Analyzing school reports with Excel 2007

Sarah Cohen National Institute for Computer-Assisted Reporting sarah@nicar.org
Updated Feb. 2010 by MaryJo Webster, mwebster@pioneerpress.com

This uses an Excel file called "schlrpt.xls"

One of the most common questions posted on computer-assisted reporting lists on the Internet is, "How can I make sense of the annual school data?"

This exercise will give you a few tools to help analyze annual school reports data using Excel. It will show you analytic techniques that don't require an understanding of sophisticated statistics. Instead, it will give you a couple of simple tools that even your editor will understand.

We'll also go through some of the more powerful features of Excel at the same time. These include using named ranges, ranking without sorting, turning numbers into words you understand, and filtering to find special situations.

There's a big footnote here: The best stories come from reporting, not data analysis. This is especially true in school reports. They are so rich with data that the potential for stories is limited largely by the early reporting you do on issues in your area.

That said, here are a few techniques you can use to cut ultra-rich, and ultra-confusing, datasets like the annual school reports down to size for your readers and your editors.

Step 1: Interviewing your data

The first step in any data analysis is making sure you understand how what you have compares with what you want to know.

We're interested in three issues: How well some disadvantaged schools are overcoming their difficulties, whether the school districts allocate extra money to that effort, and if money matters.

Let's interview our data for hints about what how we might do this:
• Open **SchlRpt.xlsx**. You'll see a list of schools and 14 indicators from a typical school report database.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	SchoolID	Students	Mobility	Poverty	Whites	Blacks	Hispanics	ESOL	Money	Teachers	AdvDeg	10Years	Salary	Math	Read
2	ELEM001	641	14.1	88.5	1	614	26	1.6	$ 4,162	28	11.4	54.3	$ 33,373	17	6
3	ELEM002	898	7.8	14.6	830	27	37	3.2	$ 3,191	31	24.4	39.5	$ 30,480	56	53
4	ELEM003	985	6.5	17.6	755	164	45	0.9	$ 3,483	34	32.7	51.0	$ 33,925	82	70
5	ELEM004	1,087	22.8	44.6	554	304	170	3.9	$ 3,275	45	63.5	76.9	$ 39,636	72	47
6	ELEM005	671	10.6	29.8	431	206	17	0.1	$ 4,884	24	27.1	42.9	$ 32,879	71	61
7	ELEM006	879	13.8	85.5	1	725	153	10.8	$ 4,138	37	26.1	37.0	$ 32,882	23	12

This dataset has already been doctored and cleaned up for use in an exercise. So you needn't look through it, as you usually would, for unusual values, missing values or values that you don't understand.

But think through each of the columns. Based on your questions, you need three pieces of information: How much difficulty does a school face? How much does the school district help? And how well does each school perform?

Poverty, English as a second language, and mobility rates may all be slightly different measures of the difficulty each school faces. Spending per student, more highly educated teachers and smaller class sizes may all indicate attempts the school system makes to overcome these difficulties. Math and reading scores might measure the success each school has had.

There are many other analyses you *could* do. You might be interested in how integrated the schools are, how much is spent on primarily black or Hispanic schools compared with primarily white schools, and how experience or advanced degrees relate to teachers' salaries. But analyzing data on deadline means avoiding the temptation to stray from your reporting unless something obviously newsworthy jumps out at you.

There are more analyses you simply *can't* do with this dataset. You can't do any kind of analysis that focuses on the kids themselves rather than the average or typical kid in each school. You can't figure out the minority representation in each school.

There even are some things you probably don't *want* to do, even though you can. You could analyze test scores by the relative percent of black kids in each school. What could you possibly learn from that kind of an analysis alone?

Step 2: Setting up your spreadsheet

Before we go any further, let's do three things to make our lives easier later on.

• Lock the title row and column on the screen so you can scroll through your data without losing your place. Select cell B2. From the VIEW menu, choose Freeze Panes (and choose the first option, also called "Freeze Panes"). For future: when you want to unlock this feature, just come back to this same place and choosing it again will unlock it.

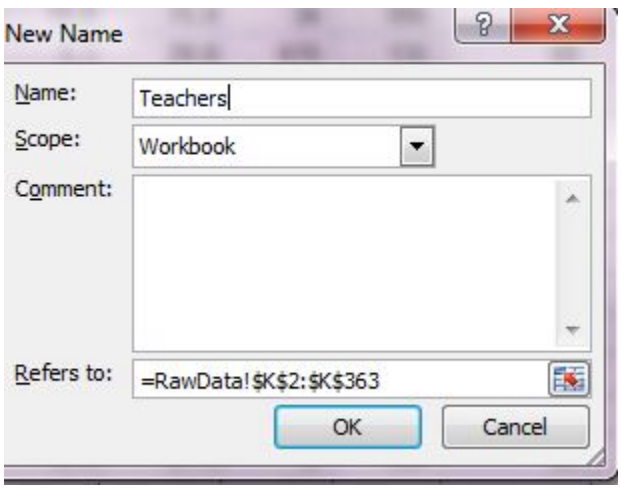
• Excel 2007 has already done part of another step for us. It has "named" some of the columns for us – using the column headers in our top row – so that we no longer need to refer to cell addresses. Instead of saying "O1" to refer to the Math column, we can simply say "math" in our formulas. To see what Excel has named for us, go to the Formulas menu and choose "Name Manager". A dialog box will come up listing all the named areas, like this:

As you can see, it only named 14 of our columns. The ones that appear to be missing are SchoolID and Teachers. We won't need SchoolID for formulas, but we might need Teachers, so let's add that one. Close the Name Manager box and highlight the teachers column from K2 to K263 (in other words, all of the data, but not the column label).

