# Making charts in Excel

Use Excel file – "MakingChartsInExcel_data"

We'll start with the worksheet called "treatment"
This shows the number of admissions (not necessarily people) to Twin Cities' treatment programs by year and by the primary substance. This counts admissions because sometimes people will go through more than one treatment episode in a year. The programs don't track people, they just track the number of admissions. And this is the primary drug—many addicts come to treatment addicted to more than one thing. It leaves out alcohol, which is – by far – the biggest reason for treatment.

Let's start by looking at the tools we'll be using in Excel.

If you click on the Insert menu (or ribbon) at the top of Excel, you'll see these options:
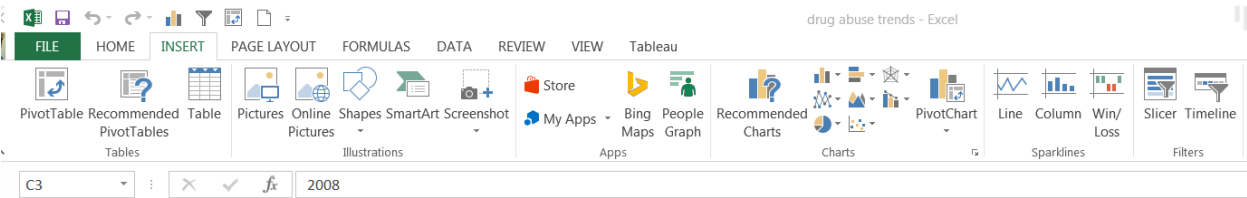
Toward the right side you'll see the "Charts" section. You can click on Recommended Charts to get Excel to make an educated guess on which type to use, or you can click on one of the images to the right of that for column chart (bars are vertical) or bar chart (bars are horizontal) or line chart (best used for time series), scatterplot, pie charts and others.

Once you create a graphic, two new "Chart Tools" ribbons are added. The main one we'll use a lot is called "Design"
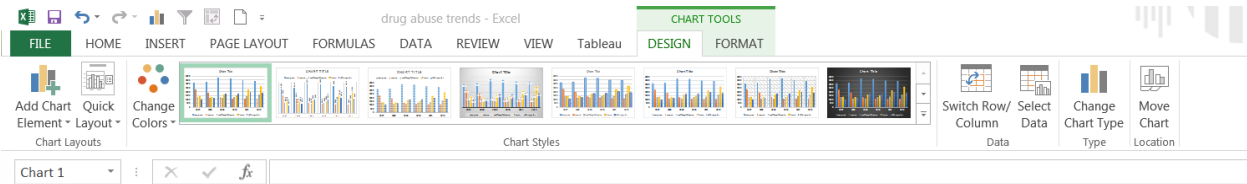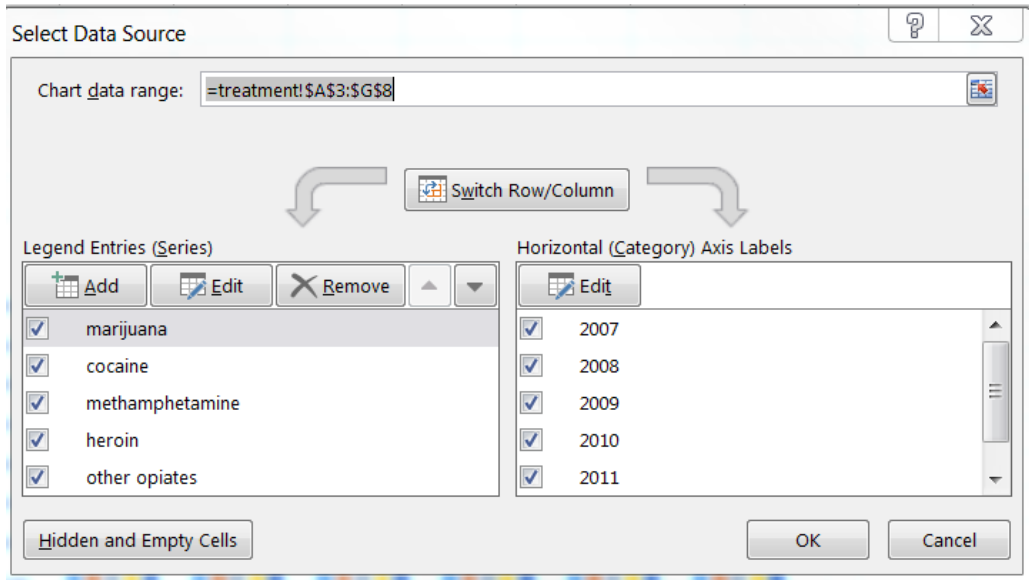

The buttons we'll use a lot are:
Switch Row/Column—reverse how the data is presented
Select Data – apply labels
Change Chart Type—change to another chart type (i.e. line chart, bar chart, etc)

Quite possibly one of the most confusing things is how to apply labels to the charts. That's what the "Select Data" button is for. When you click on that it brings up this window:


On the left will be where you create the legend. In this case, we're identifying what each color represents for our columns. (this is the vertical axis)

On the right, you set the labels for the horizontal axis (running along the bottom).

Click the "Edit" button to add the labels.

Now let's start working with the data in our worksheet – called "treatment" – to look at the various ways we can present this data.

