Math Skill #1: Graph Reading and Interpretation

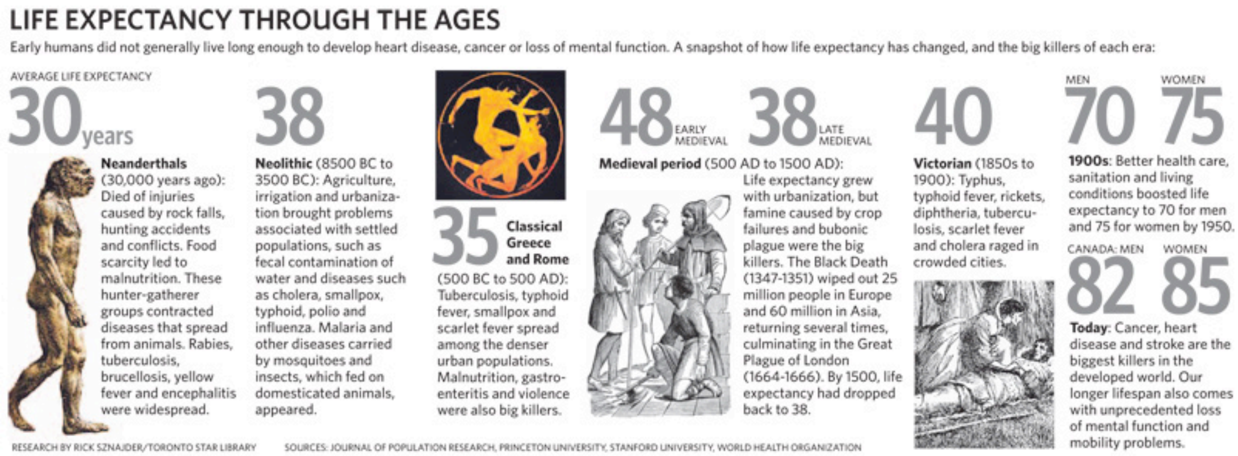
Many of your homeworks this semester will include a worksheet like this that focuses on a specific skill that we've discussed in class and that will be important throughout the semester. These skills were chosen because they are also important in (a) real life and (b) interpreting science in general, and building up the skills you need to distinguish between "good" and "bad" science in particular. Answer the questions to the best of your ability, and come see me at office hours if you struggle with anything.

One thing you will be asked to do frequently in this course is to examine graphs and charts and extract specific information and/or general trends for them.

A chart, graph, or table is only a ***tool*** *in the process of science, and they can be very enlightening OR very misleading. Graphs and charts should be used as a visualization tool, and they are often very helpful in visualizing a lot of data at once. However, they do not always tell the whole story. When considering a graph, you should always think critically about what it is really telling you as well as any factors that are NOT included.*

You can answer these on a separate sheet of paper, or you can print this out and fill in your answers as you go. You may also type in your answers and e-mail the document to me.

The figure below describes some of the general trends in human life expectancy with time. Using the data in the table, make a graph of human life expectancy from 4000BC to 2000AD using a separate sheet of paper (graph paper if you have it).