Then click on Name Manager again and push the "New" button.

It should look like this – note the name is Teachers (with a capital T) and the Refers to has the correct range of data.

If that's the case, push OK and you'll see Teachers added to the list of named ranges.



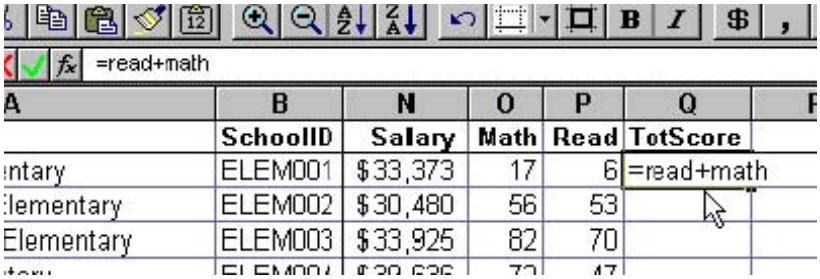
Step 3: Create indexes

In each of your categories of analysis, you have more than one measure. Many of them are legitimate and arguably important. One method of cutting your data down to size is to combine some indicators that measure similar underlying concepts.

Index 1: Performance scores

To come up with an index of the scores, we're just going to add together the median math and reading scores. That's because they're comparable and we want to use the actual test scores later on.

• Type TotScore in cell Q1, press Enter, select Q2 if it isn't already and type =read+math. Press Enter, and copy the formula down through your database:



(Hint: Type functions and names of columns using all lower case when typing formulas. That way, you can easily trace a mistake in your formula because it won't have converted to upper and lower case – to match the actual whey they appear in your column header — after you hit Enter.)

Index 2: Difficulty scores

Now we're going to put together a simple index of the difficulty that each school faces. Let's say your reporting revealed that experts think poverty rates and mobility rates (the percent of students who change schools during the year) are very important indicators of how disadvantaged kids in each school are. Your sources also say proficiency in English is important, but way less important than the other two factors.

In this data, we have the percent of students in each school who have these problems. We're going combine these indicators in one of the simplest ways possible: Rank the values in each field, and average the ranks into a simple index.

Here are four advantages to using this kind of an indexing system:

1. It doesn't require your data to begin at similar levels. Say you suspected that the size of the school had something to do with difficulty. How could you combine that level, which is measured in students, with the percentage of students taking English? When you rank data,

it doesn’t matter that the unit of measure is different.

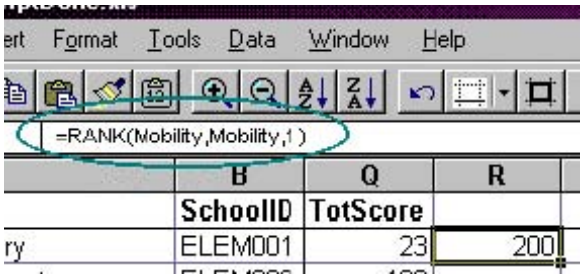
2. It’s not very sensitive to the actual values, especially very high or very low ones. This comes in handy when you’re making an index of, say, company performance or anything with income or home prices.

3. It’s easy to do and easy to explain. You could get a much more sophisticated version of this analysis by using regression, factor analysis or a comparison of z-scores. But it would be more difficult to explain, and often won’t tell you a lot more about your data.

4. It’s easy to control the direction of the analysis. If you had the percent of kids in gifted and talented programs, a high number would be an advantage, not a disadvantage, to a school. Ranking lets you control which way you want a value ranked: in ascending order or descending order.

Excel has a function called RANK(). It requires three pieces of information: The value you want ranked, the list of values you want it ranked within, and the order of the ranking. Go ahead and try a rank now, before we combine a bunch together.

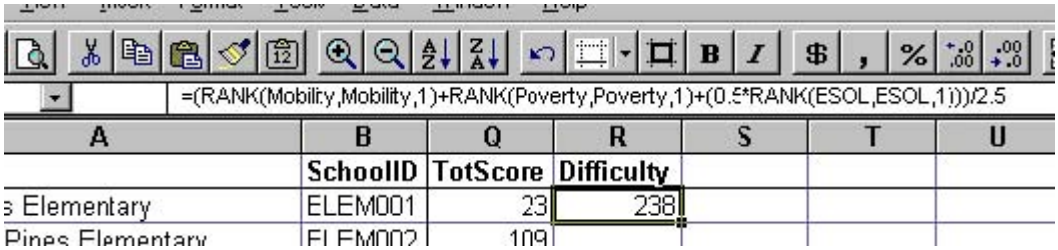
• Select R2, and type =rank(mobility, mobility, 1) and press Enter.



This means that the mobility rate, which you named earlier as the values in column D, is the 200th-lowest in the group. The first "Mobility" in the formula refers to this row’s mobility rate, or cell D2. The second "Mobility" in the formula refers to the whole column of mobility rates, or D2:363.

The "1" says that you want the reverse of a typical ranking: you want to rank the *lowest* value as No. 1, and the *highest* as No. 363. That’s because we want high index values for schools that have high difficulty scores. (If you left out the 1, and ranked it normally, the highest mobility rate would have the lowest rank value, because it would be No. 1.)

Now we’ll combine bunch of ranks in the same formula.
• Select R1, type Difficulty, and press Enter.
• Select R2. Ignore what you just typed in and typ this right over it:
=(rank(mobility,mobility,1)+rank(poverty,poverty,1)+(.5*rank(esol,esol,1)))/2.5 and hit enter.