- Line chart. There would be 5 lines, one for each drug, and 6 data points for each line, representing the years. This does a good job of showing change over time. It's harder for user to discern the gap between each drug – but they get a general idea.
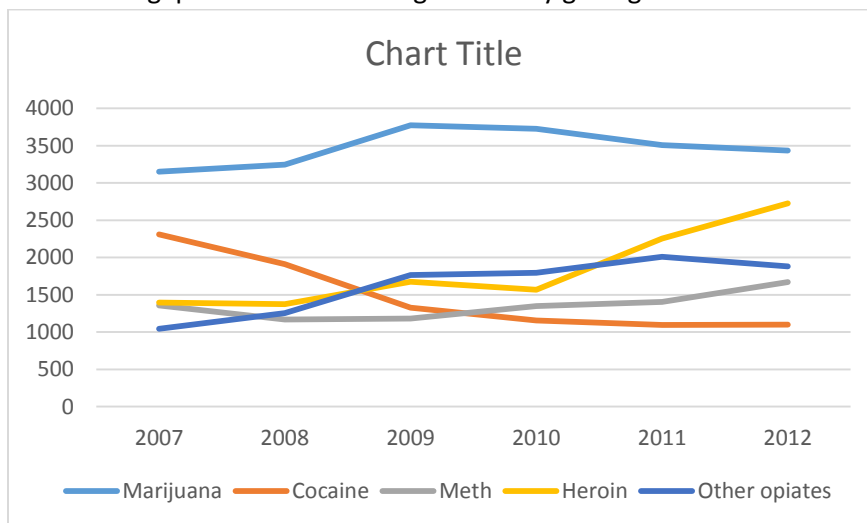
- Column chart. If you keep the years on the horizontal axis and each bar is a different drug, you end up with this – Can you tell which drug has increased most over time?
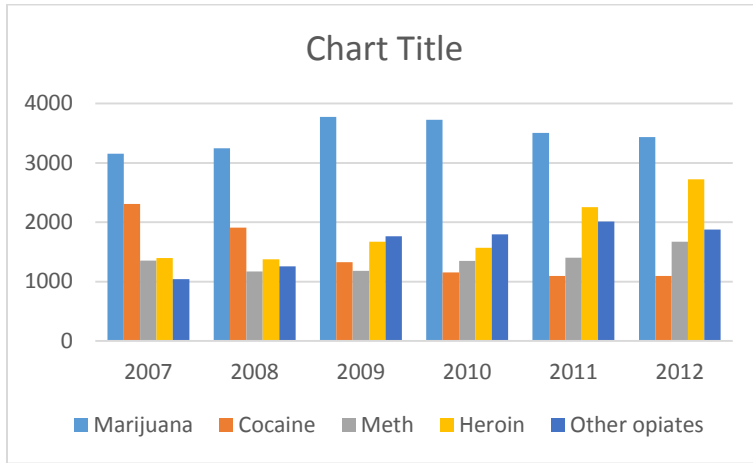

