Math Skill #1: Graph Reading and Interpretation

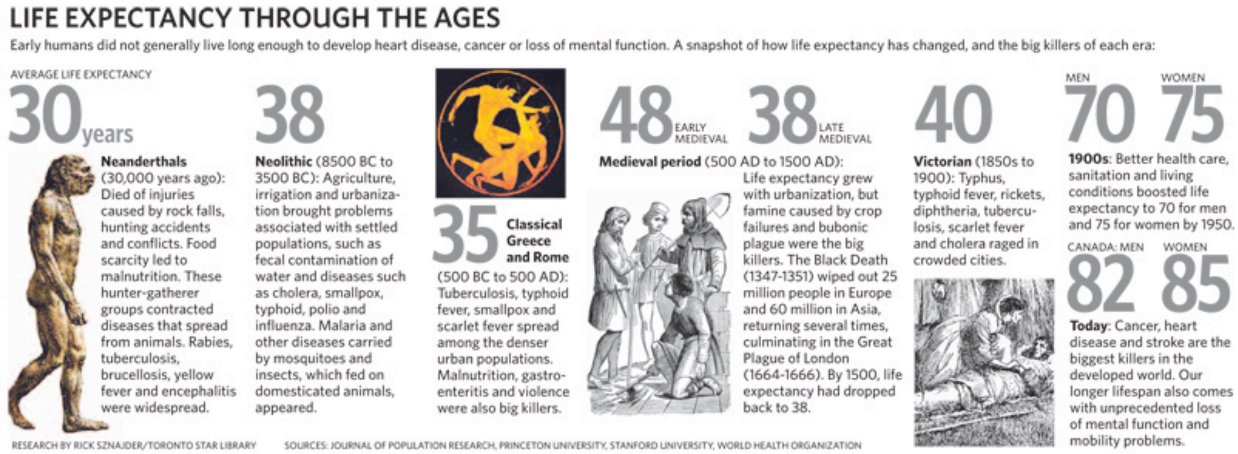
Many of your homeworks this semester will include a worksheet like this that focuses on a specific skill that we've discussed in class and that will be important throughout the semester. These skills were chosen because they are also important in (a) real life and (b) interpreting science in general, 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.

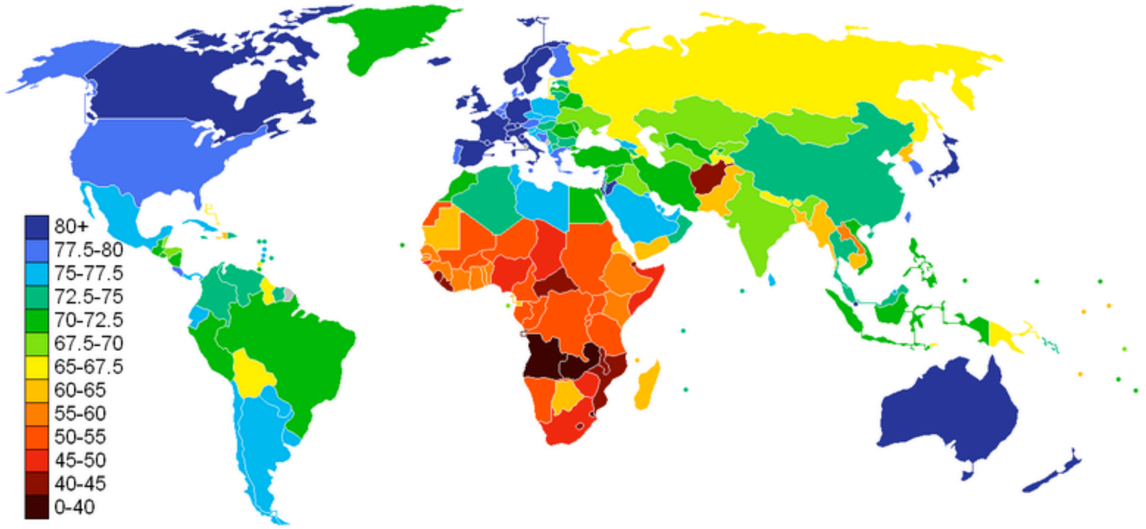
1. Why was male life expectancy higher in Neolithic times than female? Come up with a hypothesis and then explain what kind of data you might gather (from, say ancient archaeological sites or human fossils) to prove or disprove your hypothesis.

Any relevant hypothesis works, and people come up with a number of plausible ones. The best (and the real reason behind this phenomenon) is that without modern medicine, many women died in childbirth in Neolithic times. Ways to verify this might be to look for skeletal signs of childbearing, or women and infants buried close together. The goal here was to develop a reasonable and relevant hypothesis and come up with a measurable way that it could be verified. Remember good hypothesis are testable!

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?