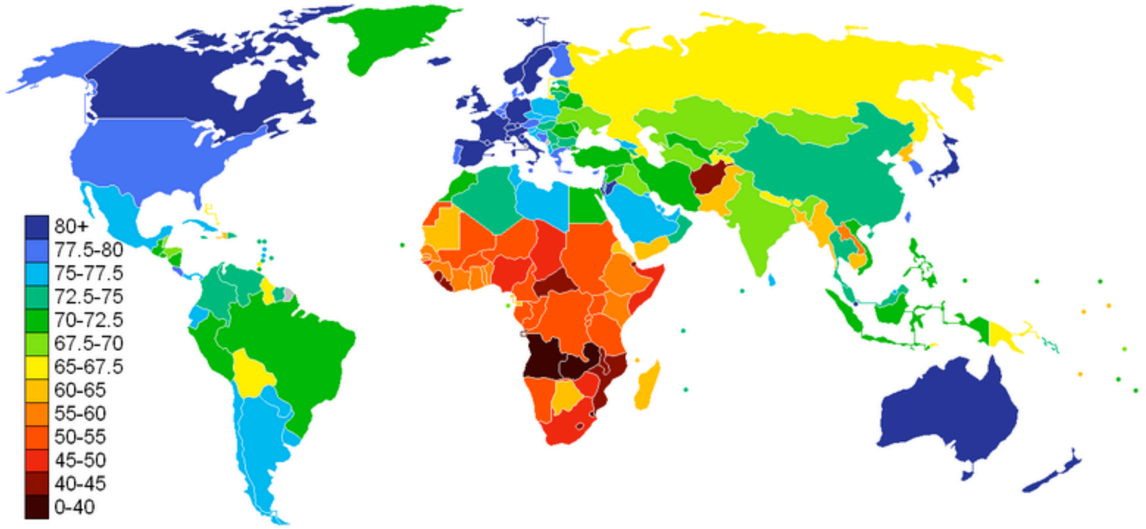


Some things to keep in mind when making your chart:
1) LABEL LABEL LABEL! You should choose a title for your chart and should label both the x and y axes with their units.
2) IT SHOULD TAKE UP THE WHOLE PAGE. Make an intelligent choice of the spacing on your axes so that you use up most of the space on the page.

1. Why was male life expectancy higher in Neolithic times than female? Come up with a hypothesis and then explain what kind of data you might gather (from, say ancient archaeological sites or human fossils) to prove or disprove your hypothesis.

2. If you connect the data points on your chart, you will see a VERY rapid increase in life expectancy in modern times. What do you think changed to cause this to happen? What data could you collect to help prove or disprove this hypothesis?

Let's practice reading graphs using more data on human life expectancy. The chart below shows a political map of the world with countries color-coded by life expectancy.



CIA World Factbook data

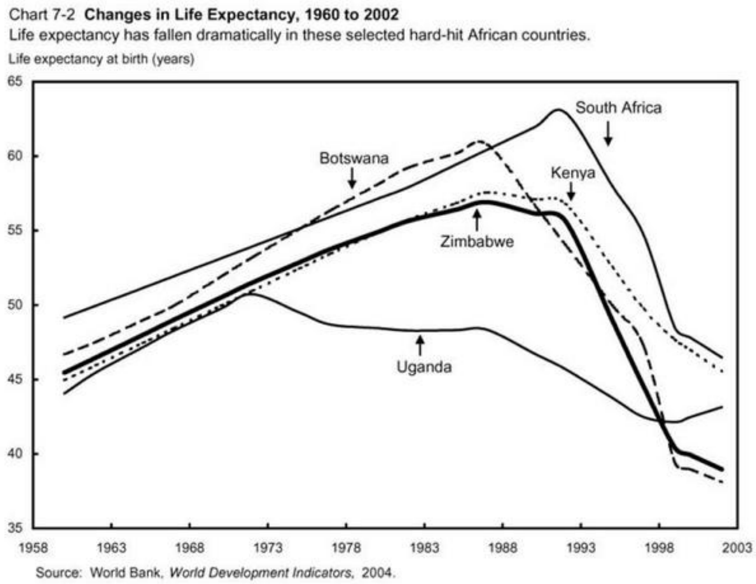
As you may have noticed, this chart contains <u>a lot</u> of data. When confronted with a graph, your first thought should always be: what is this graph <u>really</u> telling me? Note that in introductory science (and in real life!) we are talking about the big picture here. In other words, what does the graph show in general terms? As with much of the data we will see in this class, there are complex factors that affect the details of the graph, but the ability to pull out the big picture is very important.

Read the summary statement below and decide whether you agree or disagree with it:

World life expectancy varies by more than 40 years from country to country, with the longest lived (70+ years) countries concentrated in North America and Western Europe and the shortest-lived countries (<40 years) in Sub-Saharan Africa.

3. Does this summarize the "big picture" to your satisfaction? Would you add anything to the statement above?

4. Give an example <u>using specific countries</u> where the "big picture" may fall short in explaining the whole truth.



The chart above shows human life expectancy data for 5 African countries for the second half of the 20th century. To help you get started with writing your own "big picture" summaries, a general structure is provided below. Fill in the blanks and circle the appropriate options to complete it.