With the column chart still open, go to "Switch Row/Column" – so you now have 5 "groups" instead of 6. Each group represents a drug and the bars represent the years.
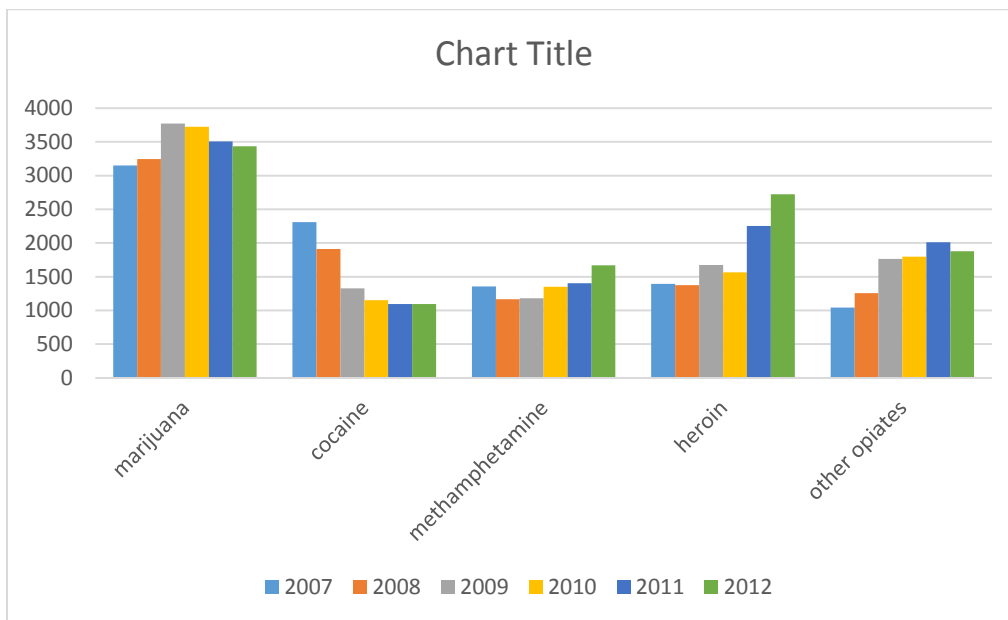
- Then go to "Change Chart Type" and switch it to a bar chart like this:
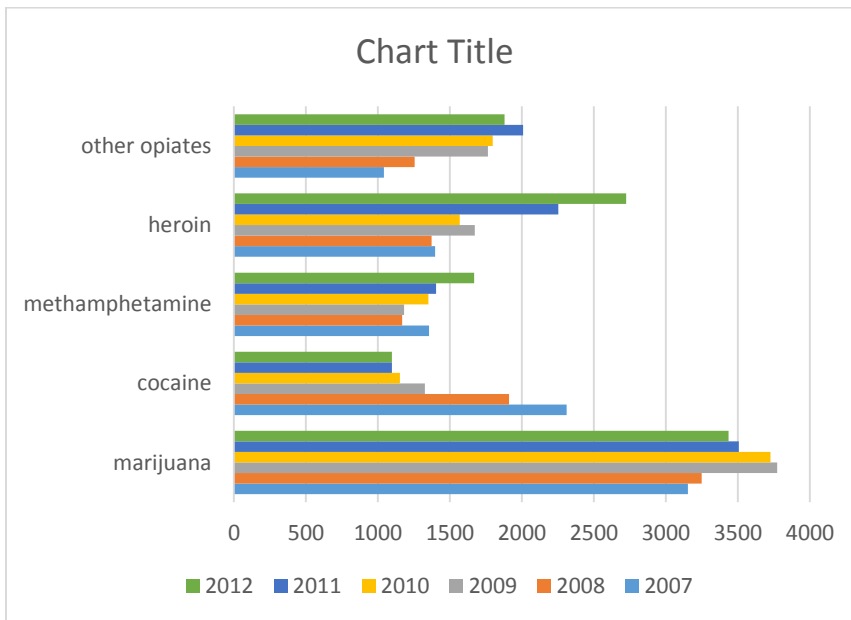
- Then change chart type again and look at a Stacked Bar Chart. This changes the picture altogether. Now each bar represents ALL of the admissions across all of the years. So now it doesn't show change over time at all.
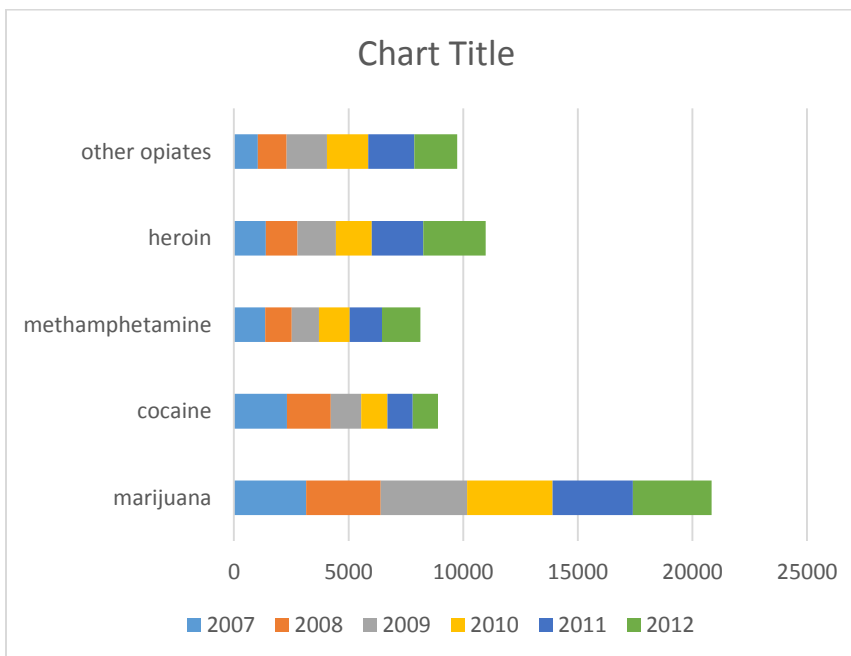

- Now change chart type again to 100% stacked COLUMN chart and, if necessary, switch row/column so that each column represents a year and the colors represent the drugs. This shows the drugs as a percentage of the whole. So you can see that marijuana has held pretty constant over the years, while treatment for cocaine is becoming a smaller portion of the whole

and heroin and other opiates are getting bigger. Note, that the thing you can't see here is whether total treatments have gone up.

Chart Title
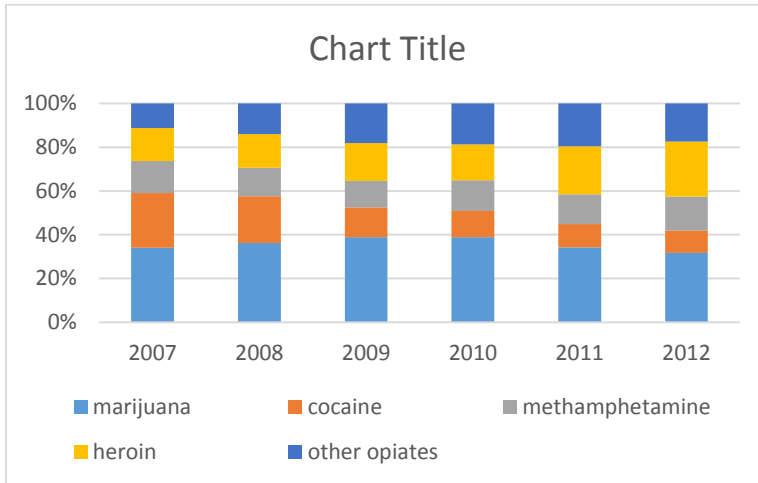
If you change it to a regular stacked column, then you also see how much total treatment admissions have gone up. Notice that it has converted back to number of admissions, not percentages.

