Rates and ratios in Excel

RATES

For journalists, learning how to compute rates is vital. That's because rates help us get closer to the truth and may even help us uncover stories that otherwise might get buried.

Here's an example, some journalists have called Chicago the most dangerous city in the United States because FBI Uniform Crime Report data say it had the greatest number of murders of any U.S. metropolitan area. One news account even called it the "murder capital" in a blog post headline. Other big cities, such as New York and Los Angeles have spots near Chicago at the top of the list.

However, focusing on those big numbers is misleading because it fails to account for population. Chicago, L.A. and New York also have the three largest metro area populations. We'd expect them to have the greatest number of homicides, or most anything else. It's more newsworthy when a small city has a high number of murders, or a big one has a small number.

When journalists took population into account, they found that Flint, Mich., was the most dangerous metro area in the United States in 2012, when it came to murders. Detroit came in second place.

In another example, The New York Times reported that that small towns like Rehoboth Beach, Del., had become gay enclaves. The Times' report relied on 2010 census data that went beyond raw numbers and took population into account.

CALCULATING RATES

To calculate a rate we take our number and divide it by the population. This gives us the per capita or per person rate. (In Latin, per capita means per head.)

Unfortunately, per capita numbers are often fractions. When we calculate them in spreadsheets we will often get numbers with many decimal places. So, we usually then multiply the per capita rate by a standard number to generate a number that's more meaningful, that reflects a defined number in the population. The FBI uses 100,000 as its standard multiplier in the Uniform Crime Reports. So we will see crime rates per 100,000 people. Likewise, the incidence and prevalence of diseases is often reported per 100,000 people.

If we are unsure what the multiplier should be, there are a few ways we might be able to dig up that information. First, contact the agency that created the data. Second, look for academic or government studies using the data. If those avenues fail, we could create our own multiplier by

examining the populations and picking an increment that makes sense. For example, with school enrollment data whose populations is in the hundreds pick 100.

Now we are ready to calculate rates. Make a copy of the CA_discipline.xlsx spreadsheet and open it. The spreadsheet contains information about California public school discipline summarized by county for the 2012-2013 school year. You can download the original file from the California Department of Education at http://bit.ly/WipqLT.
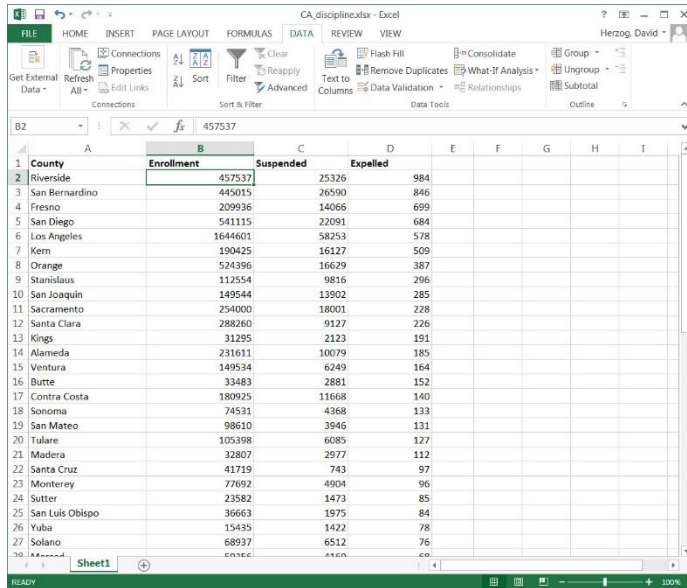
Note that we have four columns in the spreadsheet. Column A lists county, B the enrollment, C the number of students suspended and D the number expelled.

Before we calculate rates, let's sort our sheet to answer two questions: Which county had the greatest number of students who were suspended? Which county had the greatest number of

students who were expelled?

Select the data table as we did in the last lesson, then sort from largest to smallest and we see that Los Angeles County public schools suspended the greatest number of students at 58,298. Not a big surprise, because Los Angeles is also the most populous county in California. In fact, we see that its enrollment of more than 1.6 million students happens to be the largest in the state.