Aside from the rank portions of this formula, we're weighting the index by multiplying the ESOL rank by 0.5. That's because our discussions with experts indicated they believed English language skills were far less important than the other two problems facing schools. We've chosen to give it half the importance of the other two. No, this isn't scientific. Weights are often somewhat arbitrary.

We've divided the sum of three rankings by 2.5 to compute an average rank. That's because our mobility rank is implicitly multiplied by 1, our poverty rank is implicitly multiplied by 1, and our ESOL rank is explicitly multiplied by 0.5, for a sum of 2.5.

This technique is a simple way to combine lots of interrelated values that each tell you something slightly different. It is a weighted average rank of values, using arbitrary ranks you decide in advance. There are two keys to doing this well:

1. Choose indicators that tell you something different from one another, even if it's only slightly different. We can each imagine cases of high mobility that aren't associated with poverty; we can also imagine lots of cases in which students are taking English, but also aren't poor.
2. Try to limit the number of indicators to three or four. You probably won't get much new information out of each field you add after about four. Besides, it would then become hard to tell what's affecting the index if it comes out strange.

You now have two values that you want to compare: TotScore, or the sum of the test scores, and Difficulty, or the index of the difficulty that each school faces. Let's hold off on that until we calculate our third indicator.

Index 3: Resources

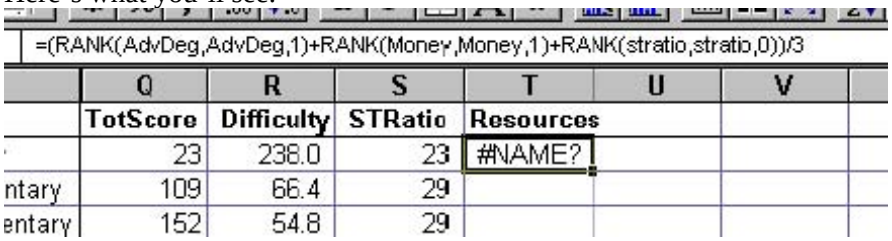
We're going to combine the ranks of the spending per student, the teacher-student ratio, and the percent of teachers with advanced degrees as our measure of the resources the school system throws at each school.

In this case, you have three indicators that begin with different units of measure: dollars, which are about $3,000 each; number of students per teacher, or a figure between about 15 and 20; and percent of teachers with advanced degrees, or a number between 0 and 100. Our ranking technique ignores these different levels and allows you to compare them directly.

• First we have to compute the student-teacher ratio. Select S1, type STRatio and press Enter. Select S2, and enter the formula =students/teachers and press Enter. Copy it down through your database.

Now we can calculate our Resources score.
• Select T1, type Resources and hit Enter.
• Select T2. Type the formula
=(rank(money,money,1)+rank(advdeg,advdeg,1)+rank(stratio,stratio,0))/3 and press Enter.

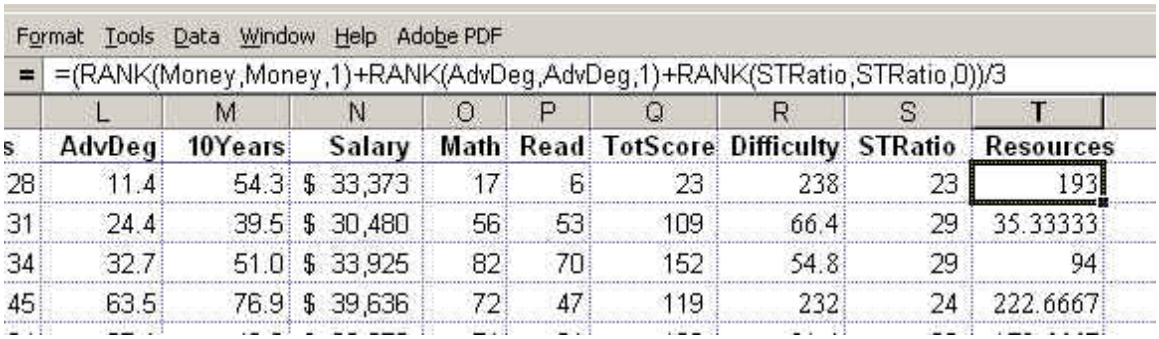
Here's what you'll see.



What happened? The error, #NAME?, shows you that the spreadsheet doesn't understand part of your formula. Moreover, the this specific error tells you that Excel is expecting a function (like AVERAGE()) or named range (like Mobility) and hasn't found it. Notice how the "STRatio" in the formula hasn't been converted to upper and lower case the way the others have. That's because we haven't named our new column yet, so it doesn't recognize the column name. Let's go ahead and rename everything, so we don't have to worry about it in the future.

• Repeat the Name Manager steps (as we did for Teachers at the beginning of the exercise) for the three columns we just created – TotScore, Difficulty and STRatio

Now it should look like this instead of the error:



There's one other big difference in this formula from difficulty index: We've reversed the order that we think matters in the student-teacher ratio. Notice the zero instead of a 1 in the last part of that rank command: RANK(STRatio,STRatio,0). It costs more to have a low student-teacher ratio (requiring more teachers) than a high one. The zero in that part of the formula means we want a high value to be associated with a low student-teacher ratio.

We're dividing by 3 because we haven't weighted this formula, and there are three variables in it.

• Select T2 and copy the formula down through your database.

HIDE COLUMNS:

• To make the rest of our work a little easier, let's hide all of the columns we won't use anymore. Highlight columns C through P by clicking on the letters at the top and dragging across until all are highlighted. Right-mouse click and choose "Hide". Then highlight the S column and right-mouse click and select "hide".