Chart Title
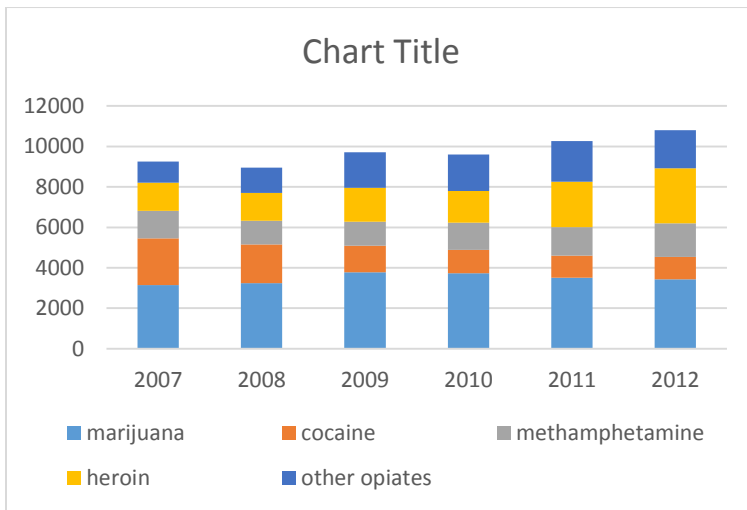
Now, let's work with the data in worksheet called "DEATHS"
This has total overdose deaths by drug, by year for Ramsey and Hennepin counties. Just below, I've added a set of lines for the total of the two counties.

The total data is easy to work with – we can do a line chart or bar chart.
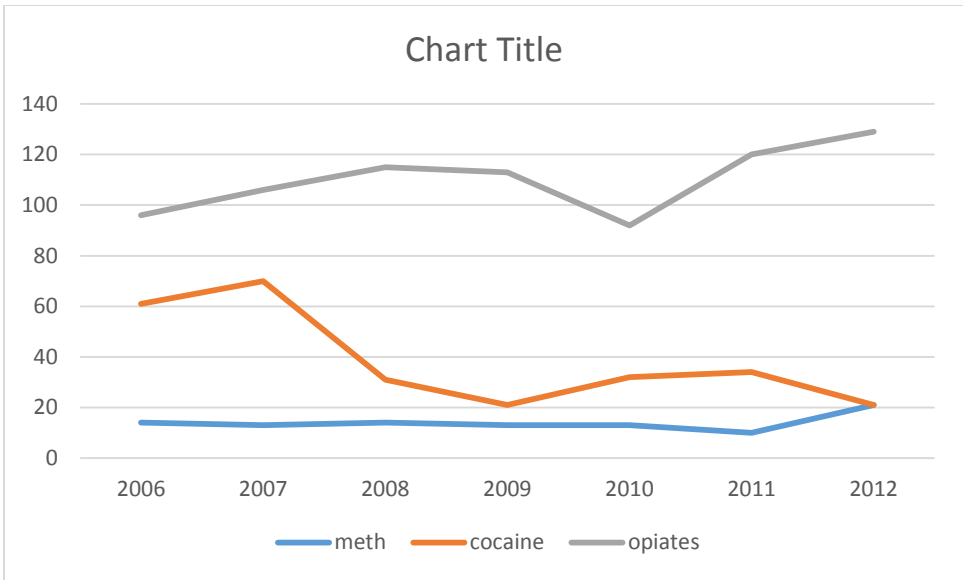
But what if we want to show both Ramsey and Hennepin separately? That gets much harder.

If we take all 6 rows of data – for the two counties – and put that into a chart, it will look something like this:
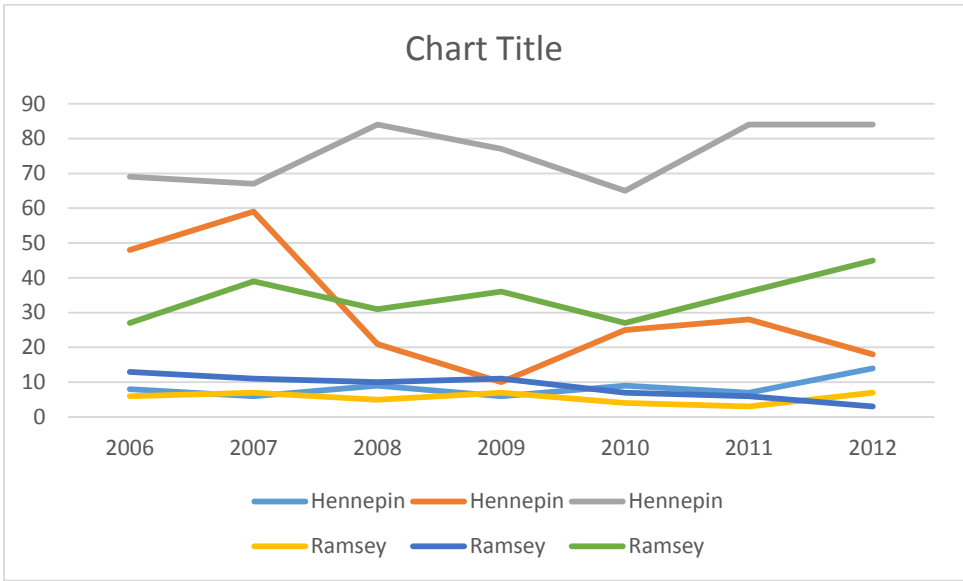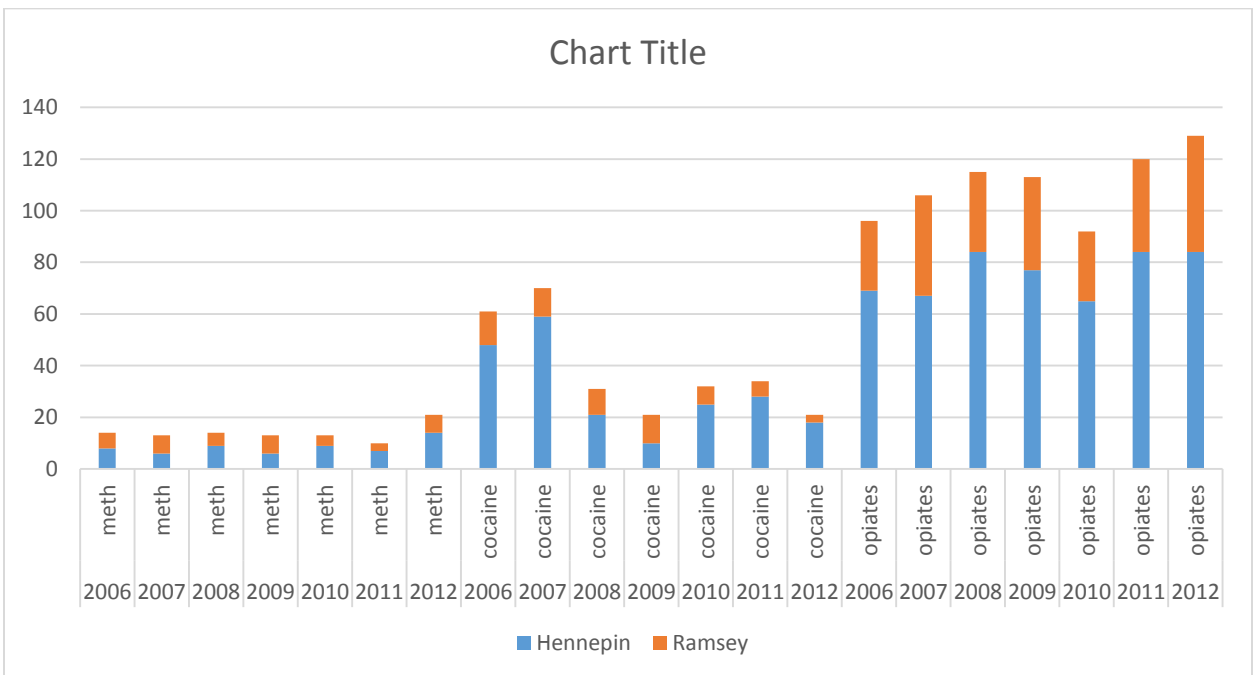

