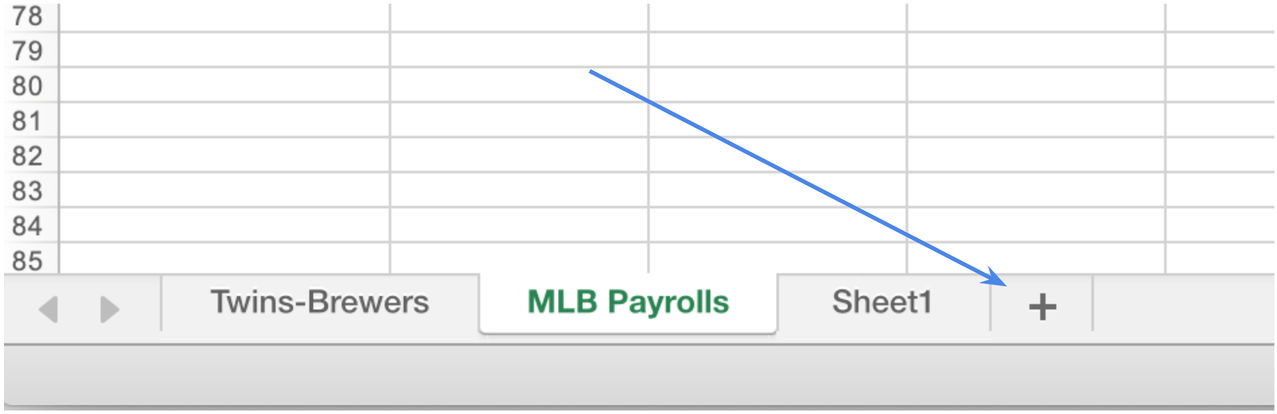
Freeze Panes: To lock your field names in place so that you can always see them when you scroll down the page. Place your cursor in the cell just below the row that you want to lock into place, and all the way to the left of the page. Then go to View menu and select "Freeze Panes." It gives you several options, such as freezing the top row or freezing the far left column.

Paste Special: Use this function when you want to get rid of the formulas behind a column or row of data (until you get rid of the formulas, that row will be dependent on the other rows/columns used to calculate the formulas). Highlight the row(s) or column(s) that have the formulas. Copy the data using Ctrl-C or the copy button. Then put your cursor where you want to paste the data (best choice is to put it in a new column or row) and right-mouse click, then choose "Paste Special." A little box will come up: Under the Paste section at the top, choose "Values". Then say OK.

Hide columns: You can hide columns to get them out of your way or to avoid printing them by highlighting the columns you want (click on the letter at the top of the column) and right-mouse clicking to choose "Hide columns". The columns will disappear. To get them back, highlight the two columns on either side of the ones that are missing and right-mouse click and choose "Unhide columns."