People are often inclined to "skip over" the chart above and take a wild guess or say they don't know, but note that the table above is essentially giving you the information you need to answer this question when it says "Better health care, sanitation and living conditions boosted life expectancy...". In other words, make sure to examine all of the information you're given before you answer the question.

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 a lot of data. When confronted with a graph, your first thought should always be: what is this graph really 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?

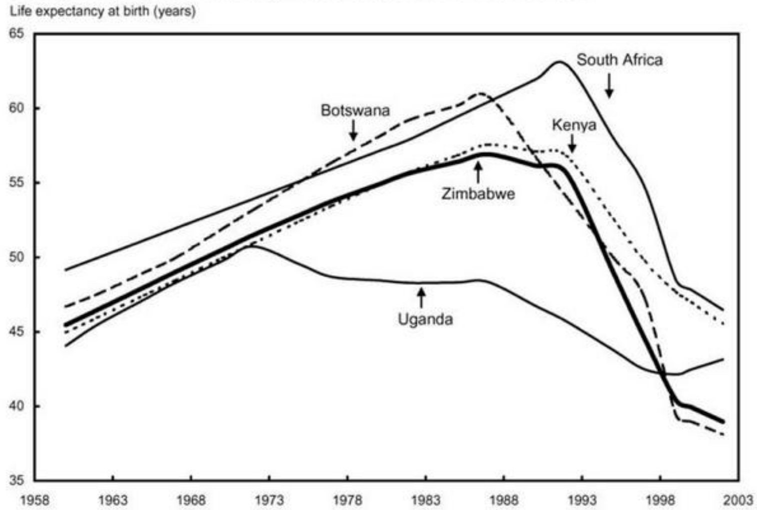
Most people get a little caught up in the details here, of which there are many. The obvious omission is simply that Australia, New Zealand and Japan are excluded from the list of longest-lived countries BUT recall that the statements says "concentrated", which implies multiple countries in a single area. If you keep going along the lines of individual countries excluded, you should include South Korea, Chile, Argentina, Iran, etc. and then your "summary" statements starts to become too long to be a real summary. Note too that the statement says that the shortest-lived countries are concentrated in Sub-saharan Africa, so Egypt, Morocco, etc. are not good examples for why this statement is false.

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

Because it is a big picture statement about a very complicated data set, there are any number of places where the statement falls short in describing the full truth, including the examples I listed above.

Chart 7-2 **Changes in Life Expectancy, 1960 to 2002**

Life expectancy has fallen dramatically in these selected hard-hit African countries.



Source: World Bank, *World Development Indicators*, 2004.

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* steadily from

   About *1958* to about *1988-1992*, then *decreased rapidly*

   (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?

   The decrease afterward is much more dramatic as revealed by the steepness (slope) of the curves.

(c) What may have caused this reversal?

Again, any relevant hypothesis works here. The real answer is the rapid increase in AIDS victims in the late 80s/90s.

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

Here again, any relevant data to support the hypothesis that you gave in (c) works! For example, statistics on cause of death.

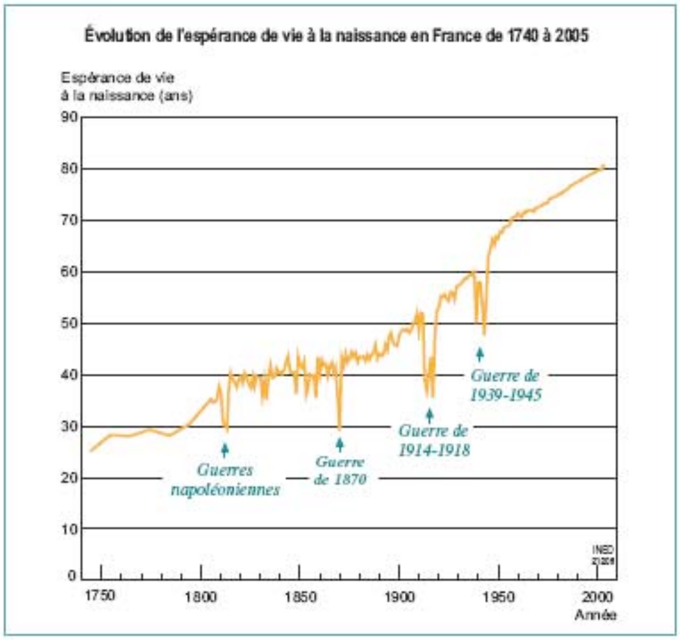
The property you were describing on the last page is called ***slope*** *and it means rate of change. Really, it is a judge of steepness in a graph, with steeper lines indicating more rapid change and shallower lines indicating more gradual change.*

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

All of then follow the same general trend EXCEPT for Uganda, which began its decrease much earlier.