It's pretty confusing trying to figure out that the yellow line and the light blue lines are both Meth and the green and grey lines are both opiates. It requires the reader to spend a lot of time studying it to see what the patterns are.

If you switch it to a column chart and then switch the row/column you would end up with something like this – again, way too confusing:
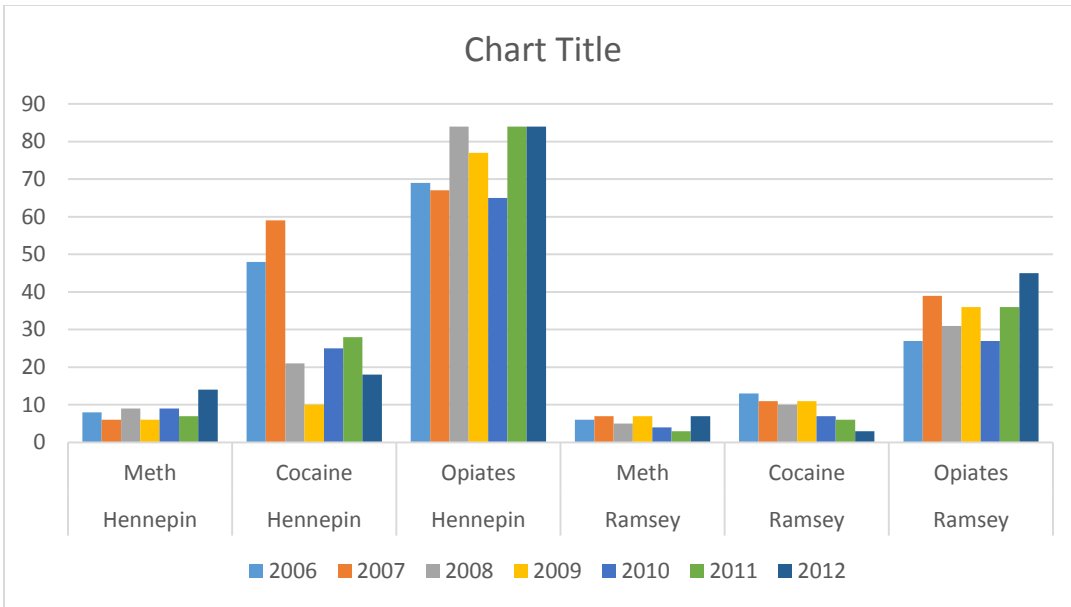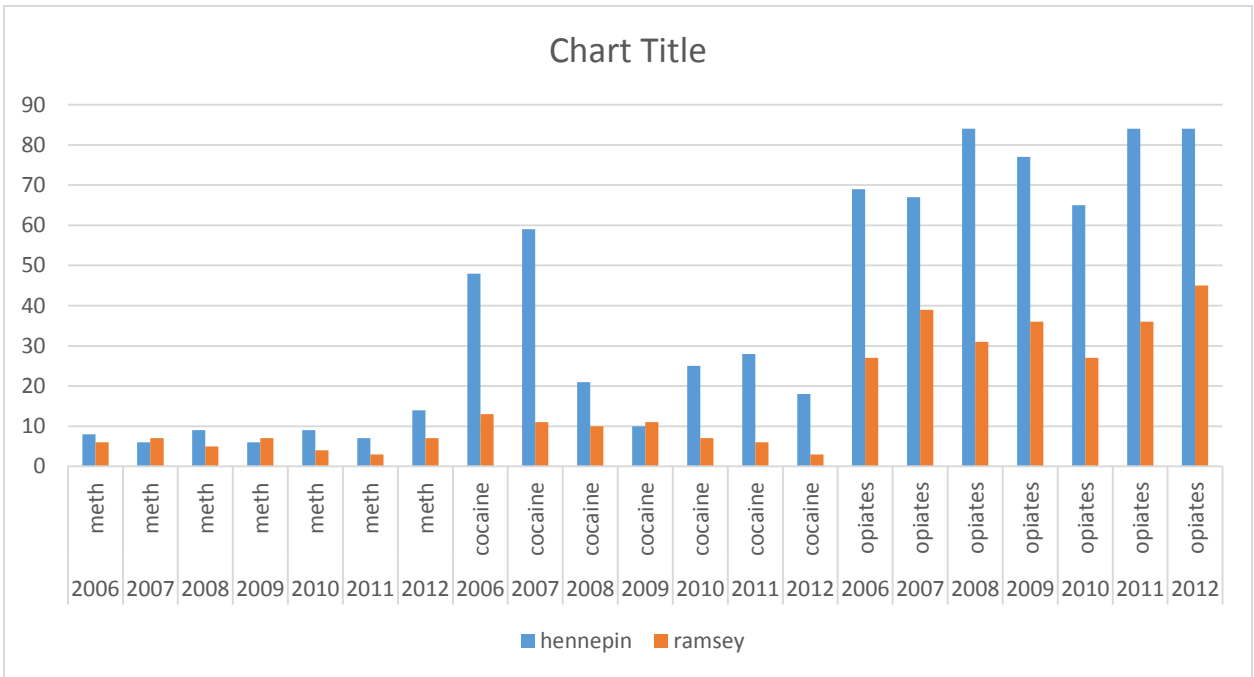
Let's try setting up the data in an entirely different way. The table at the bottom of the page has one row for each year and drug and then just two data columns – one for Hennepin and another for Ramsey.