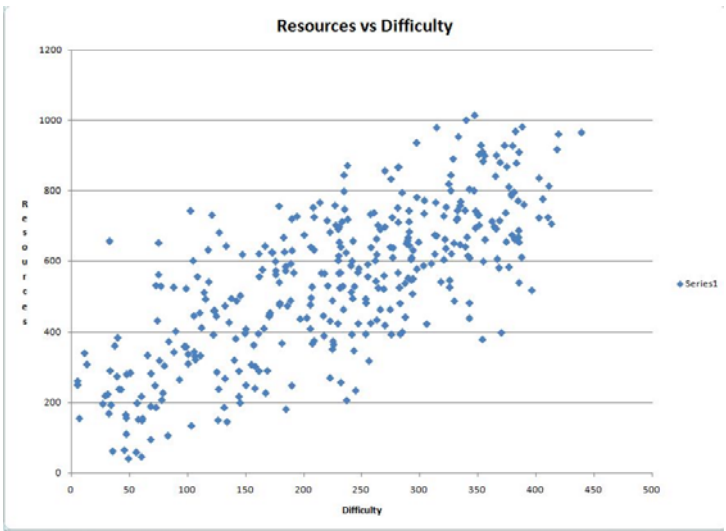
PASTE SPECIAL
Then select columns Q through T. Copy. Choose Edit, Paste Special, and check the box next to Values. This converts the formulas into the values they calculated.

Scatter plots

We're interested in three issues: Whether the school system is putting more resources into schools that face difficulty; how well the resources pay off in performance, and how performance is associated with difficulty.

The first step is to make some scatter charts testing each of these premises. You may have used these before, but they're so fundamental to any data analysis that we'll repeat it here. We'll only make one scatter plot here, but you should try this on your own for the other two.

- Select R1:T363 (since the S column is hidden, this means that you'll only have columns for "Difficulty" and "Resources" in your chart). After the area is highlighted, go to the Find & Select button on the HOME menu (far right side) and select "Go to…" and then push the "Special button" in the dialog box that comes up. A new dialog box comes up; from that choose "Visible cells only"
- From the INSERT menu, choose "Scatter". A scatter plot chart will appear within your worksheet. Let's first move it to its own worksheet. To do this, right-mouse click in the far right area of the chart (where there are no data) and a menu will come up that includes "Move chart." It will allow you to choose "New worksheet" and to name the new worksheet (the default will probably be "Chart 1", so let's leave that as the name)
- It will automatically take you to that new sheet. Go to the LAYOUT menu and the first thing we'll do is put labels on the X (horizontal) and Y (vertical) axis. So choose "Axis Titles" and the first option will be "primary horizontal axis title" and you can select "title below axis" and it will allow you to type a label under the X axis. Label this "**Difficulty**"
- Go back to Axis Titles and this time choose the "primary vertical axis title" and choose "Vertical Title." Again it will allow you to type a label. For this one, put "**Resources**"
- Now let's add an overall chart title. While still in Layout menu, choose "Chart Title" and choose "Above Chart." Label this one "Resources vs Difficulty"



Now your chart should look like this:

You can imagine a straight line going through the middle of the dots, indicating that more resources are generally given to schools that face more difficulty. (Note the order of that statement. It's not always that clear. But it this case, it's pretty inconceivable that poorer, migrant or non-English speaking kids gravitate toward schools that throw more resources at educating them.)

The next thing you'll see is that there are no schools that obviously stand out as very low-difficulty and very high-spending, or very high-difficulty and very low-spending. In other words, there are no dots in the extreme upper left or lower right corners of the graph.

You can get back to your original data by click on the tab in the lower left corner that says "Raw Data".

Picking out extremes

You'd still like to see which schools are generally out of synch with the pattern. So we're going to use the database capabilities of Excel that mirror queries in database programs to do this.

Instead of using the simple filters you learned in earlier exercises, we'll use something called Advanced Filters. These are powerful, approaching the capabilities of a database manager. By using computed criteria for your filter, you can compare values in different columns or across records.

• With the interim columns still hidden, select Q1:T1 (this will just be the headings for TotScore, Difficulty and Resources columns). Go back to the Find & Select button and choose "Go To" and hit the Special button and then choose "visible cells only". Copy. Put your cursor in cell V1 and Paste. (yes you will be leaving one blank column in between)

Q	R	T	U	V	W	X
TotScore	Difficulty	Resources		TotScore	Difficulty	Resources
23	238	193				
109	66.4	35				
152	54.8	94				
119	232	223				
132	91.4	171				
35	248	198				

This is an area that you're going to use to fill in criteria for a query. We'd like to see any school with much higher-than-average spending but much lower-than-average difficulty because we suspect that some rich parents have managed to get more money out of the school board.

We're going to use a function called "Quartile" to figure out which schools are in the top 25% of the schools, by difficulty. This tells you the top value of each cutoff point that splits the values into four categories with an equal number of values in each. For example, the 2nd quartile would be the about the same as a median, at which half the values are above the level and half are below: The QUARTILE() function requires two pieces of information: The range of cells that you want to check, and the quartile cutoff point you want back. We want the 3rd quartile.