5. Average life expectancy in this region of Africa increased/decreased steadily from

   about _____ to about ______, then _________________________________.

6. (a) Which trend is more dramatic – the increase in life expectancy prior to the late 1980s or the decrease afterwards?

   (b) How can you tell?

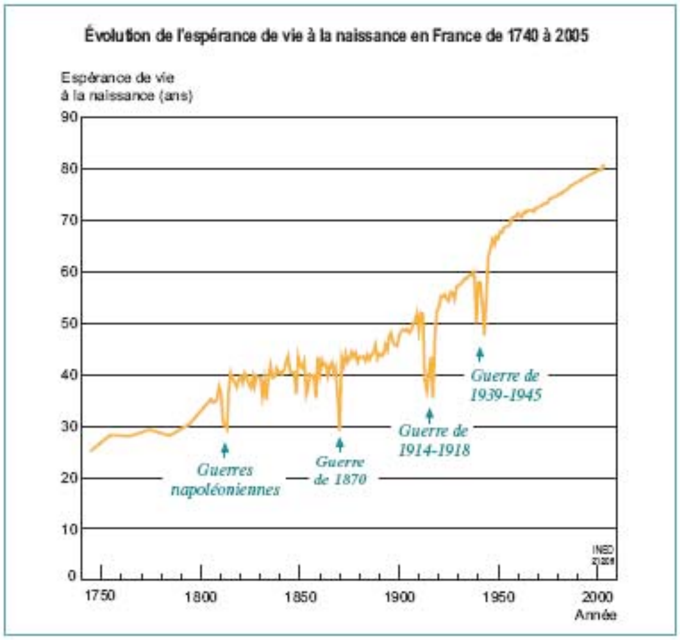
   (c) What may have caused this reversal?

   (d) What additional data would you need to test your hypothesis?

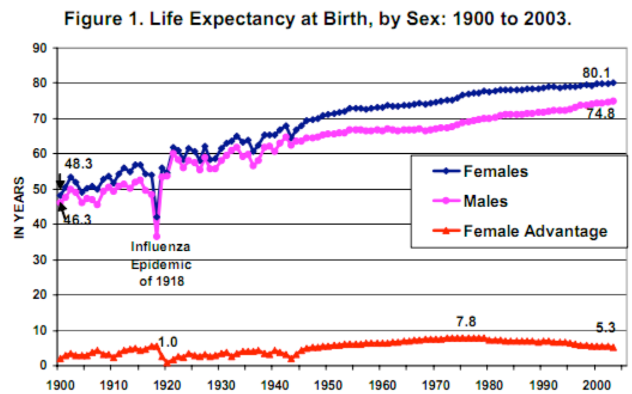
*The property you were describing on the last page is called **slope** and it means rate of change. Really, it is a judge of steepness in a graph, with steeper lines indicating more rapid change and shallower lines indicating more gradual change.*

7. Do all of the countries on the chart follow the same general trend as described above? Are there any that deviate significantly from it?

8. If so, develop a hypothesis as to what this might be. What additional data could you gather to verify or disprove your hypothesis?



Source: French National Institute for Demographic Studies



Source: For 1900-2002, CRS analysis based on data contained in NCHS, United States Life Tables, 2002, *National Vital Statistics Report,* vol. 53, no. 6, Nov. 10, 2004. For 2003, CRS analysis based on NCHS, Deaths: Final Data for 2003, *National Vital Statistics Report*, vol. 54, no. 13, Apr. 19, 2006.

Notes: Later year estimates are more reliable than those of the early 20th century.

9. Above you will find two charts of life expectancy over time. One of them is for the United States and the other for France. Describe the general trend shown in each of the charts in one sentence below (here's your practice with creating your own "big picture" summary):

US:

France:

Note that the axes on these two charts are different, so it is difficult to make a specific comparison between the two simply by looking. Is life expectancy increasing faster or slower in the US than in France? In order to really compare the rate of change, or slope, of life expectancy in these two countries, we need to put them on the same footing.

The mathematical formula "slope = rise / run" can help us with this. In the most general sense, rise/run just expresses how much the quantity on the y axis of a graph changed over a given x-axis interval. Since the x-axis here is time, the slope of the trend is how much the average life expectancy in each country changed over a given interval of time. Let's use the interval 1900 to 2000.