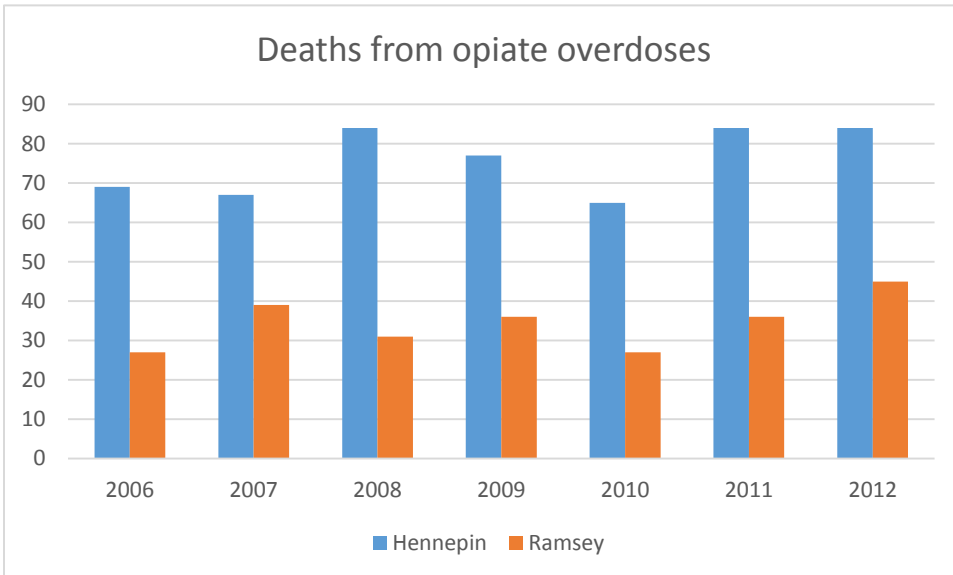A stacked column chart would get us this:

Or a clustered column chart would get us this:

Chart Title

A clustered column chart showing hennepin (blue) and ramsey (orange) values for meth, cocaine, and opiates across years 2006–2012.

That's still quite confusing and basically just shows us we have too much information that we're trying to pile into one chart. So then we could think about breaking it up.

How about this, using just the opiate deaths from the original table at the top of the worksheet:

Deaths from opiate overdoses

A clustered column chart showing Hennepin (blue) and Ramsey (orange) opiate overdose deaths for years 2006–2012.

Next, let's work with the data in the worksheet called "Hmong." This shows the Hmong population in Minnesota in 1990, 2000 and 2010 – tallied up by how many lived in St. Paul, Minneapolis, the 7-county Twin Cities' suburbs and outstate Minnesota.

The main thing I want to do with this data is to show my readers how this population has shifted in where they live. Your first instinct might be to make this into a map. It would be easy to calculate the percentage change for each geographic area and then create a map that colors those areas by that percentage change.

If you did that you'd end up with a 3030% change for the 7-county suburbs and a 563% change for outstate. Kind of crazy numbers that readers won't understand.

