

LECTURE 10: OF DATA AND DISPLAYS II

I. Scatterplot

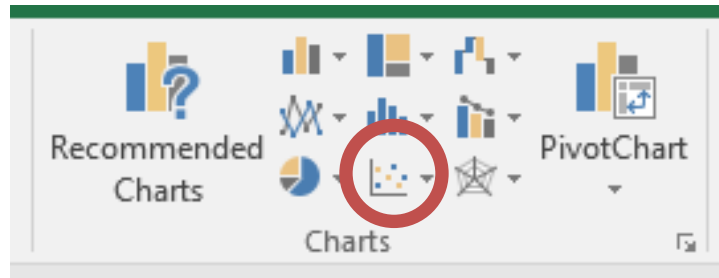
- a. The first step in any research project is finding data (this sometimes occurs even before you know what you want to investigate).
- b. The second step is determining your approach to the data.
- c. A *scatter diagram* indicates how two (or more, if you are feeling adventurous) values relate to each other.
- d. Gapminder (www.gapminder.org) is an excellent resource to explore relations between different variables. The website employs data from all over the world to various sophisticated scatter plots. The raw data are available in Excel format.
- e. You'll notice on Gapminder that you can express a variable on a linear (lin) or logarithmic (log) scale.
 - i. A linear scale means each unit is some previous unit plus a fixed value. For example: 10; 20; 30; 40; 50; etc.
 - ii. A logarithmic scale means each unit is some previous unit *times* a fixed value. For example: 10; 100; 1,000; 10,000; etc
 - iii. For values with a wide range (especially ones skewed right) logarithmic scales are a better visual choice.

II. Creating displays Practice

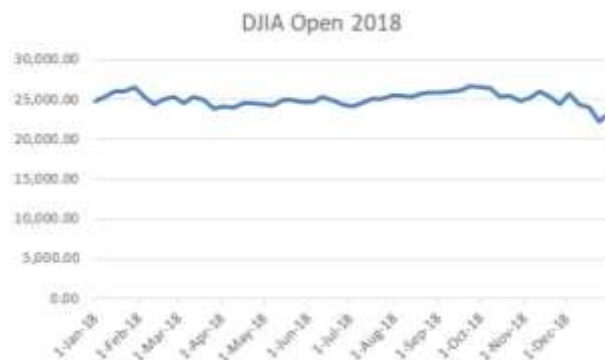
- a. Open Data Set 1, found on my website. This is cross-sectional data of 184 countries and seven variables.
 - i. Keep in mind the descriptions tab at the bottom of the page if you want to know more about what each variable is.
- b. I use ">" to indicate click order. For example, Page Layout > Margins > Normal means click Page Layout, then Margins, then Normal.
- c. It's always a good idea to add labels. You can find how to add labels (notably the horizontal label, the vertical label, and the title) in the formatting area after you make a display.

III. Scatterplot

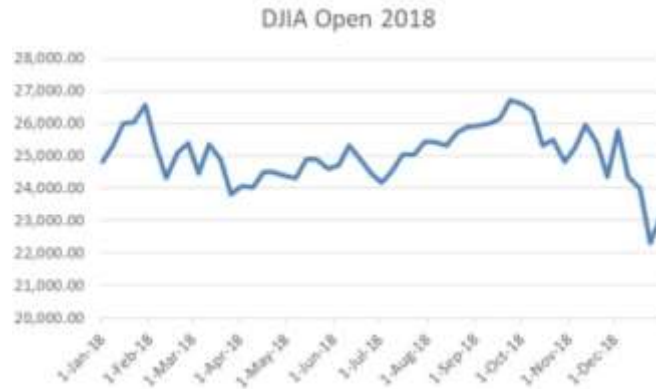
- a. [Here's](#) a video tutorial of making a scatterplot.
- b. First highlighting columns G and H (murder rate and pop density).
- c. Insert >> Scatter image >> Scatter.



- i. Excel defaults whatever variable was on the right as the vertical axis and title. Whatever's on the left is the horizontal axis.
 - d. You'll notice that while some observations stand out, we can't really tell what's going on. We need to transform population density using natural log.
 - e. Excel makes this easy. Click the population density axis and then right click it. Select Format Axis. You'll see a logarithmic option appear on the right side of the screen. Click it.
- IV. Truncating Axes
- a. The range of the axes on charts can be changed, usually done by truncating, or cutting off, part of the y axis. A truncated graph's y axis does not start at zero; this enables easier reading of the graph.
 - b. For example, considering this line graph of the opening weekly values of the Dow Jones Industrial Average for the year of 2018.



- c. It's hard to see how much the values are changing over the years. Let's change it by changing the y axis.
 - i. To do this, select anywhere on the y axis and right click, selecting Format Axis.
 - ii. Under Bounds, let's change the minimum to 20,000.



- iii. Now we can see what's going on week-to-week.
 - d. Excel defaults by truncating the y axis, though truncation comes with dangers. While the above diagram is more readable, the DJIA looks more volatile than it is. The lesson is that you should always watch the y axis for truncation. Deceptive truncation is one of the ways people lie with statistics.
 - e. Another example: labor force participation rate by gender.
- V. Printing
- a. If you want to print an image, click it and try to print it. Excel will print just the image you've selected.