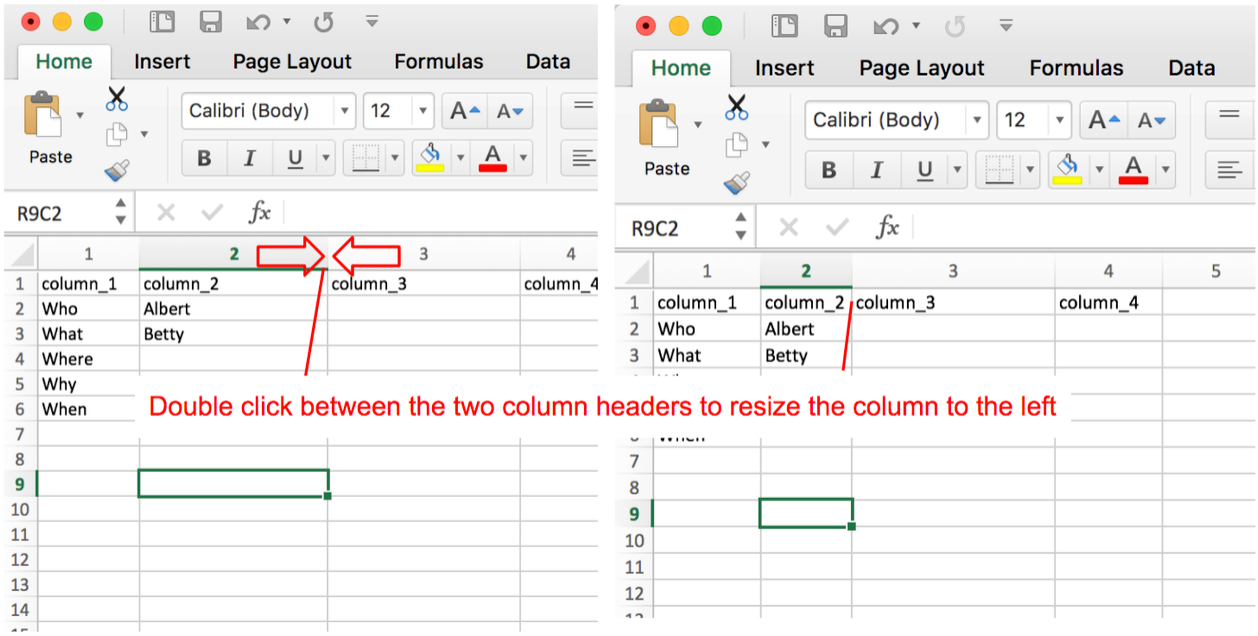
Add columns or rows: To add a new column in the middle of your data, click on the letter of the column immediately to the RIGHT of where you want the new column. So if you want a new column between columns B and C, you should click on C. Make sure the whole column is highlighted (if you clicked on "C", Excel will automatically highlight the whole column) Then right-mouse click and choose "insert".
To add a new row, click on the number of the row just BELOW where you want a new one. Make sure the whole row is highlighted, then right-mouse click and

choose “insert”

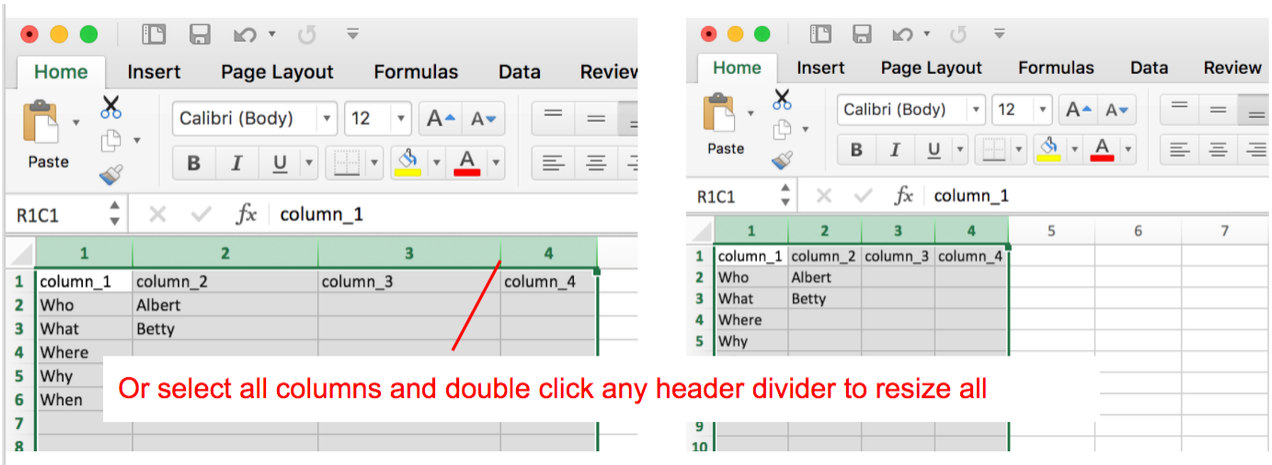


Worksheets: You can toggle between different worksheets in an Excel workbook using the tabs in the lower left corner that have default names of “Sheet 1”, “Sheet 2”, etc. To change the name, double-click on the “Sheet 1” and it will turn black. Then you can start typing to give it a new name. To add a new worksheet, you’ll see a little button to the right of the tabs that gives you the option to add a new worksheet. You can move worksheets around so they appear in a different order, by clicking on the named tab with your mouse, holding down and dragging it in whatever direction you would like.