We could try a clustered bar chart like this:
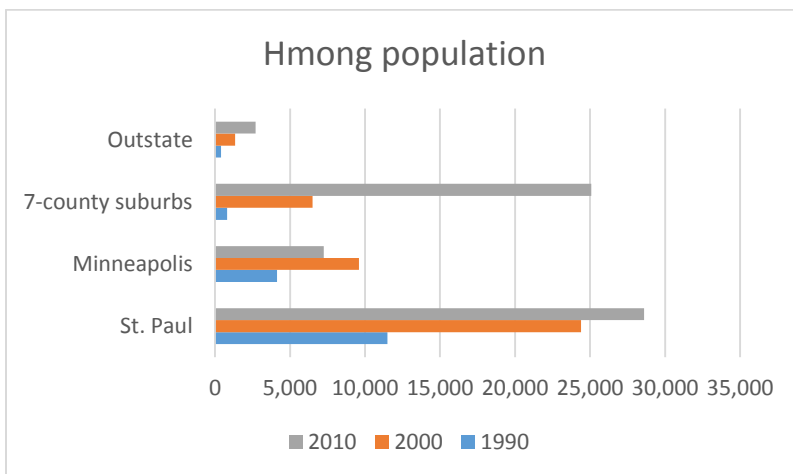
This essentially gets the job done. It's showing us that the population in the suburbs has exploded, while it's gone down in Minneapolis. The main thing I don't like about this is that Excel has put 2010 (the most recent year) on the top bar – so it's going in the reverse that a reader would expect.

You can change that by going to "Select data" and on the left side ("legend entries") you'll see a little down arrow next to the Remove button. Click on 1990 and push the down arrow button to move it down. Then move 2000 down so that it goes in reverse chronological order.

Let's try something else. Start a new chart and this time choose Horizontal bar chart – 100% stacked. It should give you a chart that has 4 horizontal bars. If so, push the Switch rows/columns button so that you end up with 3 horizontal bars.
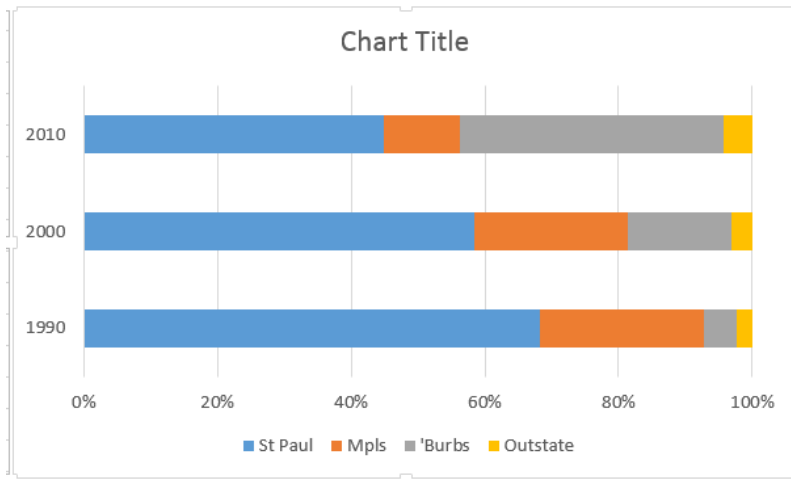
Like that other chart, we probably want to reverse the order, so that 1990 is at the top and 2010 is at the bottom. The only way I've found to do that is to change the order of the columns in the original chunk of data.
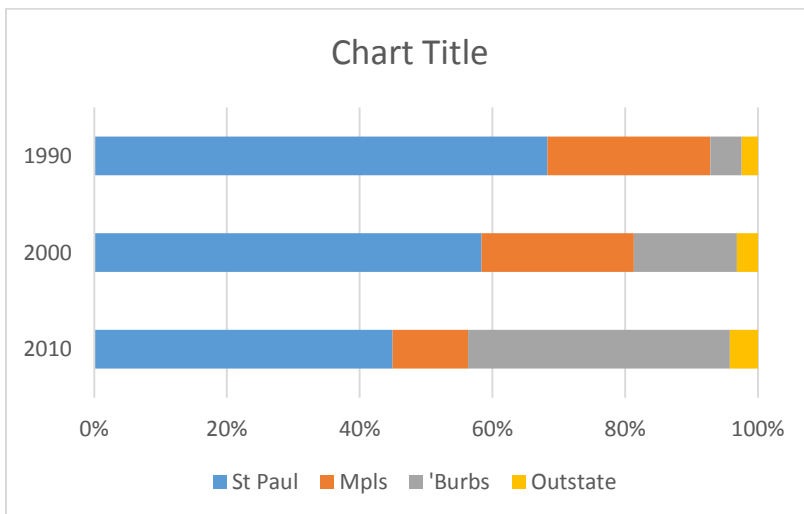

Now we're not really showing the growth in the population anymore – like we did with the last one – but we can more clearly see the redistribution in where they live. We see very quickly that St. Paul remains the predominant location, while Minneapolis has shrunk and the Suburbs have grown dramatically.

Next, let's work with the data in the tab called "Minorities." This is U.S. Census data on minority groups in Minnesota. It shows the total population for each in 1980, 1990, 2000 and 2010 (note:

Census did not add the multi-race option until 2000. American Indians were not counted separately in 1980 (they were in the "other race").

If we want to show change over time in these groups, there are a variety of ways we could do that.

