

Excel: Percentages, ratios, filtering, ranking

This is an exercise that uses a spreadsheet called “Teachers.xls”, which contains one record for each public school in the Twin Cities, showing demographics of the teachers at each one.

Each record shows:

- The number of students (“enroll”)
- The number of teachers (“totFTE” – refers to full-time equivalent)
- Teachers totaled by race and gender (series of fields between columns K and T)
- Numb of FTE teachers that are in their first year (“NewTch”)
- Average years of teaching experience (“avgyrsexp”)
- Education levels of teachers (columns W through AA)
- Average age of teachers (“AvgAge”)

You can see that this set of data is fairly large ... you need to scroll to the right and down in order to see all the columns and rows.

What questions might we want to ask this data?

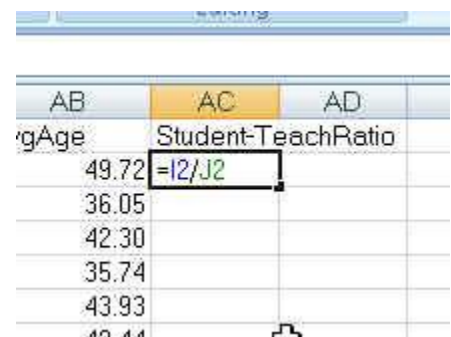
- What’s the student-teacher ratio at each school?
- What percentage of the teachers have advanced education?
- Which schools have a high percentage of male teachers?
- Which schools have the greatest diversity in their teacher ranks?

We can answer all of these questions from this data, however you’ll notice that the answers for all of these questions aren’t readily available in the data. We will need to calculate them – for example, the student-teacher ratio is not in this data, but we have the basic pieces we need to calculate that (the number of students and number of teachers).

Let’s start with student-teacher ratio. This is a type of rate that puts all the schools on an even playing field, regardless of how “big” the school is (in terms of number of students). This will give us a 1:X ratio – 1 teacher for every X number of students.

Let’s start by adding a new column to our table. In column AC, put a header in that says “Student-Teach Ratio”.

Then in cell AC2, let’s put our formula:



AB	AC	AD
gAge	Student-TeachRatio	
49.72	=I2/J2	
36.05		
42.30		
35.74		
43.93		
42.44		

Hit enter and copy down to all the records (make sure it goes all the way to the bottom!)

You might need to highlight the column, then right-mouse click and “Format Cells” to set it as a number (I’d recommend 1 decimal place)

Then put your cursor in the middle of your data and go to the Data menu and choose “Sort”. Sort by this new column ascending (smallest to largest).

You’ll see that we have a school with a 1 to 3.8 ratio. Isn’t that really low? Let’s look at the school name-- it’s a kindergarten center. That explains it. This isn’t a full elementary school, so maybe it’s an “outlier” and we might need to consider eliminating this school (and possibly others?) if they aren’t truly an “apples-to-apples” comparison to other schools.

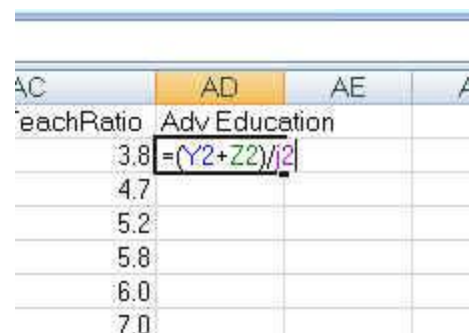
Let’s also look at the highest ratios at the bottom of the spreadsheet. We’ve got a school with 1 teacher to 558 students. Again, maybe an outlier or perhaps this is a case of dirty data? Maybe the school didn’t report the correct number of teachers? Would be something you’d want to look into before using the data for a story.

Our next question is about how well-educated the teachers are. Are there schools that have a greater proportion of teachers with advanced degrees? (a masters degree or PhD). Again we’ve got a situation where this answer isn’t already calculated but we’ve got the basic data we need. We have 5 fields showing the number of teachers with that level of education – Less than a bachelor’s degree, bachelor’s degree, master’s degree, PhD and “education unknown”.

So if we want to see the percentage of teachers with advanced education, we’d need to add together MA and PHD and then divide by the total number of teachers (totFTE).