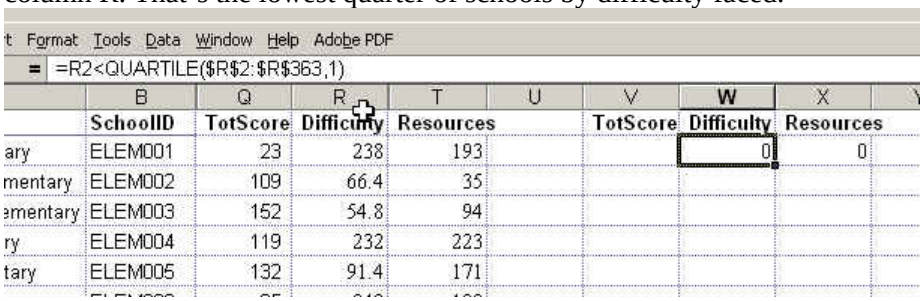
• Select cell X2 (under the Resources heading you just pasted). Type this formula: =T2>quartile($T$2:$T$363,3)

(You can't use the names you created earlier in this exercise. It's weird. But it doesn't work.)

This formula tells Excel to look at cell T2, and compare it to the top value in the third quartile of Resources. The dollar signs are needed because they'll later tell Excel to compare each row within column T to the same set of rows, row 2 through 363. After you hit Enter, the formula will show a zero, which means that the comparison came out False for the first line. If it were true, it would be a 1.

• Now select W2 (under the Difficulty heading). Type this formula: =R2<quartile($R$2:$R$363,1) and hit Enter.

This tells Excel to look at cell R2, and compare it to the cutoff point of the lowest quartile in column R. That's the lowest quarter of schools by difficulty faced.

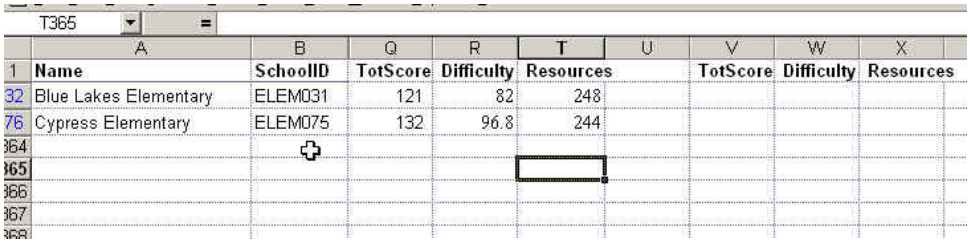


	B	Q	R	T	U	V	W	X
	SchoolID	TotScore	Difficulty	Resources		TotScore	Difficulty	Resources
ary	ELEM001	23	238	193			0	0
mentary	ELEM002	109	66.4	35				
ementary	ELEM003	152	54.8	94				
ry	ELEM004	119	232	223				
tary	ELEM005	132	91.4	171				

• Put your cursor anywhere inside your chunk of data (just so it's not in a blank area) and go to the Data menu and choose "Advanced Filter" The List Range (which refers to your database) will likely need to be revised so that it says: **$A$1:$T$363**

You have to tell Excel where to find your criteria. Select the Criteria Range section in the dialog box, and either type or select with your mouse the area $V$1:$X$2.
Press OK

• You should end up with two rows, representing two schools that get a lot of money from the school system in the form of lots of teachers, high spending and more educated teachers, but face less poverty, mobility and English language problems than most.



	A	B	Q	R	T	U	V	W	X
1	Name	SchoolID	TotScore	Difficulty	Resources		TotScore	Difficulty	Resources
32	Blue Lakes Elementary	ELEM031	121	82	248				
76	Cypress Elementary	ELEM075	132	96.8	244				
364									
365									
366									
367									

There may be good reasons for this. In fact, the school system may not have much control over how much is spent on each student. But you've gotten two schools to focus on during your reporting.

• Get rid of the filter now by going back to the Data menu and pushing the "Clear" button that is located next to the Filter button.

You may wonder at this point, "Why not just use a database manager?" for this kind of analysis. There are a couple of reasons, but the most compelling is the difficulty of ranking lots of fields and working with quartiles and percentiles in a database manager.

Turning data into words

Think about what you've done so far. You've established a very strong relationship between the difficulty a school faces and the resources the school district puts into it. You've determined that there are only two schools that buck that trend. But you still have a problem: Your editor, not to mention your readers, may have trouble understanding what the numbers in your analysis mean. So we're going to turn your values into words: High, medium and low.

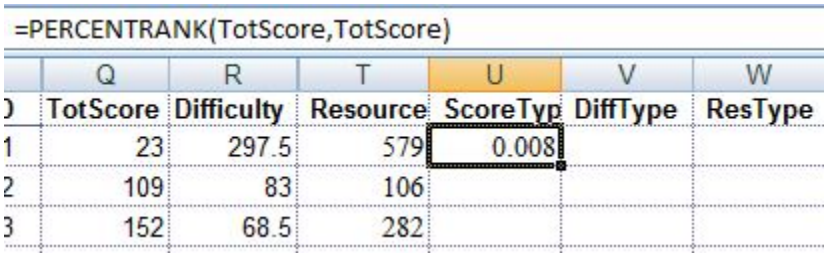
There are a few ways to do this. We're going to use one that's kind of like our ranking system, using a formula called "PERCENTRANK()" combined with a lookup table.

• Select columns U through W, and insert 3 columns. Name the columns ScoreType, DiffType, ResType. (Make sure to type the labels with upper and lower case as shown here)

First, you'll see how the PERCENTRANK() function works. Then we'll change the formula to look up the words. PERCENTRANK() is the only built-in function in Excel that allows you to work with the relative position of a value within a group of numbers. It gives you the equivalent of the percentile figure you saw on your SAT scores.

