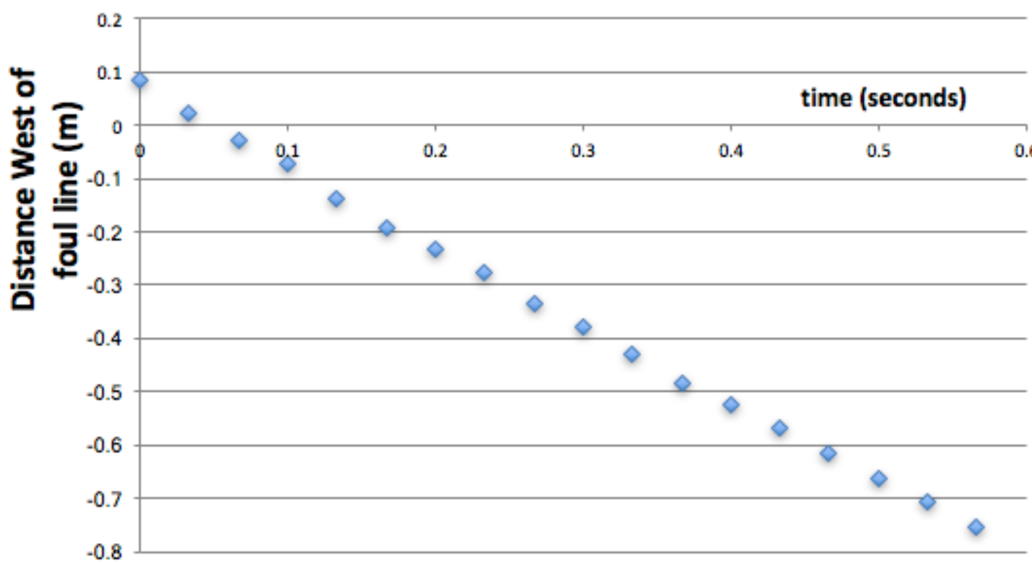
Graphs / data visualization

Names:________________________________

[In addition to writing your names above, take a *picture* with your iPad of you and your partner(s) and place it in this space.]

Imagine that you release a bowling ball just before the "foul line" (see the picture) and it rolls away from you down the lane towards the pins. The graph below represents the position of the bowling ball vs time that could have been acquired by analyzing a video.

Graph terminology



Some parts of this graph include….

The **x-axis:** the horizontal line showing what the horizontal distances on this graph mean. It has a **label,** "Time", and **units**: "seconds".

The **y-axis** is the vertical line showing what the vertical distances on this graph mean. It has a **label**: "Distance West of foul line", and **units**: "m" which is short for *meters*.

When the *x*-axis is some kind of time, this kind of graph can be called a **trend** graph.

1.) Estimate from the graph: At about what time was the ball -0.6 m West of the foul line?

Most of the distances shown in the graph are *negative*. Answer the following questions to get at what it means to be "-0.6 m West of the foul line" on this graph:

2.) The label of an axis always tells you what *positive* numbers mean. At t =0 seconds, when the graph starts, the ball has a positive position. So does it start on the West side of the foul line or on the East side of the foul line?

3.) A little later the ball's position is exactly 0. What does that mean about whether it's West or East of the foul line?

4.) So, assuming that the ball keeps moving, for times after its position was 0, which side of the foul line is it on? (and is its position positive or negative according to the graph?)

5.) Putting all of this together: If the ball is "-0.6 meters to the West of the foul line", a different way of giving its position would be to say:

The ball is +0.6 meters ______ of the foul line.