Let’s do that in column AD. Since we’re doing two math calculations in one formula, the first one needs to be surrounded by parentheses.

Hit enter and copy the formula down. Make sure it goes to the bottom. Then you will likely need to Format Cells to change to percentage.



The screenshot shows a spreadsheet with columns AC, AD, AE, and AF. Row 1 contains headers: 'eachRatio' under AC, 'Adv Education' under AD, and empty cells under AE and AF. Row 2 shows a value of 3.8 in column AC and the formula $= (Y2 + Z2) / J2$ in column AD. Rows 3 through 7 show values 4.7, 5.2, 5.8, 6.0, and 7.0 in column AC, with corresponding empty cells in column AD.

AC	AD	AE	AF
eachRatio	Adv Education		
3.8	$= (Y2 + Z2) / J2$		
4.7			
5.2			
5.8			
6.0			
7.0			

Sort your results and see which schools have the highest percentages.

Our next question looks at the gender mix. Traditionally, teaching has been a female-dominated profession. Are there any schools that buck that trend?

Let’s start by figuring out where the data shows us the number of male teachers. It’s a bit trickier than we’d like – there are 5 fields (one for each race) tallying up the number of men. So the first thing we

should do is create a new column that shows the total men, and we might as well make one for total women too (just in case we need it for something else)

Let's create two new columns – Men and Women. We can write one formula that will work to populate both of them. Notice that the race columns start with the men first, then the women. We need to follow that pattern in order to just use one formula.

So put “Men” in the AE column and “Women” in the AF column.

Now we'll just do some simple addition by adding together all the “male” columns:

=K2+M2+O2+Q2+S2

Hit enter. Then grab the copy tool and copy the formula across to the AF column. Look at the formula and note that it automatically adjusted to all the female columns:

=L2+N2+P2+R2+T2

This trick only worked because all the fields were lined up next to each other.

Copy the two formulas down the page (remember they do have decimal points because these are FTE counts – some teachers are part-time, which is counted as less than 1)

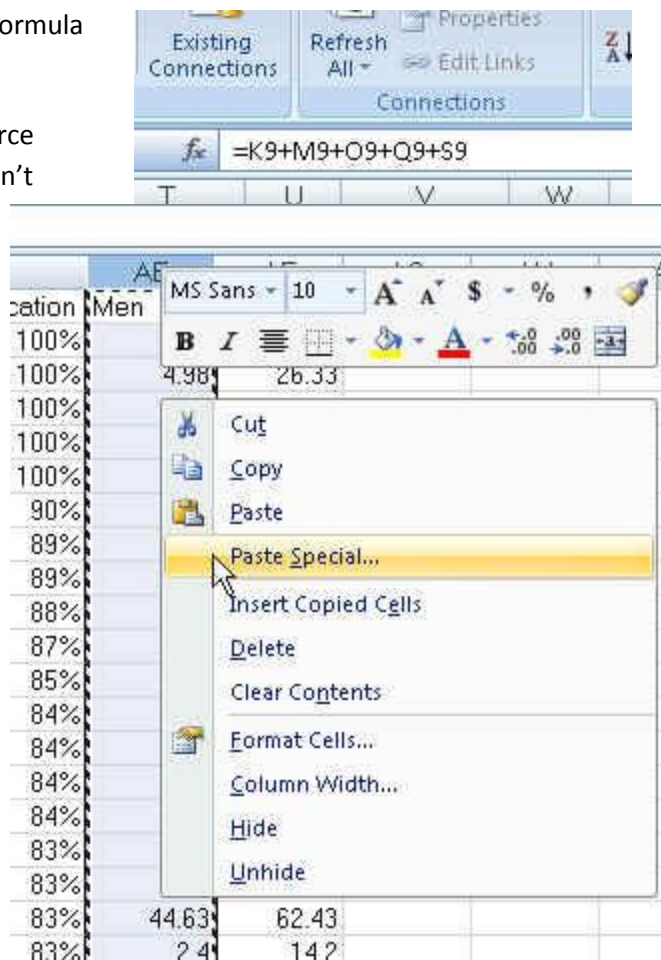
When you work with formulas like this, one thing you have to remember is that the answer showing up in your new columns is still dependent on the source columns. Click on any cell in the new “Men” row and up in the formula bar you'll see that there's a formula there.