Adjusting column width: For better readability, you can resize column widths by dragging the line between columns left and right.
Or by selecting a single column head and double clicking the line between column headers. This resizes the column to fit the widest data point in that column.



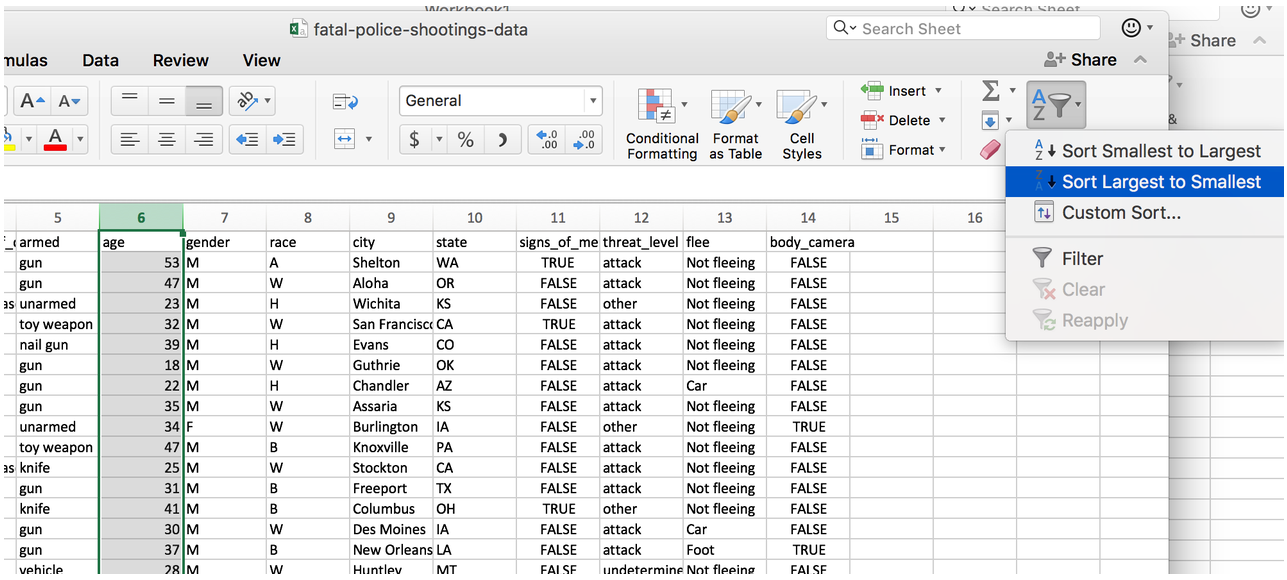
OR! Just click that square where the row and column headers meet (this selects all the cells in the sheet) and then double click just one line between a pair of columns. This will apply the resizing to all of the columns now.