	A	B	C	D
1	County	Enrollment	Suspended	Expelled
2	Los Angeles	1644601	58253	578
3	San Bernardino	445015	26590	846
4	Riverside	457537	25326	984
5	San Diego	541115	22091	684
6	Sacramento	254000	18001	228
7	Orange	524396	16629	387
8	Kern	190425	16127	509
9	Fresno	209936	14066	699
10	San Joaquin	149544	13902	285
11	Contra Costa	180925	11668	140
12	Alameda	231611	10079	185
13	Stanislaus	112554	9816	296
14	Santa Clara	288260	9127	226
15	Solano	68937	6512	76
16	Ventura	149534	6249	164
17	Tulare	105398	6085	127
18	Monterey	77692	4904	96
19	Sonoma	74531	4368	133
20	Merced	59356	4160	68
21	San Mateo	98610	3946	131

	A	B	C	D
1	County	Enrollment	Suspended	Expelled
2	Riverside	457537	25326	984
3	San Bernardino	445015	26590	846
4	Fresno	209936	14066	699
5	San Diego	541115	22091	684
6	Los Angeles	1644601	58253	578
7	Kern	190425	16127	509
8	Orange	524396	16629	387
9	Stanislaus	112554	9816	296
10	San Joaquin	149544	13902	285
11	Sacramento	254000	18001	228
12	Santa Clara	288260	9127	226
13	Kings	31295	2123	191

Our descending sort of the number of students expelled shows us that Riverside County led the way with 984. That's surprising because Riverside County's enrollment is just a little more than one-quarter the size of Los Angeles'.

Los Angeles is in fifth place with 578 expulsions.

Now we are going to calculate suspension and expulsion rates for each county. We will first do that by taking baby steps: Per capita first, then per 1,000 students.

In cell E1 type a label: "Suspend_percap". Type "Suspend_Per1k" in F1.

To calculate the per capita suspension rate for Riverside County, we will divide its number of suspensions by its number of enrolled students. In cell E2 type "=C2/B2" and hit Enter. The result tells us there are .0553529 expulsions per person. Copy that down for all of the counties and we see similar results that are meaningless.

Now we will calculate per 1,000 students in the next column by multiplying. In cell F2 for Riverside County type "=E2*1000" and hit Enter. We see the county's suspension rate is 55.35290042 for every 1,000 students.
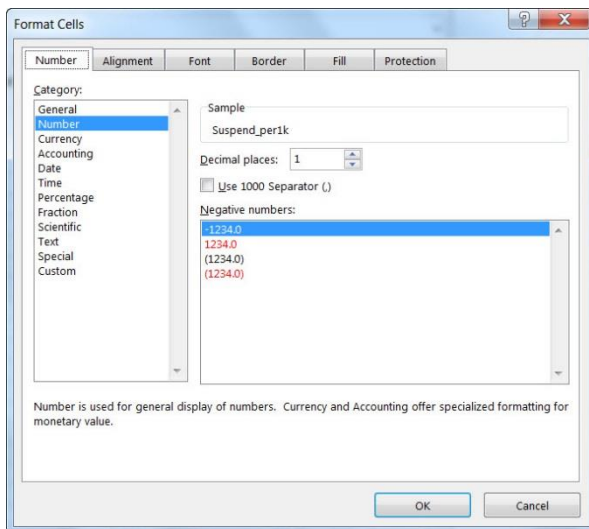
Copy the formula down for the other counties. Format the contents of the F column by right clicking on the letter F and picking Format cells. Under Category pick Number, then set the decimal places to 1.

We see that some counties with smallish enrollment numbers are at the top of the list. Del Norte leads the way with nearly 125 suspensions for every 1,000 students. Alpine comes in second with nearly 105. But it also has a small enrollment. Nonetheless, we would want to look into what's going on in these counties. And Los Angeles? It's in the 50th spot with more than 35 suspensions for every 1,000 students.

Now that we've learned how to calculate the per 1,000 student rate in two steps, we're going to combine them into one formula for the expulsion per 1,000 student rate.

Type "Expelled_per1k" in cell G1 for our label. In G2, enter this formula, which calculates the per capita expulsion rate and then multiplies that by 1,000. " =D2/B2*1000" . Excel shows Del Norte's rate as 8.668593021. Copy the formula down for all of the counties and format the G column as Number with 1 decimal place.