In other words, if you delete the data from the source columns, your answer will disappear. Although I don't recommend deleting source data (it's best to keep it), sometimes you might want to “save” your answer just to ensure that nothing happens to it.

You do that with a little thing called “paste special-values”

Here's how to do that:

Highlight the entire “Men” column (do this by clicking on the “AE” at the top of the column). Then right-mouse click and choose “Copy”. Then you'll notice the column will be surrounded by

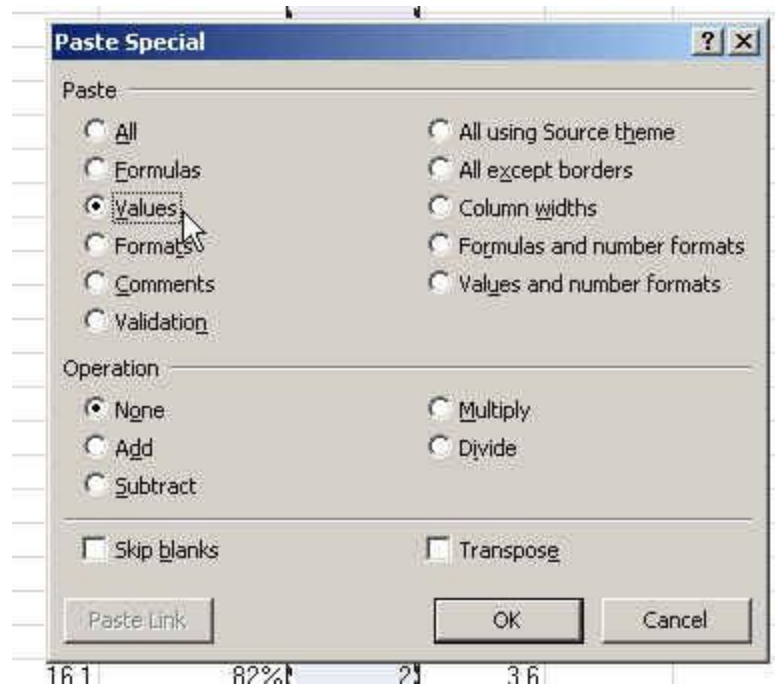


what I call “running ants”. Then right-mouse click again and choose “paste special”

When you click on Paste Special, it will bring up a dialog box where you can select “Values”

Now when you click on a cell in the “men” column, the formula bar will display the answer – no formula!

Repeat this for the “Women” column.



To get the answer to our question, we still need to calculate the percentage of male teachers at each school. We’ll do that by taking the number of men and dividing it by the Total FTE number.

Hit enter and copy the formula to the bottom of the data.

You might need to “Format Cells” to change it to a percentage.

	AE	AF	AG	AH
	Men	Women	Pct Men	
	1.17	1.51	=AE2/J2	
	4.98	26.33		
	1.02	12.01		
	0.99	7.79		
	1	3		

Then sort the data by this new column. Yep, we do have some schools that defy the stereotype.

Our last question was to find out which schools have the greatest diversity in their teacher ranks. Again, this is a situation where we have several fields that tell us the race of the teachers but no single measure of “diversity.” There are a couple ways you could go about this. The simplest would be to calculate what percentage of the teachers are white. The schools with the lowest percentage would have the highest diversity. The other way would be to add together all the teachers who are NOT white and divide that by the total teachers – this would give you the percentage of teachers who are a minority race.

Let's do the second approach and we'll do the whole calculation in one formula (unlike the gender one, where we did added up the totals in new fields first, before dividing by the total teachers)

Create a new column in AH and call it "diversity"

Our formula will be to add together all the numbers in columns K through R – so we can use the SUM function (since this is a continuous range of cells) – and then divide by the total FTE number in J2.

`=SUM(K2:R2)/J2`

Hit enter and copy the formula down. You might need to "Format Cells" to change to a percentage.

Sort the data by this new column.

P13		Outline		
F	AG	AH	AI	AJ
ten	Pct Men	Diversity		
0	100.0%	<code>=sum(K2:r2)/J2</code>		
1	80.8%			
3	76.9%			
1	75.0%			
2.98	66.6%			
9	65.6%			

RANKING:

Sometimes we want to know where a school "ranks" on one of these measures that we've just calculated. The most accurate way to do that – especially taking ties into consideration – is to use the RANK function.