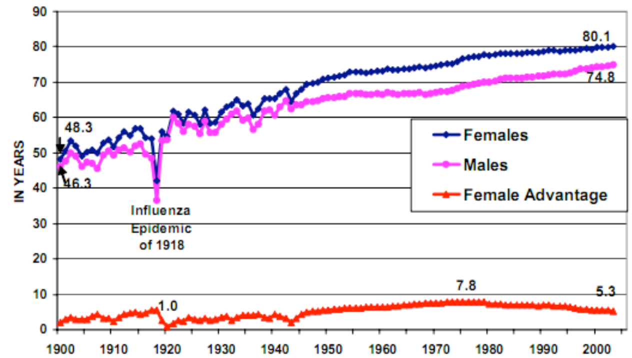
7. If so, develop a hypothesis as to what this might be. What additional data could you gather to verify or disprove your hypothesis?

   Again, any relevant hypothesis works. War, disease and famine are the classic causes of decreases in population, and in the case of Uganda, it was war and genocide that started the population in decline even before AIDS became widespread. A good way to test this (or any of the other) hypothesis would be to review news articles, primary sources, etc. to figure out what was going on in Uganda at the time.



Source: French National Institute for Demographic Studies

Figure 1. Life Expectancy at Birth, by Sex: 1900 to 2003.



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.

8. 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):

Keep in mind that the question asks for the "big picture", so this should only be a sentence or so!

US: *With one noticeable deviation due to the Influenza epidemic of 1918, Life expectancy in the US has been increasing steadily from 1900-2000, with women averaging slightly longer than men.*

France: *With four noticeable deviations due to wars, life expectancy in France increased steadily from 1750-2000.*

***Careful to look at the axes of the graph carefully. Many people are inclined to say that the increase in life expectancy in France is more rapid in general than in the US, but if you look carefully you will see that the time intervals are different, with the interval shown for US life expectancy data significantly shorter. When you compare apples to apples (what was the increase over the same time interval) you will see that they are nearly identical!*

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".

9. 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:	*written on graph, but average male and female values* 47.3	77.5
France:	47	80

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

US: 77.5-47.3=30.2

France: 80-47=33

11. 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: 0.30

France: 0.33

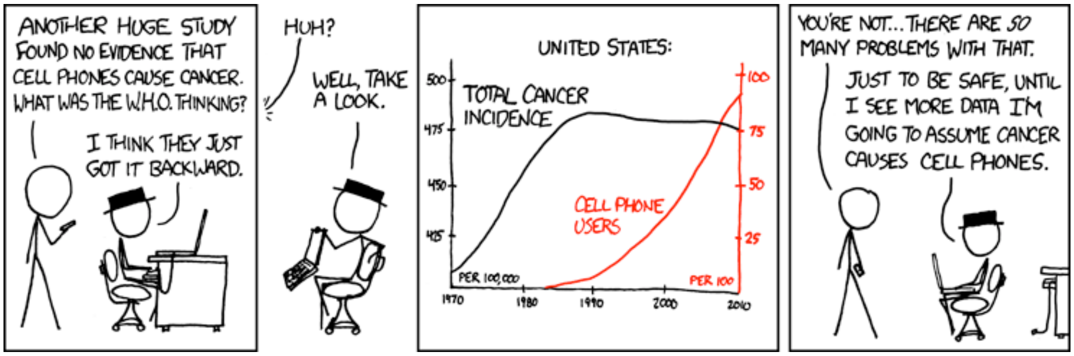
12. Is it possible for this change (your slope) to have been a negative number? Why or why not?

Yes, if there had been a decrease in life expectancy, the slope would have been negative.

13. 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?

The slope calculation you did above tells you that the average life expectancy in the US increases by 30 years over a 100 year time period. Thus, over half that time period (50 years), it should increase by half as much, or 15 years (assuming that the slope remains the same). Add 15 to 77.5 (average US life expectancy in 2000), and you get 92.5.

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.

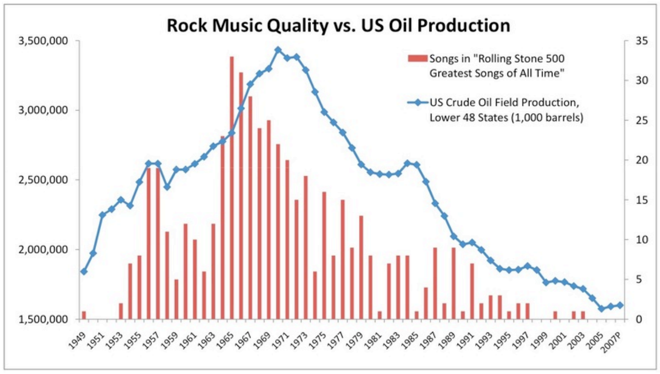
Two quantities that are correlated track each other, in the sense that as one increases so does the other OR as one increases the other decreases proportionally. If two quantities are correlated they will show similar trends over the same[1] *interval of time.*

"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.