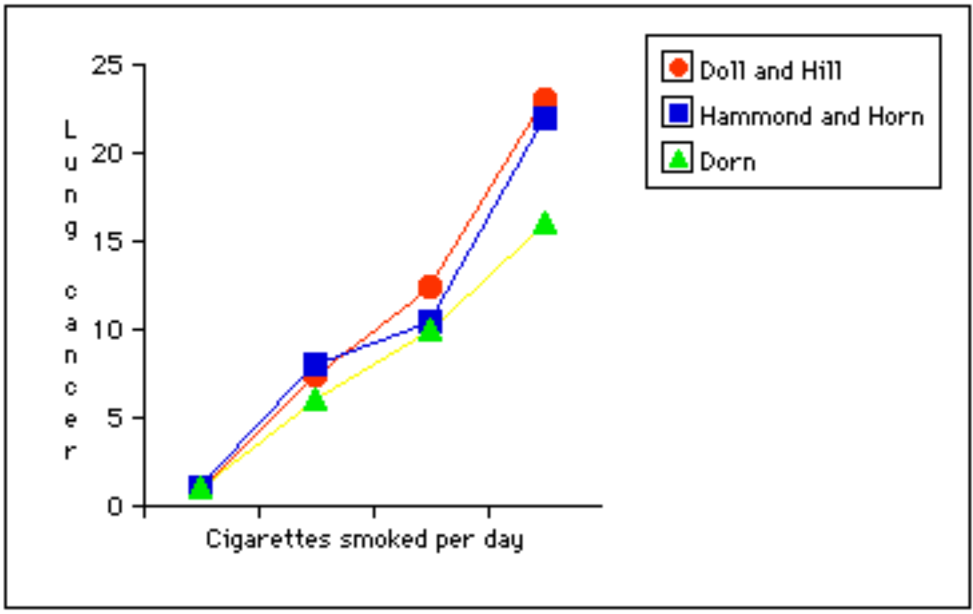
6.) Is the ball moving towards the East or towards the West?

7.) I asked you to watch that Kurt Vonnegut video and read what Robert Krulwich wrote (**http://tinyurl.com/cindyq**). What labels did Kurt Vonnegut put on the x- and y- axes of his story graphs? (He may have called them quantities, let's call them "labels"…)

8.) How might you measure what Vonnegut put on his y-axis? Does this suggest any particular **units** to use? [Be creative!]

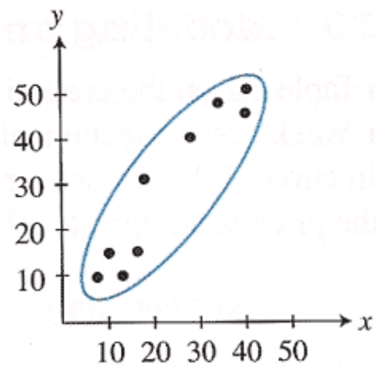
The bowling ball distance graph above shows a "**negative correlation"**: as the quantity (time) on the x-axis increases, the quantity on the y-axis (distance) decreases.

The graph at right shows a **positive correlation**: as the quantity (number of cigarettes) on the *x*-axis increases, the quantity on the *y*-axis increases. [I think the y-axis labels and units are something like "lifetime chance of contracting lung cancer (percentage)".\

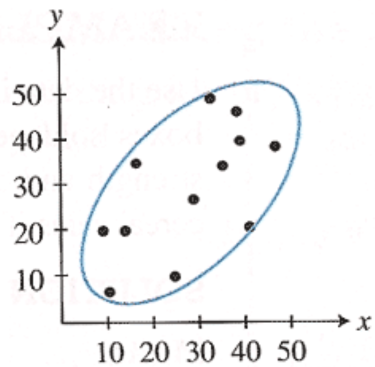


If you suspect that something (e.g. cigarettes smoked per day) **causes** something else (e.g. lung cancer), then the **suspected cause should be on the *x*-axis**, and the suspected effect goes on the *y*-axis.

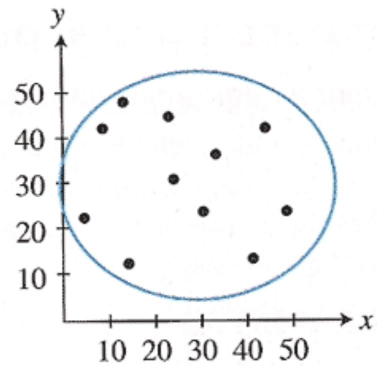
Graphing data can quickly show if there is positive or negative correlation between two quantities, or if there is perhaps *no correlation*. (There are also correlations that are not *linear*, but we need not concern ourselves with those right now.)