Importing Fatal Force data

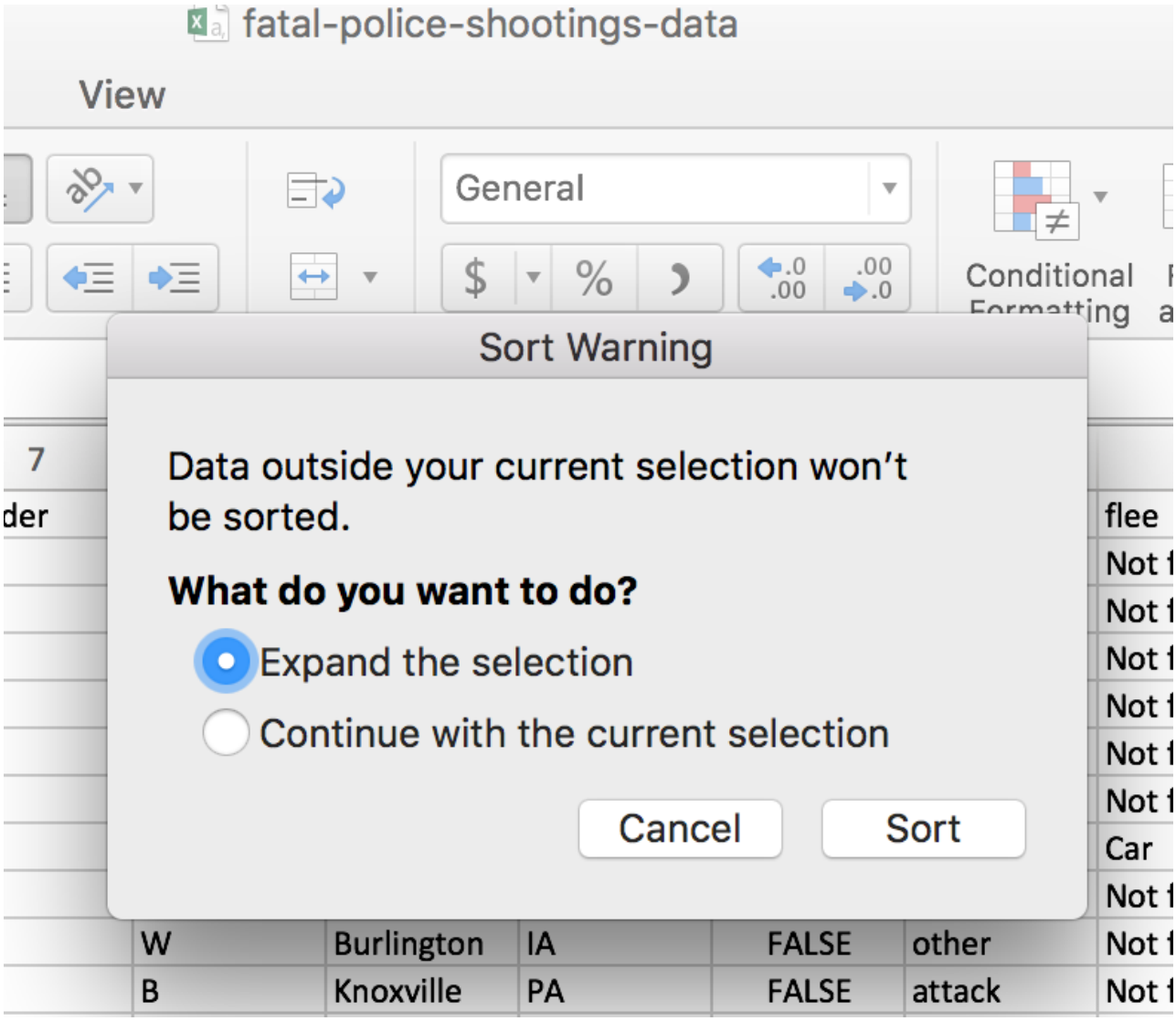
Let's explore another data set: [fatal-police-shootings-data.csv (Links to an external site.)]

Download the file and bring it into Excel so we can analyze it.
How?
Select `File > Open and find the file`
Or just double click the file in your file finder and let Excel open it automatically.
Or right click the file and say open with Excel.
Wait, what's a CSV file?
It stands for **C**omma **S**eparated **V**alues because if you look at the raw file, the data is separated by commas. In contrast, XLS stands for e**X**ce**L** **S**heets. If you look at the raw file, it's just a bunch of code that only Microsoft's Excel program can decipher.
Sorting
We're looking at police shootings.
First things to do when exploring data is to see the range. Max and min and that sort of thing.
Question 1 we can quickly answer: What's the youngest age someone has been shot by police?
We can do that by sorting the `Age` column.
Hover on the right side of the column header and a **small triangle** will appear. Click on it and a bunch of options will show up.
Select sort `A - Z` (this means descending order, like the alphabet). Vice versa, `Z - A` means ascending order.