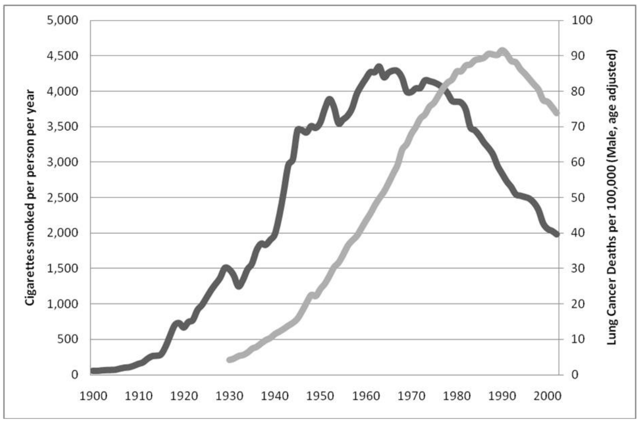
Examine the trends shown in the following four graphs. All four graphs show a correlation, but only two have a legitimate claim for causation. In the space to the right of each graph, write a one-sentence explanation of the trend being shown (your one sentence "big picture" explanation of the graph)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

---

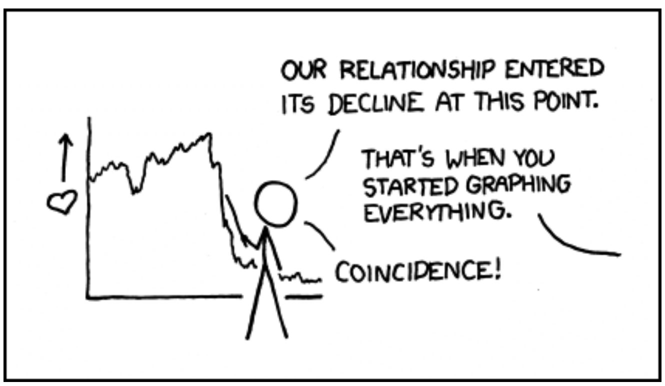
[1] Note that it is also possible for one quantity to lead or lag the other in time. For example, a "leading indicator" of the number of new houses in the US could be the number of applications for building permits for new homes. An example of a "lagging indicator" is unemployment, which changes only 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.



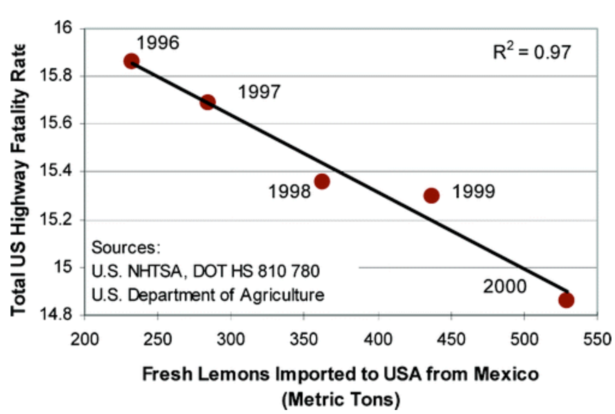
This graph shows that US crude oil production rose and fell with the number of songs released each year among Rolling Stones' "500 Greatest Songs of All Time". There is no conceivable reason why these two should effect one another, so this is a correlation, not a causation.



This graph shows that the consumption of cigarettes rose from 1900 to around 1970, and has been falling since, while the incidence of lung cancer followed the same pattern, but trailing by trailed by approximately 25 years. These two things can and should be related to one another, as your intuition should tell you, so this is not only a correlation, but a causation. The delay between the two is due to the time it takes lung cancer to develop, thus lung cancer is a trailing indicator of cigarette consumption.



Yes, this graph is silly. If you look carefully at the caption though, you can see that it is a plot of love (represented by the heart on the y axis) as a function of time. In particular, the graph suggests that the tendency of the character to graph things is the cause of this trend. This is a causal relationship.



This graph shows an inverse relationship between the number of fresh lemons imported from Mexico and the total US fatality rate. In other words, as more lemons are imported from Mexico, it shows that the US highway fatality rate falls. If you think about this for a moment, it should occur to you that it is bogus. Why should the highway fatality rate DECREASE when the number of lemon trucks on the road INCREASES. Lemons trucks are not likely to be a high proportion of the trucks on the road either. This is purely a correlation.

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.

14. 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?

You assumed that the trend would continue undisturbed. In mathematical terms, you assumed that the slope would remain constant. This is probably a reasonable assumption considering that this slope is similar to the 300 year trend in France and that the US data shows an increase over the last 100 years. We don't yet know if there is a "ceiling" to the increase in human life expectancy.

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.

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

You should! The line is a pretty nice fit to the data, but to assume that it will continue this long is, in this case, not reasonable. There is a limit to the speed that the human body can achieve!