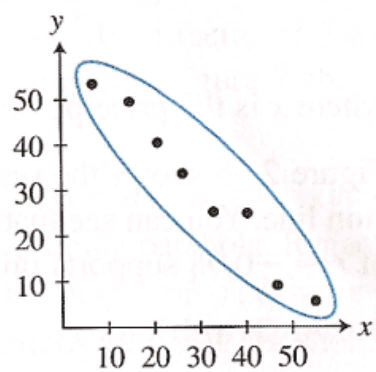
Strong positive linear correlation



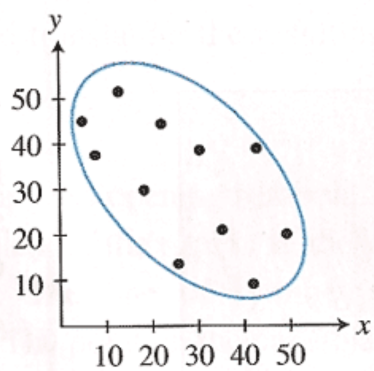
Weak positive linear correlation



Little or no linear correlation



Strong negative linear correlation



Weak negative linear correlation

9.) Think of a pair of quantities that are *positively* correlated. What label, and what units would you put on your graph's *x*- and *y*-axes

10.) Think of a pair of quantities that you think are negatively correlated, and repeat the exercise above.

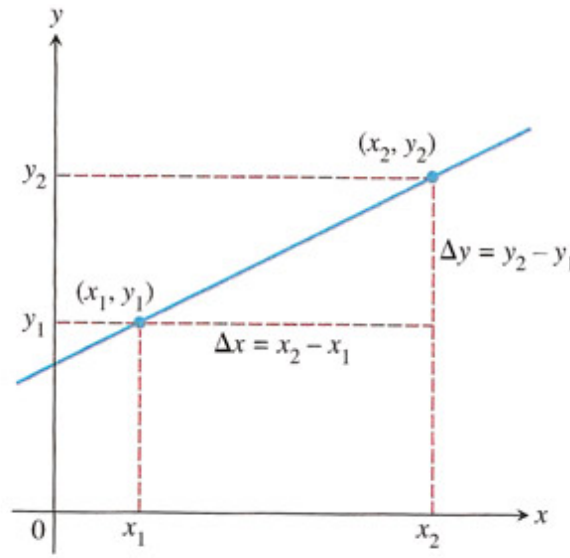
Slope

The **slope** of a **line** is the *ratio* of vertical change to the horizontal change. For example the slope of a road that rises 10 m for every 200 m forward is (multiply a decimal ratio by 100 to get %):

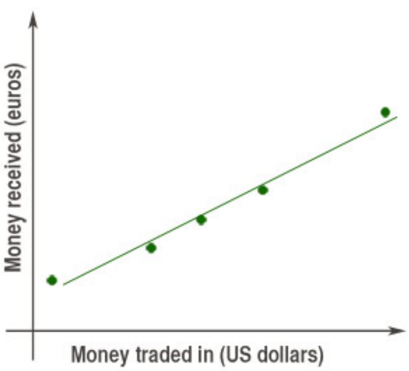
$$\frac{10\,m}{200\,m} = 0.05 = 5\%$$

If the coordinates of a point are written as (x,y), then you can use the "rise over run" formula to calculate the slope from any two pairs of coordinates of points on the line.