Here's how it works:

`=RANK(This Number, $Start Range:$End Range$, Order)`

This Number should be the cell where your data starts.

Start Range should be the cell where your data starts. Anchor with dollar signs.

End Range should be the last cell of your data. Anchor with dollar signs.

Order is either a 1 (smallest value will get assigned #1) or a 0 (largest value will get assigned #1).

Let's try it out in our data by ranking the diversity percentage we just created.

We'll put our formula in cell AI2.

AG	AH	AI	AJ	AK
t Men	Diversity	Diversity Rank		
0.0%	100.00%	<code>=RANK(AH2, \$ah\$2:\$ah\$675,0)</code>		
0.0%	83.33%			
33.3%	70.37%			
1.6%	64.52%			

The formula requires us to first tell it what

we want to rank – in this case the data in cell AH2 (the diversity percentage for the first school in our list). Then we need to tell it what to rank that against – we need to give it the full range of the column. In this case that is AH2 to AH675 (the last row in our data). Notice that there are anchors (\$) in front of the letter and the number of each cell address in that range. Then the third piece to the formula is whether we want it ranked in ascending order or descending order. We want the highest percentage ranked as #1, so we'll use the 0 for descending order.

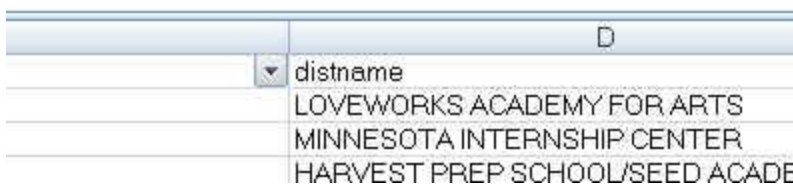
Copy the formula down and make sure it goes to the bottom of the page. Then notice the last records – all the ones with zero diversity. They are all ranked #419. This means they are tied. So why didn't we have any ties higher up in the list? The reason is that our calculation for the diversity percentage is carried out to many decimal places (that's how Excel operates, even if you're not DISPLAYING all those decimal places).

FILTERING:

Because this is such a large dataset, we might want to use Excel's filters to zero in on particular schools. Go to the Data menu and click on the Filter button.



Notice that now all the headers have little pull-down menus next to them.



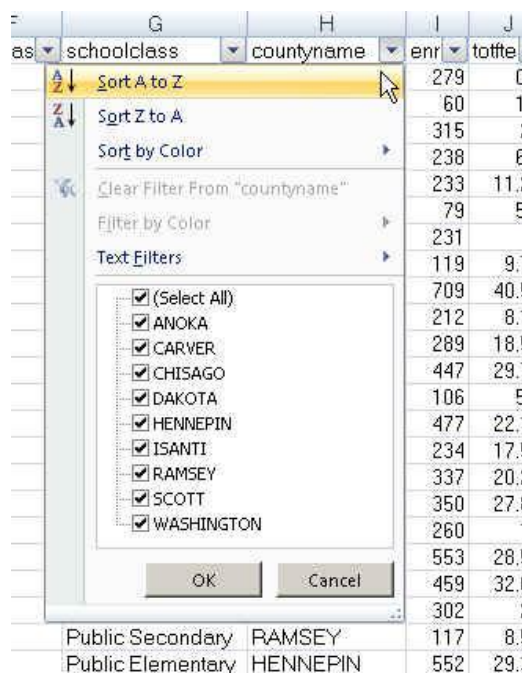
Go to the County column and click on the pull-down arrow. It will give you a list of all the values that exist in that column.

Notice the checkboxes in front of each county name. We can uncheck the "select all" option and then select just one – let's pick Ramsey.