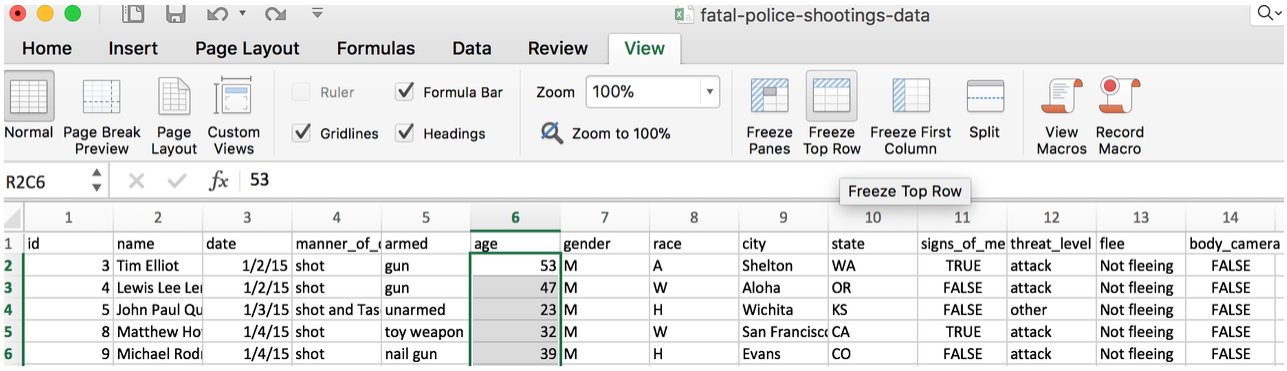


armed	age	gender	race	city	state	signs_of_me	threat_level	flee	body_camera
gun	53	M	A	Shelton	WA	TRUE	attack	Not fleeing	FALSE
gun	47	M	W	Aloha	OR	FALSE	attack	Not fleeing	FALSE
unarmed	23	M	H	Wichita	KS	FALSE	other	Not fleeing	FALSE
toy weapon	32	M	W	San Francisc	CA	TRUE	attack	Not fleeing	FALSE
nail gun	39	M	H	Evans	CO	FALSE	attack	Not fleeing	FALSE
gun	18	M	W	Guthrie	OK	FALSE	attack	Not fleeing	FALSE
gun	22	M	H	Chandler	AZ	FALSE	attack	Car	FALSE
gun	35	M	W	Assaria	KS	FALSE	attack	Not fleeing	FALSE
unarmed	34	F	W	Burlington	IA	FALSE	other	Not fleeing	TRUE
toy weapon	47	M	B	Knoxville	PA	FALSE	attack	Not fleeing	FALSE
knife	25	M	W	Stockton	CA	FALSE	attack	Not fleeing	FALSE
gun	31	M	B	Freeport	TX	FALSE	attack	Not fleeing	FALSE
knife	41	M	B	Columbus	OH	TRUE	other	Not fleeing	FALSE
gun	30	M	W	Des Moines	IA	FALSE	attack	Car	FALSE
gun	37	M	B	New Orleans	LA	FALSE	attack	Foot	TRUE
vehicle	28	M	W	Huntley	MT	FALSE	undetermine	Not fleeing	FALSE

Wait, what's this?