Other functions, like QUARTILE(), tell you the cutoff points in a dataset; they don't tell you into which group each value falls. So we'll use PERCENTRANK()to figure out a position, between 0 and 1, of each value in the column.

• Select U2. Type the formula =PERCENTRANK(totscore,totscore) and press Enter. (no need to copy down just yet)



	Q	R	T	U	V	W
)	TotScore	Difficulty	Resource	ScoreTyp	DiffType	ResType
1	23	297.5	579	0.008		
2	109	83	106			
3	152	68.5	282			

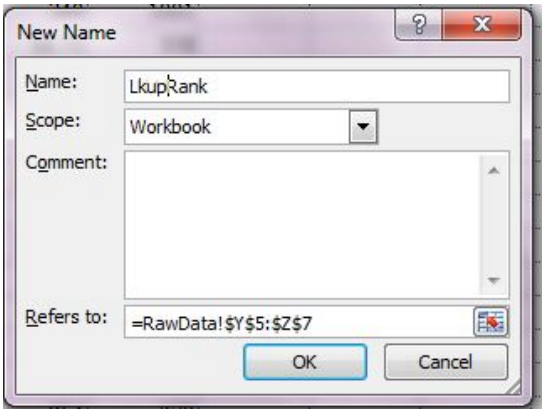
The resulting value 0.0 (when rounded) means that the total score of 23 for this line puts the school in the lowest percentile, or 0. The first TotScore in the formula refers to the value for this row in the column TotScore. The second TotScore in the formula refers to entire column of values.

• Now move to a clean area of your worksheet, to **cell Y4**. Here's where we're going to set up the information Excel needs to associate words with various levels of percentiles. Type two columns that look like this:

	Y	Z	AA
	TotScore	Difficulty	Reso
		0	
	PctRank	Word	
	0	Low	
	0.333	Medium	
	0.667	High	

• Select the area **Y5:Z7**. Go to the FORMULAS menu and choose "Define name." In the dialog box that comes up change the name to LkupRank

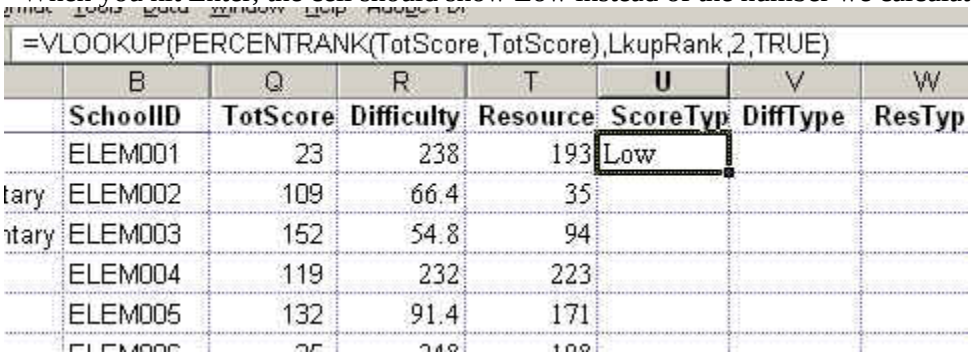
From now on, any time you want to use this area on the worksheet, you don't have to remember where it is. Just refer to its name, LkupRank.



• Go back to your formula in U2. Press F2 to edit it, and move to the beginning of the formula. Change the formula to look like this: =vlookup(percentrank(totscore,totscore),lkuprank,2,true)

VLOOKUP() checks the value of the PERCENTRANK() result against the table (lkuprank) you just made. If it's under .333, then it gives you back the word "Low," which is in the second column of that area. If it's between .333 and .666, it gives you back the word "Medium." If it's .667 or higher, it gives you back the word "High". The "true" in the formula means that the LkupRank values are sorted, which is required when you don't want to type in every possible value but want to check ranges of values instead.

• When you hit Enter, the cell should show Low instead of the number we calculated earlier.



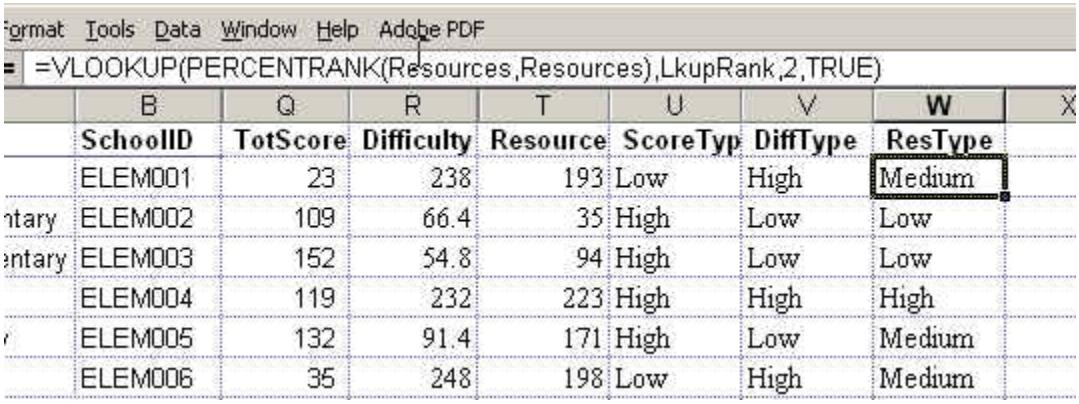
	B	Q	R	T	U	V	W
	SchoolID	TotScore	Difficulty	Resource	ScoreTyp	DiffType	ResTyp
	ELEM001	23	238	193	Low		
tary	ELEM002	109	66.4	35			
ntary	ELEM003	152	54.8	94			
	ELEM004	119	232	223			
	ELEM005	132	91.4	171			

• Select U2 and copy it across to the other two columns. Don't copy down yet. Now edit the remaining two columns, replacing TotScore with Difficulty in the first and with Resources in the second.