Note that by studying a long-term trend, any temporary changes in life expectancy such as those due to war, disease or famine are "smoothed out".

10. Fill in the life expectancy of US and French citizens at each date below, which you can read directly off the graph (a ruler will help with this).

	1900	2000
US:		
France:		

11. To calculate the slope, first calculate the difference between the life expectancies in 2000 and 1900.

US:

France:

12. Now, divide by the "run" of 100 years to get the average change in life expectancy per year between 1900 and 2000 in the two countries

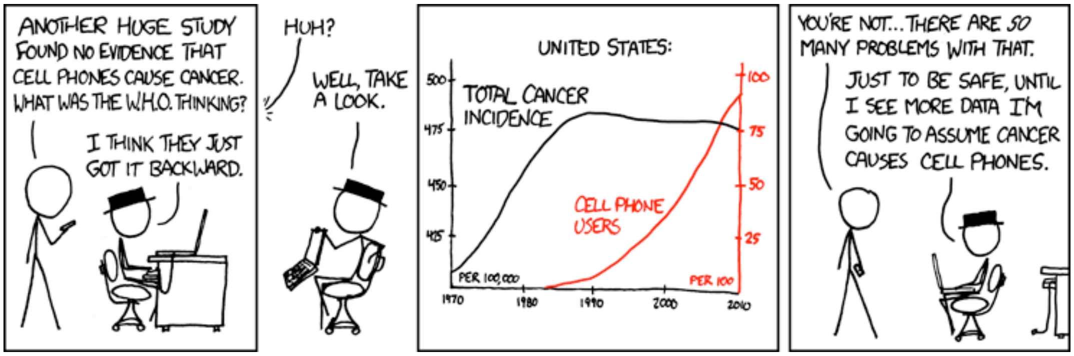
US:

France:

13. Is it possible for this change (your slope) to have been a negative number? Why or why not?

14. Use the slope you calculated to answer the question: If this trend continues, what will the average life expectancy be of a US citizen be in the year 2050?

Correlation vs. Causation



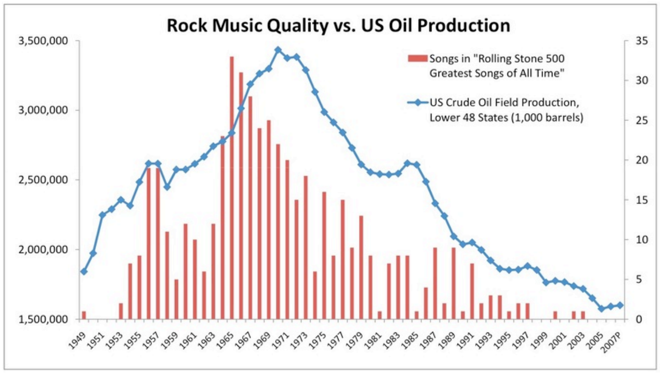
It is also important when examining graphs, particularly those that are presented to convince you of a trend (that two quantities are related to each other somehow) to consider correlation and causation.

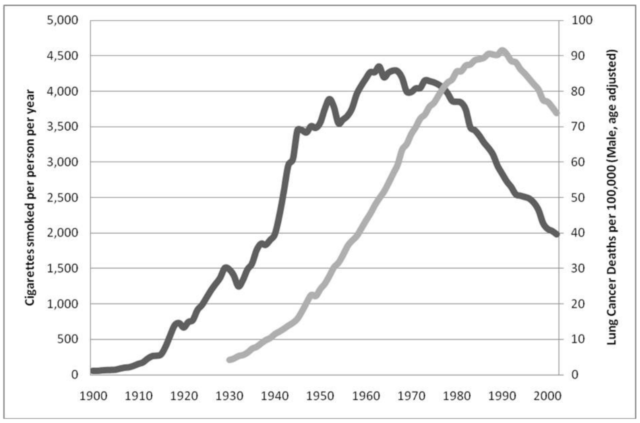
Two quantities that are correlated track each other, in the sense that as one increases so does the other OR as one increases the other decreases proportionally. If two quantities are correlated they will show similar trends over the same[1] interval of time.