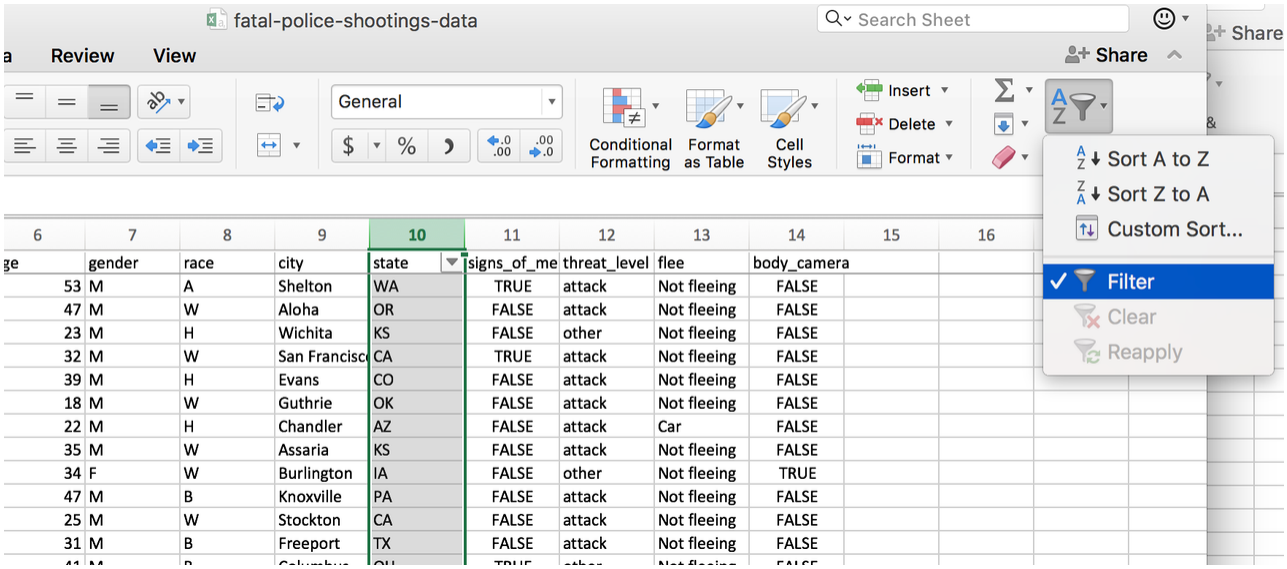


If **Continue with the current selection** was selected, then only the column would be sorted and all other columns would stay the same.
This is not good.
Instead, select **Expand the selection** so when the spreadsheet is sorted, the associated columns will sort with it.
SOMETIMES EXCEL WONT GIVE YOU THIS OPTION (or it won't work as it should)
If that happens, we need to freeze the top row.
Select the **View** tab at the top and click on the **Freeze Top Row** button.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	id	name	date	manner_of_	armed	age	gender	race	city	state	signs_of_me	threat_level	flee	body_camera
2	3	Tim Elliot	1/2/15	shot	gun	53	M	A	Shelton	WA	TRUE	attack	Not fleeing	FALSE
3	4	Lewis Lee Le	1/2/15	shot	gun	47	M	W	Aloha	OR	FALSE	attack	Not fleeing	FALSE
4	5	John Paul Qu	1/3/15	shot and Tas	unarmed	23	M	H	Wichita	KS	FALSE	other	Not fleeing	FALSE
5	8	Matthew Ho	1/4/15	shot	toy weapon	32	M	W	San Francisc	CA	TRUE	attack	Not fleeing	FALSE
6	9	Michael Rod	1/4/15	shot	nail gun	39	M	H	Evans	CO	FALSE	attack	Not fleeing	FALSE