	A	B	C	D	E	F	G
1	County	Enrollment	Suspended	Expelled	Suspend_percap	Suspend_per1k	Expelled_per1k
2	Mariposa	2087	211	22	0.10110206	101.1	10.5
3	Del Norte	4499	560	39	0.124472105	124.5	8.7
4	Kings	31295	2123	191	0.067838313	67.8	6.1
5	Lake	9951	1016	51	0.102100291	102.1	5.1
6	Yuba	15435	1422	78	0.09212828	92.1	5.1
7	Butte	33483	2881	152	0.086043664	86.0	4.5
8	Sutter	23582	1473	85	0.062462895	62.5	3.6
9	Humboldt	18936	1177	68	0.062156738	62.2	3.6
10	Madera	32807	2977	112	0.090742829	90.7	3.4
11	Fresno	209936	14066	699	0.067001372	67.0	3.3

Mariposa County, another school with small enrollment, is at the top of the list, with 10.5 expulsions for every 1,000 students. Del Norte is second with 8.7.

Los Angeles is 52nd in expulsions with .4 for every 1,000 students.

Save your work. We're going to run one more set of calculations that will help us place our county-level results in context.

What if we want to know the average rate for all of the California counties? We cannot accurately calculate an average rate using individual rates that were based on different populations. We could calculate a weighted average, which would take those differing populations into account. However, that's fairly complex and outside the scope of this lesson. Instead we will create totals for the state enrollments, suspensions and expulsions, and then use those to calculate **total rates**, which we can use as benchmarks for the counties.

Navigate to the bottom of your spreadsheet. In cell A61 enter "Totals" for the label. In cell B61, use the =SUM() function to add all the county enrollments. Copy that formula into cells C61 and D61.

Now copy the contents of cell E59 down to E61. Likewise copy F59 to F61 and G59 to G61. Delete the contents of cells E60-G60. "#DIV/0!" is the error message that Excel issues when we attempt to divide by zero or a cell that's empty.

50.0	1.2

In cell F61 we have the statewide total suspension rate of 50 per every 1,000 students. Cell G61 tells us the statewide expulsion rate of 1.2 per every 1,000 students. We can compare individual counties and see where they fall in relationship to those averages.

That's it for rates. Save your work and close the spreadsheet.

RATIOS

Ratios are also useful because they allow us generate a number that expresses the relationship between two different items. One example: Let's say we have six apples and three oranges in our shopping cart. We could use a ratio to compare our apples to oranges (even though we've been warned countless times against doing this). To compare apples to oranges, we divide our number of apples (6) by our number of oranges (3) to get our result of 2. So we can say apples outnumber oranges 2 to 1.

A more useful application for us as journalists would be calculating a ratio that might help determine whether public school classrooms are crowded. We'd calculate a student-to-teacher ratio by dividing the number of students by the number of teachers.

Another application would be using a ratio to compare the number of traffic stops made by police by race. Journalists have divided the number of minority drivers stopped by the number of white drivers stopped to show racial disparities.

CALCULATING RATIOS

Now we're ready to go back to Excel. Make a working copy of RI_student_teacher.xlsx and open it. The file, downloaded from the National Center for Education Statistics and modified for this lesson, holds data about the number of teachers and students at Rhode Island public schools. (Link: http://nces.ed.gov/ccd/elsi/)

 A four-corners check shows that we have 291 rows of data, including one for headers. Column A contains the school names, B the number of full-time equivalent teachers and C enrollment. The FTE tells us how many teachers are working full time, plus the number of part-time teachers, expressed as if they were full timers. So, if a school has 30 full-time teachers and two half-time, it would have 31 FTEs.

We'll calculate the student-to-teacher ratio in Column D. Enter "Ratio" in cell D1 for the label.

In Cell D2, we will enter the formula to calculate the ratio for the Agnes B. Hennessey School. Remember, we're after the student-to-teacher ratio, so we are dividing the number of students by the number of teachers. The formula is "=C2/B2". Excel tells us the school has 14.47619 students for every teacher. Enter the formula and copy it down for all of the schools.

We see that our results are formatted inconsistently. Let's reformat the D column as Number with 1 decimal place. Now save your spreadsheet, because we are about to sort to find out which schools have the highest and lowest ratios.

