Excel I – Sorting and filtering

Revised February 2013

Nerd notes: total number of columns in a worksheet = 256

total number of rows in a worksheet = 65,536 (old)/1 million (Excel 2007)

total number of characters per cell = 32,000

♦ **Total number of rows:** If you receive a spreadsheet with exactly 65,536 rows, contact the source. It's likely the table is larger and their Excel version cut it off when it reached its limit.

THE EXCEL GRID - CELL ADDRESSES

When you first start up Excel, here's how it looks. Across the top are letters, and down the side are numbers. Just like on a map, you refer to each square or cell by its intersection of column and row:
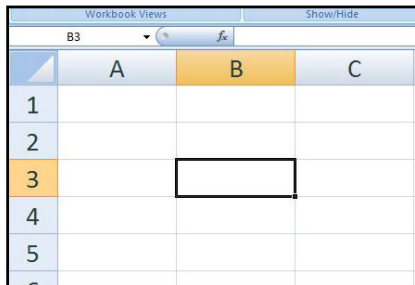
B3 is the cell that's currently active. You can tell because its address is shown on the upper left corner, and the cell itself is outlined with a black border, which has a little knob on the lower right corner.

EXCEL MOUSE SHAPES

Here are the four main mouse shapes you'll see:

♦ The Big Fat White Plus Sign, or BFWPS:

You can never get into trouble with this symbol. Whenever you see it, you can select a cell or a group of cells.

♦ The Copy Tool, or the thin black cross:

When you see this, you'll copy anything that's selected in whichever direction you drag. This can be good or bad. Normally, you only copy formulas.

♦   The Evil Arrow:

This will actually *move* whatever you have selected, sometimes directly on top of something else. Don't use this most of the time.

Undo is your friend.

♦   The Column- and Row-adjuster

You'll see this only when your mouse is hovering in the gray, title, area of either a row or column. It widens or narrows a column, or lengthens or shortens a row.

OK, YOU HAVE DATA, NOW WHAT?

Open *Countypop.xls* and scroll across to review the fields:

	A	B	C	D
	State			
1	**State**	**County**	**2000pop**	**2007pop**
2	Alabama	Autauga County	43,671	49,960
3	Alabama	Baldwin County	140,415	171,769
4	Alabama	Barbour County	29,038	27,941
5	Alabama	Bibb County	20,826	21,535
6	Alabama	Blount County	51,024	56,614
7	Alabama	Bullock County	11,714	10,781
8	Alabama	Butler County	21,399	20,157
9	Alabama	Calhoun County	112,249	113,103
10	Alabama	Chambers County	36,583	34,764
11	Alabama	Cherokee County	23,988	24,560

- *Pop* shows total population (from the 2000 census and the 2007 estimates).
- *Births* shows the total number of births by county.
- *Deaths* shows the total number of deaths by county.
- *Internat* shows the number of new county residents that year who moved from outside the United States.
- *Netmigration* shows the total population change that year involving people who moved into the county (either from outside the United States or from some other place in the United States)

DATA FORMATTING

♦**Lock header row: View →** **Freeze panes →** **Freeze panes or freeze top row**
♦ Resize the width of a column by clicking and dragging on the vertical bar between column letters. Do multiple columns at once.
♦ Click on a row number to highlight (define) a whole row to make formatting changes; same with columns:

DATA DESIGN AND MANIPULATION TECHNIQUES

♦ Insert a row by highlighting a row and then **Insert → Row.**
♦ Insert a column by highlighting a column and then **Insert → Column**
♦ Delete the contents of a cell by highlighting and using the **Del** key
♦ Hide and unhide a column or columns using the menu bar method -- Unhide hint: define the column letters on either side of the hidden area
♦ Hide and unhide a row or rows using the mouse method or menu bar method

DATA MANIPULATION THROUGH SORTING

In this exercise, we will learn how to rank information in a variety of ways. Think about all of the possible applications with data as we work through this section.

Let's find the **top 10 counties** in the country based on 2007 population (column D).

To sort the data, highlight the whole spreadsheet by clicking on the square above Row 1.