=VLOOKUP(PERCENTRANK(Difficulty,Difficulty),LkupRank,2,TRUE)

=VLOOKUP(PERCENTRANK(Resources,Resources),LkupRank,2,TRUE)

If you get an error message on any of them, make sure that your columns Difficulty and Resources have been named by opening up the Name Manager. Add them if necessary as we did earlier with Teachers.



	B	Q	R	T	U	V	W
	SchoolID	TotScore	Difficulty	Resource	ScoreTyp	DiffType	ResType
	ELEM001	23	238	193	Low	High	Medium
ntary	ELEM002	109	66.4	35	High	Low	Low
entary	ELEM003	152	54.8	94	High	Low	Low
	ELEM004	119	232	223	High	High	High
	ELEM005	132	91.4	171	High	Low	Medium
	ELEM006	35	248	198	Low	High	Medium

• Now select U2:W2 and copy them down through your database (this allows you to copy all three at the same time)

You now have words that your editor will understand. Although they're not precise measures of the quality of school, the resources thrown into the school or the difficulty a school faces, you have an excellent general idea of how they stack up in each of these three areas. Your simple measure now encompasses 8 different variables. And you don't have to look at numbers anymore to understand how they fit.

Making crosstabs

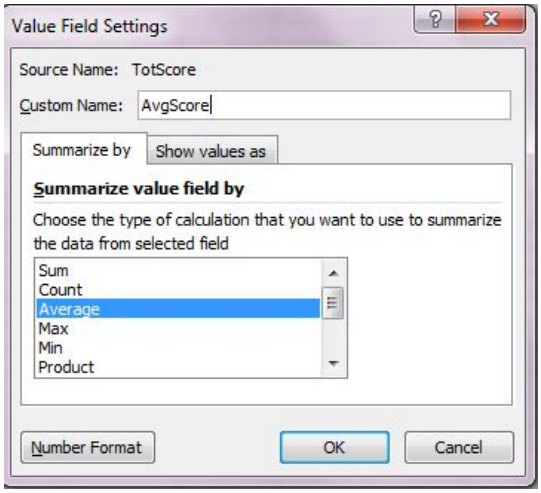
We're going to make two crosstabs that should help you understand the relationship between test scores, spending and difficulty.

First, remember how we measured the test scores. Instead of using our ranking system, we simply added up the scores. That means the average score still has meaning. So our first crosstab will compare test scores by the spending and difficulty categories we figured out in the last step.

• Put your cursor anywhere in the database. Go to the INSERT menu and choose "Pivot Table." In the first step make sure the correct data range is shown and that "New worksheet" is selected. Then push OK.

• It will bring up the Pivot Table designer window.

•Drag the Diff Type box to the ROW LABELS area. Drag the ResType box to the COLUMN LABELS area. Now drag the TotScore box to VALUES area. It will change to "sum of TotScore". Click on it and choose "Value Field settings" and in the dialog box that comes up change it to Average (instead of Sum) You can also change the name to something shorter, like Avg Score



• Press Number Format button (in lower left), and choose Number format instead of General in the format box. This will limit the number of decimal places you'll see in your pivot table. Press OK, then OK again.

•Now drag the box SchoolID to the VALUES area. The title will change to Count of SchoolID. That's fine. This will tell us how many schools make up each average score.

• The last adjustment we'll make to the pivot table is to change the order of the categories. Excel has assumed we want them in alphabetical order. Instead, we want them to go from Low to Medium to High.

Right now, it's showing High then Low, then Medium. So we need to move High across to the right. To do that, select B4:C9. This will take both the average score and the count of schools for the High category.

Get your cursor along the top edge of the highlighted area and it will change to a four-way cross. Then drag until you can drop it into column H.

You can do the same with the rows. Right now the rows go High, Low, Medium. Highlight the entire "high" row (row 6) and get the four-way cross cursor and drag it down until it falls below Medium.

You can also edit your pivot table a bit so you can understand it better. Right now it says "Row Labels" and "Column Labels". Click on Row Labels and change it to "Difficulty" and click on Column Labels and change it to "Resources"

B8 79.1666666666667

	A	B	C	D	E	F	G	H	I
1									
2									
3		Resources							
4		Low		Medium		High		Total Avg Score	Total Count of SchoolID
5	Diffculty	Avg Score	Count of SchoolID	Avg Score	Count of SchoolID	Avg Score	Count of SchoolID		
6	Low	124.4	86	120.0	31	133.4	5	123.6	122
7	Medium	92.0	29	82.7	55	86.9	35	86.2	119
8	High	79.2	6	59.2	34	60.5	81	61.1	121
9	**Grand Total**	**114.4**	**121**	**85.7**	**120**	**71.2**	**121**	**90.4**	**362**
10									
11									

Now you have average test scores by difficulty and resources.

What's the first thing that jumps out at you? Check the values in the corners of the crosstab. Although schools that face a lot of difficulty also get a lot of resources, their test scores don't seem to improve with them. And although putting resources into schools with less difficulty may improve the scores a little, they're remarkably high to begin with.

Look also at the average scores down the right side and across the bottom. (The counts are meaningless here: our method of splitting up the schools into Low, Medium and High categories guarantees they will be about equal.) The scores consistently move down as you increase difficulty AND as you increase resources.