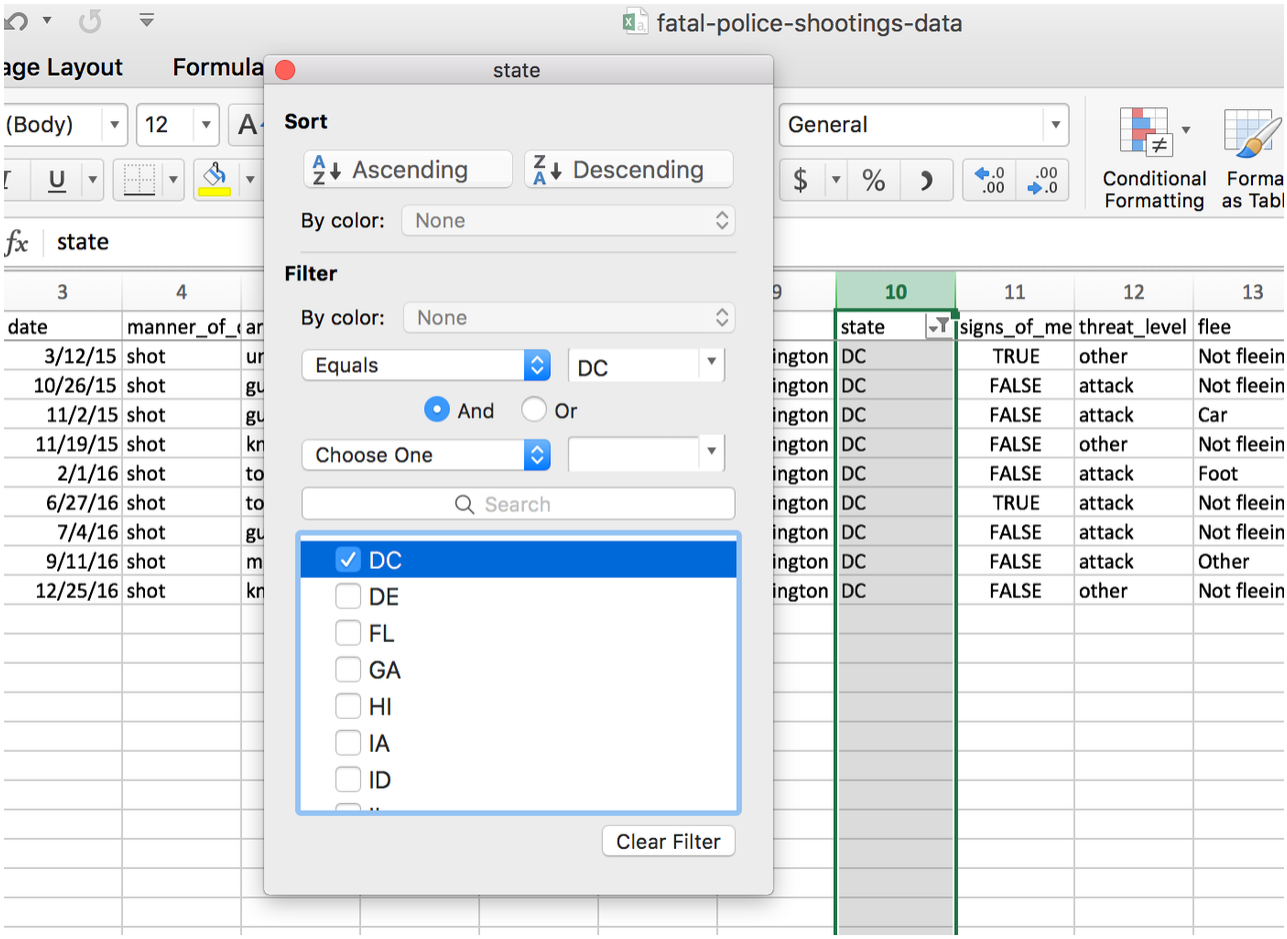
You can also freeze first column of data.
This is particularly useful if you have a lot of data and rows and want to scroll through the data without forgetting what column header was what.
Filtering
We have thousands of rows which can be overwhelming.
Let's narrow it down by focusing just on the shootings that happened in DC.
We have to enable filtering first.
For a single column, select the column header of the column you want to filter `state` and then click on the **Sort & Filter** button (in the **Home** tab).



6	7	8	9	10	11	12	13	14	15	16
ge	gender	race	city	state	signs_of_me	threat_level	flee	body_camera		
53	M	A	Shelton	WA	TRUE	attack	Not fleeing	FALSE		
47	M	W	Aloha	OR	FALSE	attack	Not fleeing	FALSE		
23	M	H	Wichita	KS	FALSE	other	Not fleeing	FALSE		
32	M	W	San Francisc	CA	TRUE	attack	Not fleeing	FALSE		
39	M	H	Evans	CO	FALSE	attack	Not fleeing	FALSE		
18	M	W	Guthrie	OK	FALSE	attack	Not fleeing	FALSE		
22	M	H	Chandler	AZ	FALSE	attack	Car	FALSE		
35	M	W	Assaria	KS	FALSE	attack	Not fleeing	FALSE		
34	F	W	Burlington	IA	FALSE	other	Not fleeing	TRUE		
47	M	B	Knoxville	PA	FALSE	attack	Not fleeing	FALSE		
25	M	W	Stockton	CA	FALSE	attack	Not fleeing	FALSE		
31	M	B	Freeport	TX	FALSE	attack	Not fleeing	FALSE		

The little triangle will pop up permanently now in the header.

Alternatively, you can enable filtering on all the columns by making sure nothing is selected before clicking the **Sort & Filter** button.
Now, select the little blue triangle in the `state` column and then click **`Select all`** to get rid of all the states and scroll down and click `DC` so there's a check next to it.
Click `OK`.



Alright, we have turned thousands of rows to just nine. Between 2015 and 2016, nine people were shot by police officers. It looks like all of them were holding some form of weapon (or something that could be mistaken for one).
A couple of them exhibited signs of mental health issues and half were perceived as attacking officers. All of them were black. Only two of the officers involved were wearing body cameras.

Next steps

Congratulations, now you're ready to dig a little deeper with formulas and then eventually pivot tables, which summarizes the data. This enables you to answer

questions like: Which states have the most shootings? What's the average age of victims?