	A	B	C	D
1	School	Teacher_FTE	Enrollment	Ratio
2	NFIACE PROGRAM	--	--	#VALUE!
3	OCEAN TIDES, INC.	--	--	#VALUE!
4	BROAD ROCK MIDDLE SCHOOL	22	520	23.6
5	RINI MIDDLE COLLEGE	6	133	22.2
6	WILLIAM D'ABATE ELEM. SCHOOL	19	402	21.2
7	RESERVOIR AVENUE SCHOOL	15	306	20.4
8	URBAN COLLABORATIVE PROGRAM	7	141	20.1
9	FRANK D. SPAZIANO ANNEX	11	220	20.0
10	ROBERT F. KENNEDY EL. SCHOOL	28	529	18.9
11	ASA MESSER EL. SCHOOL	32	599	18.7
12	WOONSOCKET HIGH SCHOOL	92	1717	18.7
13	FIFTH AVENUE SCHOOL	14	261	18.6
14	HARRY KIZIRIAN ELEMENTARY	33	607	18.4
15	CUMBERLAND PRESCHOOL CENTER	4	73	18.3
16	MARTIN LUTHER KING EL. SCHOOL	33	592	17.9
17	GEORGE J. WEST EL. SCHOOL	42	750	17.9
18	CLASSICAL HIGH SCHOOL	63	1105	17.5
19	VEAZIE STREET SCHOOL	37	642	17.4
20	THE GREENE SCHOOL	7	121	17.3
21	ALAN SHAWN FEINSTEIN ELEM.	25	430	17.2
22	WEBSTER AVENUE SCHOOL	22	372	16.9
23	STONY LANE EL. SCHOOL	26	439	16.9
24	NATHANAEL GREENE SCHOOL	33	554	16.8
25	COMMUNITY SCHOOL	37	621	16.8
26	NATHANAEL GREENE MIDDLE	56	938	16.8
27	CARL G. LAURO EL. SCHOOL	53	885	16.7

ELSI Export

Select the block of data and sort largest to smallest using the Ratio column. Two entries with no data – NFIACE Program and Ocean Tides, Inc. –appear at the top of our list.

Broad Rock Middle School, with 23.6 students for every teacher, is the leader.

	A	B	C	D
1	School	Teacher_FTE	Enrollment	Ratio
2	BLOCK ISLAND SCHOOL	26	114	4.4
3	MILDRED E. LINEHAM SCHOOL	7	35	5.0
4	DRUM ROCK EARLY	23	133	5.8
5	JACQUELINE M. WALSH SCHOOL	20	137	6.9
6	JAMESTOWN SCHOOL-LAWN	25	196	7.8
7	GORTON JUNIOR HIGH SCHOOL	56	451	8.1
8	PARK VIEW MIDDLE SCHOOL	54	443	8.2
9	CALCUTT MIDDLE SCHOOL	49	433	8.8
10	LIPPITT SCHOOL	28	251	9.0
11	VETERANS MEMORIAL ELEMENTARY	40	360	9.0
12	PROVIDENCE CAREER TECHNICAL	55	498	9.1
13	NORWOOD SCHOOL	29	268	9.2
14	MEADOWCREST EARLY CHILDHOOD	6	56	9.3
15	HUGH B. BAIN MIDDLE SCHOOL	41	385	9.4
16	PARK SCHOOL	27	255	9.4
17	JOSEPH A. WHELAN SCHOOL	22	209	9.5
18	SEGUE INST FOR LEARNING	21	201	9.6
19	WINMAN JUNIOR HIGH SCHOOL	56	540	9.6
20	HOLDEN SCHOOL	26	256	9.8
21	GEORGE HANAFORD SCHOOL	18	178	9.9
22	NARRAGANSETT HIGH SCHOOL	50	495	9.9
23	ROBERTSON SCHOOL	20	198	9.9
24	SCITUATE MIDDLE SCHOOL	41	406	9.9
25	TIVERTON HIGH SCHOOL	57	565	9.9
26	BLACKSTONE VALLEY PREP E. 2	8	80	10.0
27	DR. EDWARD RICCI SCHOOL	34	340	10.0

ELSI Export

Now sort smallest to largest. This tells us that at Block Island School the student-to-teacher ratio is 4.4. It is also a small school in terms of overall enrollment.

It would be nice to have an average ratio to which we could compare our individual schools. However, we can't accurately average ratios that are based on different denominators, unless we calculated a weighted average. Instead, we'll calculate a total rate for the state.

Insert a new totals line by highlighting Row 292. Right click and pick Insert two times.

BROAD ROCK MIDDLE SCHOOL	22	520	23.6
NFIACE PROGRAM	–	–	#VALUE!
OCEAN TIDES, INC.	–	–	#VALUE!
Totals	10801	138124	12.8

In cell A293 enter "Totals" as our label. Sum the B and C columns. In cell D293, enter "=C293/B293" to calculate the total ratio. Our result is 12.8. We now have a number for comparison.