How would you interpret this? One possibility is that the disadvantages faced overwhelm any spending by the schools to overcome them. Another is that the spending isn't effective. A third is that perhaps the spending is detrimental! This is a good starting point for your reporting on where the money is going, which teachers are in difficult schools and how big programs are working.

SECOND PIVOT TABLE

In the second pivot table, we'll look at how the high, medium and low-scoring schools are distributed instead of average scores.

• See if you can make a pivot table like this on your own. To do this, use the tabs in the lower left corner to go back to your original data – the "RawData" sheet. It has **Diff Type** across the top and **Score Type** down the left. Its data element is the count of SchoolID, which is renamed as #.

	A	B	C	D	E
1					
2					
3	Count of SchoolID	Difficulty			
4	Scores	Low	Medium	High	Grand Total
5	Low	3	35	83	121
6	Medium	32	59	33	124
7	High	87	25	5	117
8	**Grand Total**	**122**	**119**	**121**	**362**
9					

This by itself tells us something pretty interesting: There are five schools that are scoring quite well even though they face a lot of difficulty. We'll go back to this later, and use a filter to discover which schools they are.

For now, though, we want to know a little more precisely how the pattern works. Most analysts use something called "row" and "column" percents in a crosstab to pick out patterns. First, you identify which event might depend on the other. In our case, scores probably depend on difficulty; difficulty probably doesn't depend on the scores. (One hint to choosing a dependency is to look at which comes first. The difficulty predated the scores, so it would be impossible for the difficulties to depend on something that hasn't happened yet.)

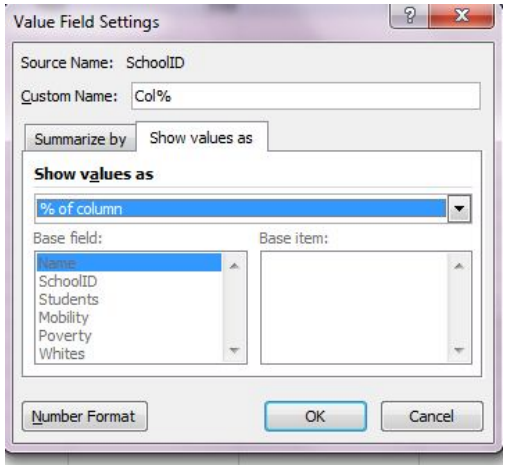
So Diff Type is what we'll call an "independent" variable. Scores depend on them, so we'll call scores "dependent." You can have lots of independent variables, including our resources type if we wanted. You only identify one dependent variable at a time.

So we want to know whether the percentage of schools that score in each category are different by the level of difficulty they face. A rule of thumb in crosstabs is, set them up so that the independent variable is shown across the top, and the dependent down the side. Then use column percents. You can compare them by reading across the rows.

Go back to your Pivot Table designer (if you have your cursor somewhere in the pivot table it will show the designer on the right side of your screen)

• Drag another version of the SchoolID box into the VALUES area. Its title will change. click it and choose "value field settings". In the title area, change its name to Col %.
Select the "show values as" tab in the dialog box and select the one that says "% of column."

Select OK to return to your pivot table.



Now you can just look across each row to see how consistent the pattern is.

	A	B	C	D	E	F	G	H	I
1									
2									
3		Difficulty							
4		Low		Medium		High		Total Count of SchoolID	Total Col%
5	Scores	Count of SchoolID	Col%	Count of SchoolID	Col%	Count of SchoolID	Col%		
6	Low	3	2.46%	35	29.41%	83	68.60%	121	33.43%
7	Medium	32	26.23%	59	49.58%	33	27.27%	124	34.25%
8	High	87	71.31%	25	21.01%	5	4.13%	117	32.32%
9	**Grand Total**	**122**	**100.00%**	**119**	**100.00%**	**121**	**100.00%**	**362**	**100.00%**
10									
11									

This method gives us the ability to say, for example, that only 69 percent or two-thirds of the high-difficulty schools had low test scores, but only 2 percent of low-difficulty schools did. It's a powerful yet simple way to illustrate what educators often say – that you have to take into account the barriers to learning before you tar a school for underperforming.

Our last step is to go back to our database, and identify the five high-difficulty schools that score well.

• Click the RawData tab at the bottom of your spreadsheet. Select a cell anywhere in your database and choose FILTER from the Data menu. Little arrows will appear at the top of each column.

• We want to select the schools with high test scores and high difficulty. So in the column marked **Diff Type**, press the arrow button and choose High. Do the same thing on the **Score Type** column You should be looking at five lines that look like this:

	A	B	Q	R	T	U	V	W
	Name	SchoolID	TotSco	Difficul	Resour	ScoreT	DiffTyp	ResTyp
	Air Base Elementary	ELEM004	119	290	668	High	High	High
2	Belvedere Elementary	ELEM021	113	354	379	High	High	Low
0	Greynolds Park Elementary	ELEM119	111	291	684	High	High	High
9	Roosevelt Elementary	ELEM288	127	294.25	632	High	High	Medium
7	West Gate Elementary	ELEM346	113	328.75	891	High	High	High
4								
5								
6								

They might be magnet schools, in which case the test scores might not be that meaningful as an average. Or they might be discouraging the worst-performing kids from coming to school the day of the tests. That hardly suggests success. But they may just have interesting principals or experiments you might have heard of elsewhere – people or policies that seem to be making a difference.

TO COMPLETE ASSIGNMENT:

1) Submit your Excel file showing your work

2) Write a brief summary of your findings to your editor, including suggestions for how to proceed in the reporting on this story. Be sure to explain your findings in as simple a way as possible, without bogging it down with numbers.