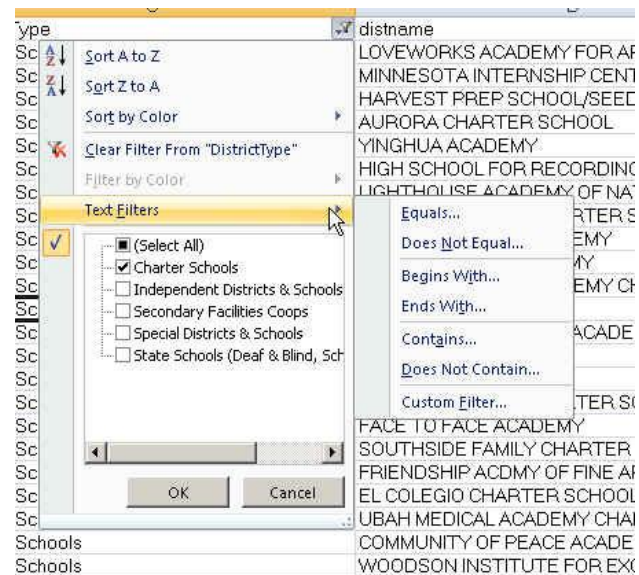
And the data will change so that only the Ramsey County schools are displayed. Notice that the numbers on the left-side of the screen are now in blue – this is how you can tell that you are viewing filtered data.

You can also choose more than one county – go back and add "Hennepin"

You can also filter on multiple columns at the same time. So let's say we only want to see charter schools in Ramsey and Hennepin. We've already got the county filter working. Now we just need to go to the DistrictType field (on the far left of the data) and choose "Charter Schools"



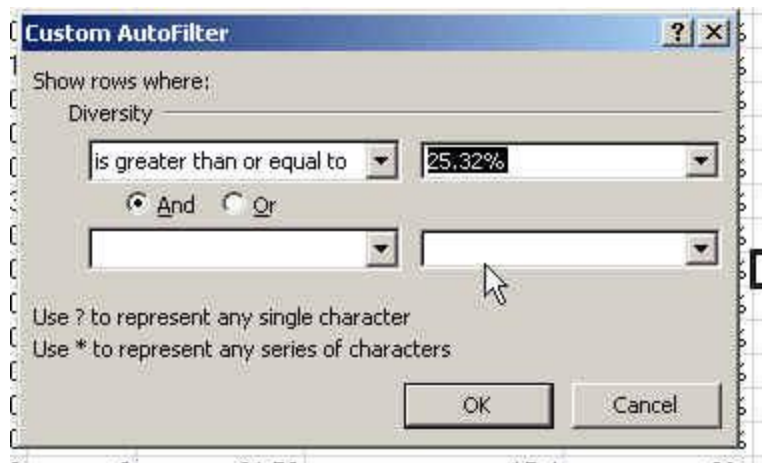
These additional options are great when you have dirty data and you need to look for variations in spellings – especially in situations where your list of options in the checkboxes are really long.



For example, we might want to limit our list of charter schools to only those with “high” diversity. How do we determine what’s high? That’s a bit of a judgment call, but usually you can find experts who will help you determine where to set the bar.

You can also eyeball the data and see where there's a “natural” break. In this case, the diversity of our charter schools starts at 100% and drops down to the 50% range really fast. The majority of schools are certainly below the 25% threshold.

The screenshot shows the 'Number Filters' task pane in Excel. The list of numbers to filter by includes (Select All), 0, 0.8, 0.94, 0.99, 1, 1.01, 1.17, 1.2, and 1.37. The '1.37' filter is currently selected. The background data is a PivotTable with columns labeled AC, AD, AE, AF, AG, AH, and AI. The 'AF' column is active, displaying numerical values. The 'AG' column displays corresponding percentages. Other visible values in the table include 14.2, 13.9, 20.4, 14.0 in the first column, and 22%, 35%, 40%, 28% in the second column.



Let's go to the Diversity Percentage field pull-down and choose "Number filters" and then choose "Greater than or equal to". It will bring up this dialog box where you can use the pulldown menu to find the number closest to the cut point. In this case, I chose 25.32% -- so we're going to filter so that we only see records where the diversity percentage is 25.32% or higher. (and don't forget, we've already filtered down to only

charter schools in Hennepin or Ramsey counties)

So now our list of schools is down to maybe a dozen.

Filtering is a good way to find examples in your data. So maybe you want to find all the elementary schools that have diversity greater than 25%. Or maybe you want to see all the schools where the majority of teachers are male. Or maybe you want to see the schools with student-teacher ratios below 15. The options are endless.

Created by MaryJo Webster
 St. Paul Pioneer Press
mwebster@pioneerpress.com
 January 2013