$$slope = \frac{y_2 - y_1}{x_2 - x_1}$$



Units of slope



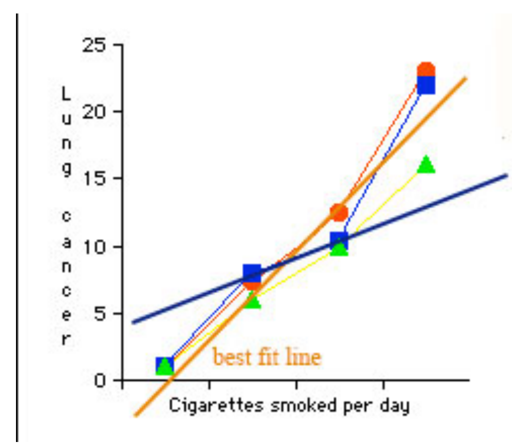
The units of slope are the *y*-units divided by the *x*-units. For example, the graph at left shows several (imaginary) currency exchanges that I made on an (imaginary) European vacation. The units of the slope are Euros / Dollars, or Euros **per** dollar. This particular slope has a special name: it is the average *exchange rate*.

For trend graphs, the slope represents a *rate of change*. or a "growth rate". If I graph the number of hairs on my father's head over time, with units of years, then the slope of a line would have units of *hairs per year*, which is a rate of change of hair. (It could be positive or negative, depending on his age!)

Choosing the points to find the slope

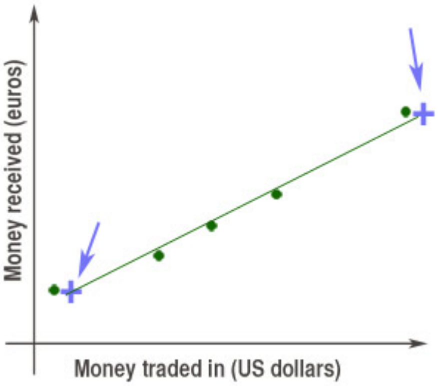
The line in the exchange rate graph is an example of a **best-fit line**. It does not necessarily touch *any* of the data points, but tries to capture some sort of "average behavior" for data that has some variation. There should be about as many points above the line as below the line.

In the "Lung cancer" graph, the orange line is a good "best fit line" for the blue data. It's reasonably close to each of the blue squares. The blue line connects two of the data points, but is a poor best fit line: It's very far away from the first and last data points. In general, your line need not actually touch *any* of the data.



Sometimes we'll use software (Logger Pro or Desmos) to find the best fit line. But you should always visually check that a line gets as close as possible to as many data points as possible.

Here are some ways to calculate slopes:



By hand: You shouldn't use the data points themselves to calculate the slope of such a line. (In this example, none of the data points are even on the line) Instead, make two new points (the crosses shown) that are

- On the line, as precisely as you can make them, and
- As far apart as possible from each other. Estimate the coordinates of these two new points and use those to calculate the slope.

Wolfram Alpha: Use 'linear fit (1.2, 3), (2.4, 6)….' Each parenthesis is the (*x*,*y*) coordinates of one data point.

Desmos: If you've got data in a table with columns labelled x_1 and y_1, then a new cell with:

