- When doing SLR, we're often interested in 1) Determining if there is a linear relationship between these two variables and (if there is indeed a relationship), 2) Modeling their relationship to see how raising the value of one variable might affect the other variable.
- The two variables are denoted as X (Predictor variable) and Y (Response variable) and are plotted against each other on a scatterplot. Sometimes, it doesn't matter much which is which, but in experiments, the predictor (X) variable is what we think *might* be the causal agent.

- An Example! Time Outside vs. Time on Social Media
  - Open up the PhET least squares regression 1.1.4 simulation (try a Google search).
  - Play with the sim for a few minutes.
  - After playing around, think about the following scenario: 15 college students are asked how many hours a week they spend on social media and how many hours a week they spend doing outdoor activities (walking, playing sports, getting some sun, etc.). Use the scatterplot to predict some of the responses you might get.

  - Do you think there is a relationship between these two variables? If so, how would you describe the kind of relationship these two variables have?

  - How strong a relationship do you think they have?

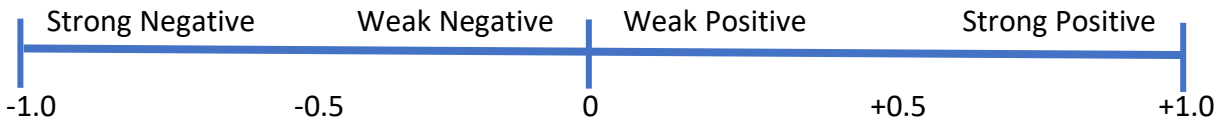
  - Reveal the "correlation coefficient" box. Try moving dots around and see if you can figure out what the correlation coefficient measures. Why does it change? What is the highest and lowest value it can be?

- Defining some terms

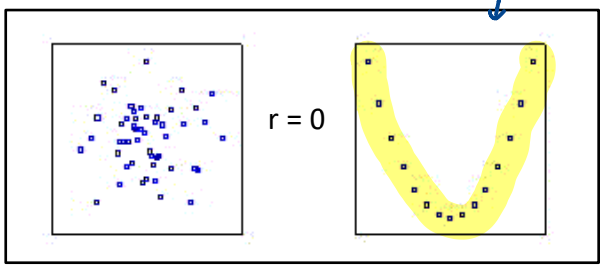
  - **Positive Correlation**: As one variable increases, the other variable also increases (directly proportional).
  - **Negative Correlation**: As one variable increases, the other variable decreases (indirectly proportional).
  - **No correlation**: There is no relationship between the variables (for this section, we are specifically looking at linear relationships.
  - **Correlation Coefficient**: Measures the strength of the LINEAR relationship between two variables (from -1 to +1).

Measuring a Linear Relationship

- A Closer Look at Correlation
  - The correlation coefficient is a measure of how strong and tight the linear correlation is between two variables.
  - The correlation coefficient is abbreviated r (for sample statistic) or ρ (for population parameter) and can take any value between -1 and +1.
  - Negative values imply that as one variable increases in value, the other decreases in value. (Negative correlation). Positive values imply that as one variable increases, the other variable increases as well (Positive correlation).



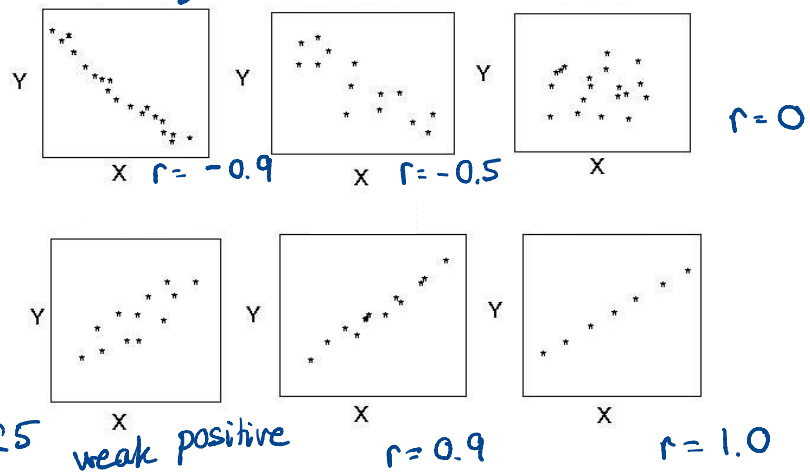
  - As touched on earlier, not all relationships are linear in form.
    - A low correlation coefficient signals no LINEAR relationship.
    - The graph on the right shows two scatterplots with r = 0.
    - The one on the left is a scatterplot for data that have NO relationship at all.
    - The one on the right is a scatterplot with NO LINEAR relationship, but some kind of quadratic relationship, but r = 0 for this as well.



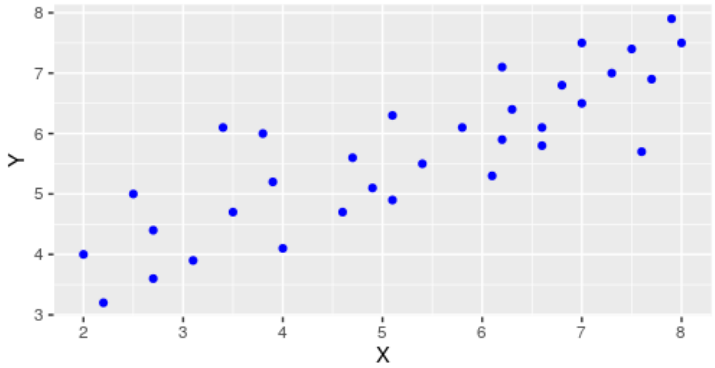
  - For this reason, be careful not to confuse correlation with measuring **ANY** relationship. We would need polynomial regression methods and a non-linear correlation coefficient to measure those relationships.
  - Interpreting correlation coefficients
    - Formula: $r = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})(y_i-\bar{y})}{s_x s_y}$ (you **won't** need to calculate r by hand!)
    - The formula just measures whether or not the x values and y values are "moving together."

Practice: Match the correlation coefficient with the scatterplot it represents.

r = -0.50
r = 0.90
r = -0.90
r = 0
r = 1.00
r = 0.50



Digging Deeper: Let's assume two variables X and Y are positively correlated, with r = 0.5. Which situations would change r?



Adding 10 to every X value. Would it change r? If so, how? (if in doubt, try drawing a picture or thinking about the formula again!)

Does not change r!

↳ correlation coefficient

Switching the Y values linked with the 3 lowest X values with the 3 highest X values

r ↘ decreases!

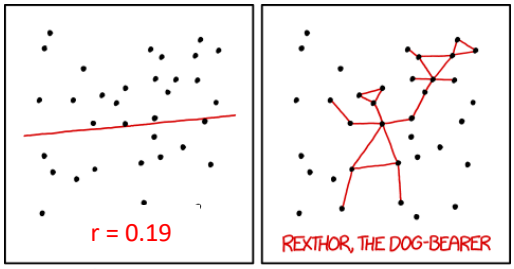
Switching the predictor and response variable around (e.g., which is on the X axis and which is on the Y axis).

Does not change r!

Multiplying all of the X values by -1

negative number

r = −0.5



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

For those of you who are competitive, you can make a game of guessing correlations here: http://guessthecorrelation.com/

- The Regression Line (**Line of Best Fit**)
  - Example: Consider the following study in which a store owner is interested in seeing the relationship between th of salmon and the number of sales (Y) for that week.
  - He conducted the study over 5 weeks with the price being held constant at a different price each week.