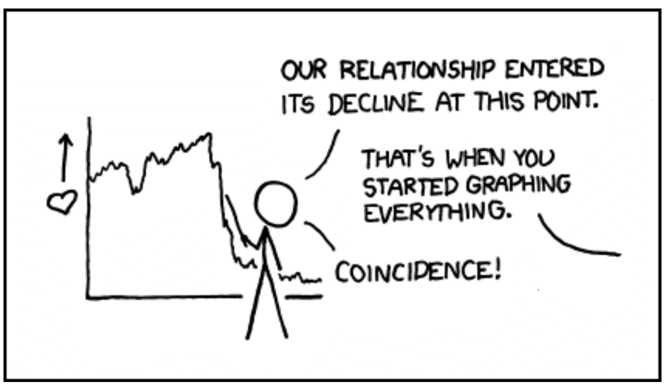
"Correlation does not imply causation" is a warning that is often given when considering results, scientific or otherwise. Just because two quantities increase or decrease together in a particular way, does not necessarily mean that one of the quantities is changing BECAUSE the other is changing. In other words, just because two quantities are correlated, that does not mean that one is the cause and the other the effect. They could follow similar trends for a variety of reasons, including random chance or a relationship to another quantity not considered. Proving that two quantities are correlated is the first step on the way to showing that they have a cause-effect relationship, but is by no means the last.

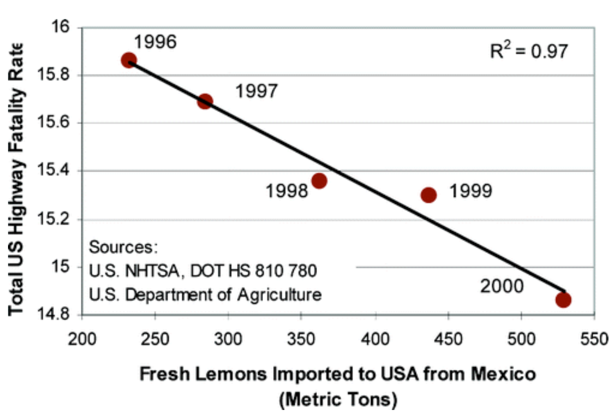
Examine the trends shown in the following four graphs. All four graphs show a correlation, but only two have a legitimate claim for causation. In the space to the right of each graph, write a one-sentence explanation of the trend being shown (your one sentence "big picture" explanation of the graph) and then write one sentence about whether and why (or why not) this trend is likely to be a cause and effect relationship.

[1] Note that it is also possible for one quantity to lead or lag the other in time. For example, a "leading indicator" of the number of new houses in the US could be the number of applications for building permits for new homes. An example of a "lagging indicator" is unemployment, which changes only after the economy has changed, and businesses have had time to adjust their workforce. This is the case with one of the four graphs you will examine in the next section.


Rock Music Quality vs. US Oil Production
Songs in "Rolling Stone 500 Greatest Songs of All Time"
US Crude Oil Field Production, Lower 48 States (1,000 barrels)


Cigarettes smoked per person per year
Lung Cancer Deaths per 100,000 (Male, age adjusted)


OUR RELATIONSHIP ENTERED ITS DECLINE AT THIS POINT.
THAT'S WHEN YOU STARTED GRAPHING EVERYTHING.
COINCIDENCE!


Total US Highway Fatality Rate
Fresh Lemons Imported to USA from Mexico (Metric Tons)
$R^2 = 0.97$
Sources:
U.S. NHTSA, DOT HS 810 780
U.S. Department of Agriculture

Trend Lines

Extrapolation is an important component of science. It involves taking a trend shown in data and projecting it forward to predict the future. This technique is used, for example, to calculate the date when the US will run out of domestic oil reserves or when the global population will reach 7 billion.

15. You were using this technique when you estimated the average life expectancy of a US adult in the year 2050 earlier. What assumption(s) did you make in doing that extrapolation?

Extrapolation, sometimes called regression, has its limitations, and is often applied in a misleading way in the media. You should always consider how reasonable extrapolation of a certain trend is and how long it is likely to stay valid before other factors come into play.

16. As an example, can you see anything wrong with the extrapolation shown below? Why or why not?