Then click on the Data tab at the top. Next, click on the Sort icon next to the Filter funnel:
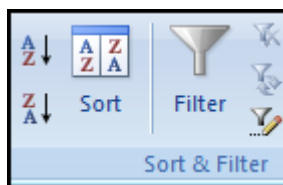
In the next window, use the pull-down menus to fill in the choices this way:

Column		Sort On		Order	
Sort by	POPESTIMATE2007	Values		Largest to Smallest	

Note that the "My data has headers" box is checked. If Excel guessed incorrectly, you could turn that off to make sure you're sorting all of the records.

Look at your results after you run the Sort function.

Your top 10 counties should look like this:

	A	B	C	D
	A1	State		
	State	**County**	**2000pop**	**2007pop**
1				
2	California	Los Angeles County	9,519,338	9,878,554
3	Illinois	Cook County	5,376,741	5,285,107
4	Texas	Harris County	3,400,578	3,935,855
5	Arizona	Maricopa County	3,072,149	3,880,181
6	California	Orange County	2,846,289	2,997,033
7	California	San Diego County	2,813,833	2,974,859
8	New York	Kings County	2,465,326	2,528,050
9	Florida	Miami-Dade County	2,253,362	2,387,170
10	Texas	Dallas County	2,218,899	2,366,511
11	New York	Queens County	2,229,379	2,270,338

Other sorts: by state, then county; by state, county, then highest pop; by state, netmigration, etc.

DATA MANIPULATION THROUGH FILTERING

Next, click on the Filter funnel next to the Sort button:

Turn on the automatic filter by clicking on the Filter icon at the tip (it looks like a funnel):
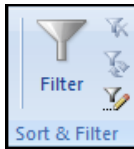
Notice that little pulldown arrows show up at the top of each column.

♦   custom filter wildcard searches -- filter so only those records associated with a certain county or state are visible -- review of mathematical operators as Excel understands them! (=, >, <, >=, <=, <>)
♦   custom filter **"and"** vs. **"or"**
♦   while a subtotal or filter is turned on...copying only visible cells via

In this exercise, we will learn how to extract/filter subsets of the data and copy them into other worksheets inside our main workbook. This is a common part of data analysis, especially if you're using national or statewide data sets and want to extract data for a smaller area. Once you have carved out smaller subsets of data, the analysis can be much simpler and less error-prone.

On the State column, filter for your state; I did New Mexico. Copy those records – **Home → Find and Select → Go To Special… → Visible cells only –** and paste them into a new worksheet.

 Name the tab at the bottom the name of your state.

Go back to the Counties worksheet, so we'll be looking at data for every county in the United States.

Let's say that we want to do an analysis that focuses on counties with more than 1 million residents (as of the 2007 population estimates). Turn the filter on this worksheet.
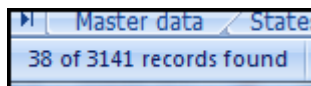
You need to go into the filter for Column D and slide down to "Number Filters / Greater Than…"
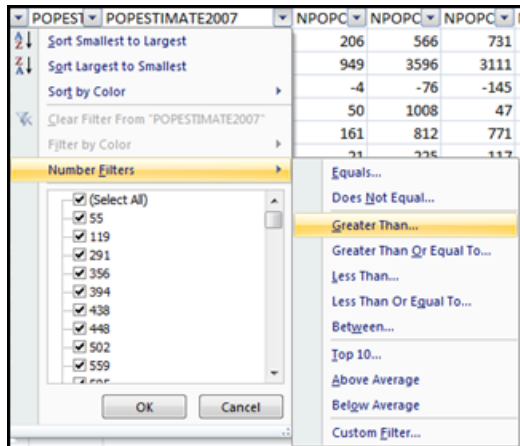
Another window will open and let you type in your criteria (e.g. greater than 1000000):
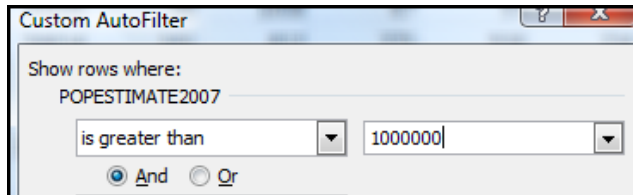
How many counties had 2007 population estimates of more than 1 million people?

If you really were doing this analysis, you could copy those 38 records and paste them into another worksheet for your analysis.

Turn OFF the filter by clicking on the Filter icon.

(Tip: I generally move the filters before turning them off.)