We could use stacked columns to show the growth in minorities overall, using just the total people (note: here it's OK to use total numbers, rather than rates, because we aren't comparing disparate groups):
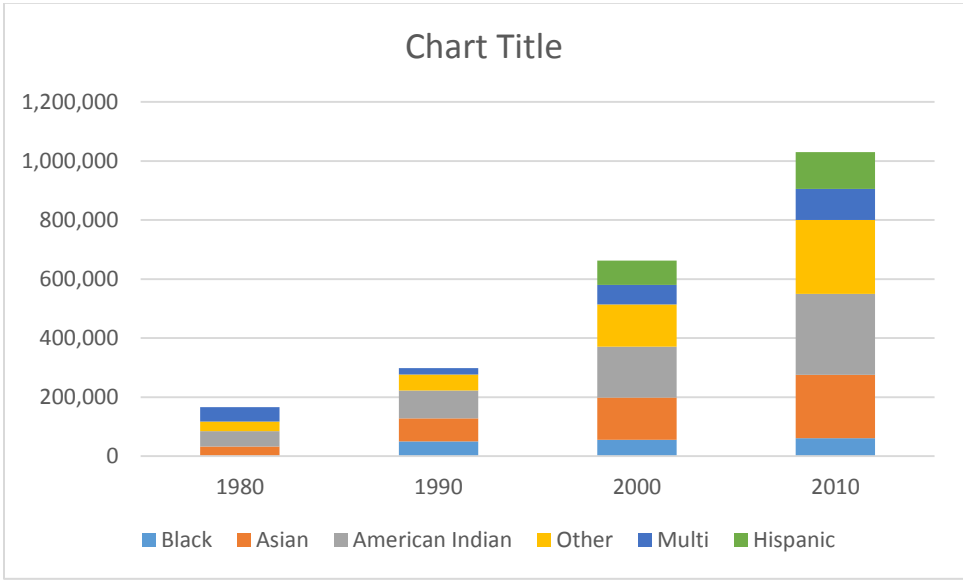
If you want to give your readers a better idea of which groups are the biggest, or have grown the most in terms of raw number of people, you could plot the totals like this (I chose to skip the 1980 data since two groups don't have data for that year). This chart is a bit more complicated and a bit harder for the reader to grasp quickly.
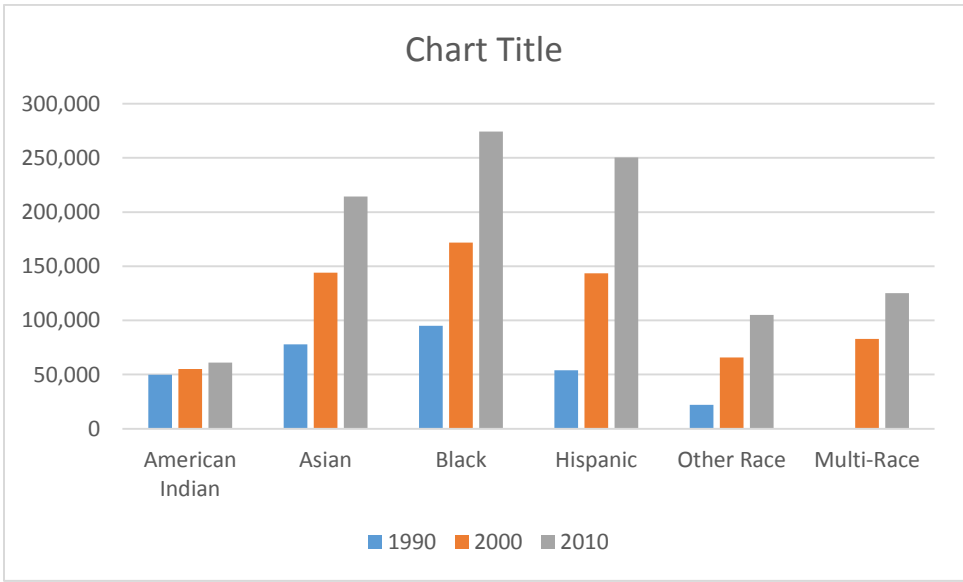
If we want to see which group has grown the fastest, the simplest route – and most technically correct in terms of the mathematics -- would be to plot the percentage change. For this one I calculated the percent change between 1990 and 2010. I chose to avoid 1980 since there are two groups that don't have values for 1980.

Chart Title
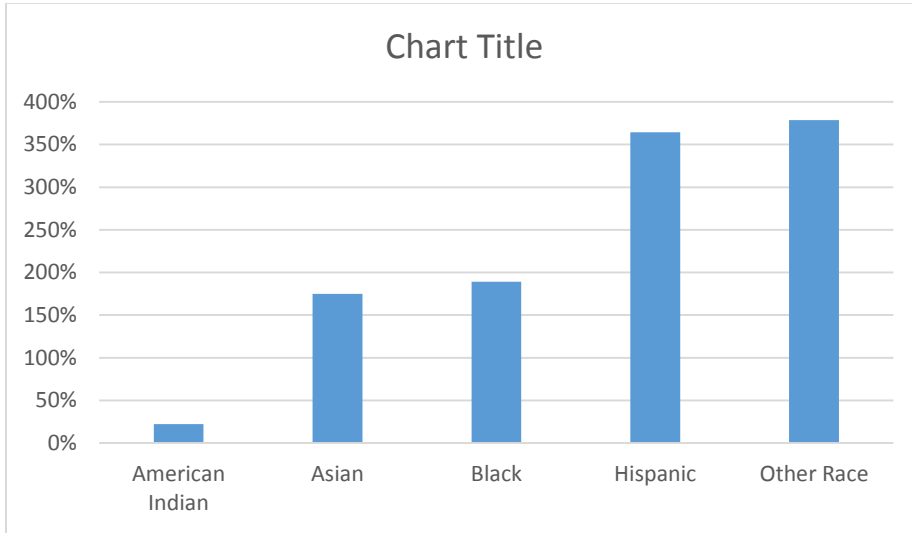
Created by: @MaryJo Webster
April 2014
mjwebster71@gmail.com