

## Practice: Rates and Ratios

---

Let's solidify the skills we just learned by analyzing a new dataset – transit.xls.

When a Chicago Transit train derailed at the O'Hare Airport in March 2014, reporters rushed to cover the breaking news. More than 30 people were injured. You can watch the derailment [here](#).

NBC Chicago reporter Phil Rogers wanted to know how Chicago derailments compared to other major cities and asked for data from the federal Department of Transportation the day of the derailment. Transit.xls is exactly what he received. You can watch his story [here](#).

Please keep in mind that this is practice data and should not be used for a story. Contact your state department of transportation or the federal DOT to request similar data.

Sometimes you will get Excel sheets from entities that will be “dressed-up.” They may use different fonts or color code items. Make sure you understand what information is contained in this “pretty” version of the data; never hesitate to contact the agency you got it from and ask them for clarification.

So in this fancy sheet, we can see the DOT has done some calculations for us. Our past work tells us we should never trust other's math. The data in columns A through G include raw numbers of derailments and fatalities. In Column H the DOT calculated six year averages of those annual numbers for each city's transit.

On the right side of the sheet, we see the total vehicle revenue miles and below that are a bunch of rates (fatality over miles and derailment over miles).

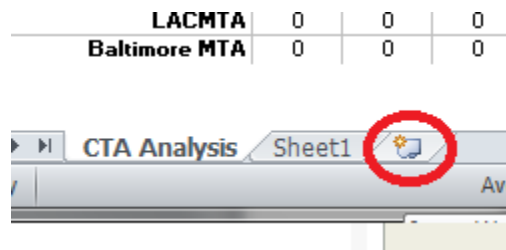
One thing to keep in mind when using rates, is it doesn't have to involve a population of people (or per capita). You can calculate rates using different measurements. In this instance, you notice in their calculations the federal government uses vehicle revenue miles traveled. For example, cell K15 is telling us that Chicago Transit had seven derailments per 100 million miles in 2008. Like our previous rate exercises, by calculating the derailments or fatalities over miles, we are able to compare all of the transits (big and small) at the same level. (Apples versus apples). If we compare only each city's count of derailments, it's not a fair comparison. A transit like Chicago puts on more miles so of course, it's more likely to have more derailments. For example, Transit A may have 7 fatalities in 2007 and Transit B has 5. However, Transit B has a higher rates of fatalities because it's only traveled 10 million miles that year versus Transit A's 100 million miles.

If you want to get rid of all the fancy stuff like the colors and the formatting, you can copy the data and open a blank sheet. In cell A1 of that blank sheet, right-click and choose Paste Special and values.

If you want to calculate any of the data, you will want to copy the numbers (not the averages) and put into a separate sheet. It's much easier to focus on the task at hand and get rid of the noise of all of the other figures.

Let's take a look at derailments in 2008. Looking at the number of total derailments in 2008, CTA or Chicago Transit is at the top with five that year. Let's level the field and compare each city's transit by taking into account the number of miles traveled.

Using your white plus sign, select cells A14 through B22. Copy using Command+C. You will have the marching ants. Head to the bottom of Excel and click on the sheet that looks like it's on fire (below). It will open a new spreadsheet and paste (Command+V) the data into the sheet.



Go back to the "CTA Analysis" sheet and select cells L2 through L10. We do not highlight the transit names because they are in our new sheet and in the same order. But it is worth double-checking. Copy so we have the marching ants. Paste on the new sheet next to the derailment data.

In D1, label the column "DerailRate". If we take the number of derailments over total revenue miles (=B2/C2), we get a lengthy decimal.

Clipboard		Font	
D2		fx =B2/C2	
A	B	C	D
	2008	2008	DerailRate
CTA	5	67,450,040	7.41289E-08
MBTA	3	21,919,009	
MARTA	0	22,208,017	

In order to get a number we can understand (and inform our audience about), we will want to multiply that decimal by 100 million like below.

Clipboard		Font	
D2		fx =(B2/C2)*100000000	
A	B	C	D
	2008	2008	DerailRate
CTA	5	67,450,040	7.412894047
MBTA	3	21,919,009	

Copy that formula down. We can clearly see that MBTA or Massachusetts Bay Transportation Authority had a much higher rate of derailments in 2008 when you take into consideration how few miles it traveled compared to CTA.

Let's test these skills by asking our data some questions.

---

### ASK YOUR DATA QUESTIONS

---

Let's pretend we are a Chicago reporter and answer the following questions. We recommend copying the cells of data into new sheets to do your own analysis. It's cleaner and offer less room for mistakes.

1. Which city has the highest rate of fatalities (excluding suicides) over the last six years and where does Chicago rank?
2. Over the six years, did Chicago transit have more derailments than other major city transit systems? Is it getting better or worse?
3. Which year was the worst for all major transit in terms of fatalities (excluding suicides)?
4. How many suicides happened at CTA in 2013?
5. What questions should I ask the DOT data clerks regarding the data?
6. What other data might be useful to mine after this story runs?

###