$$y1 = m * x1 + b$$

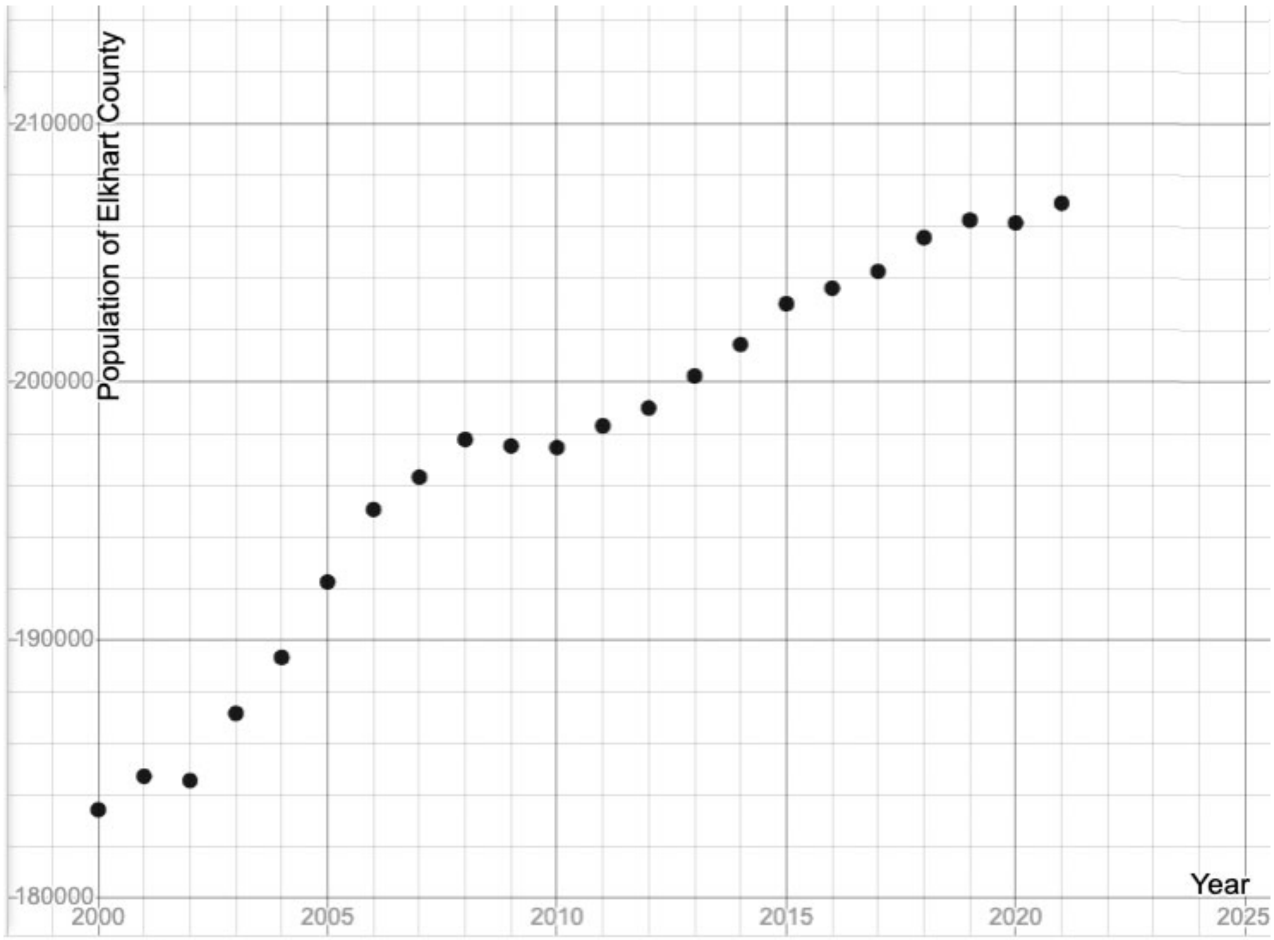
11.) For the data below, you can visually position a straight line with Notability's pencil:

1. Choose the pencil tool.
2. Then, Finger down, drag for a long line, wait until it turns into a straight line.
3. Use the scissors (select) tool to draw a circle around any part of the line to select it,
4. Grab your selection with two fingers and you rotate and move the line.

Mark the two points where your ruler/line crosses the left and the right edge of the graph. (You want points on the line that are as far from each other as possible.) Estimate the coordinates of the two points, and write those on the graph.

Use the coordinates you found to calculate the slope of the line that connects your two points. (Use the slope formula above.) Include your interpretation of what this slope *means*.

[Data from census.gov via datacommons.org. See https://www.desmos.com/calculator/oewuirujnp for a desmos graph displaying this data.]



Population of Elk. County (use the space below to show your calculations)

Slope of best fit line= (include units)

Write a sentence interpreting what this slope number *means*.

11.) Go back to the bowling ball graph. Again, use your Notability "ruler" and mark points at either edge of the graph where your ruler crosses for your "fit by eye" line. Calculate the slope of the line that connects those points. (Since this is a negative correlation, the slope should also be a negative number.) Give your number appropriate units.

12.) For this particular graph, the units of slope are [distance]/[time]. What is another name for this quantity? (Hint—what if you measured distance in miles and time in hours and then used the word "per" for the symbol "/"?)

Any comments / suggestions / questions for this lab?