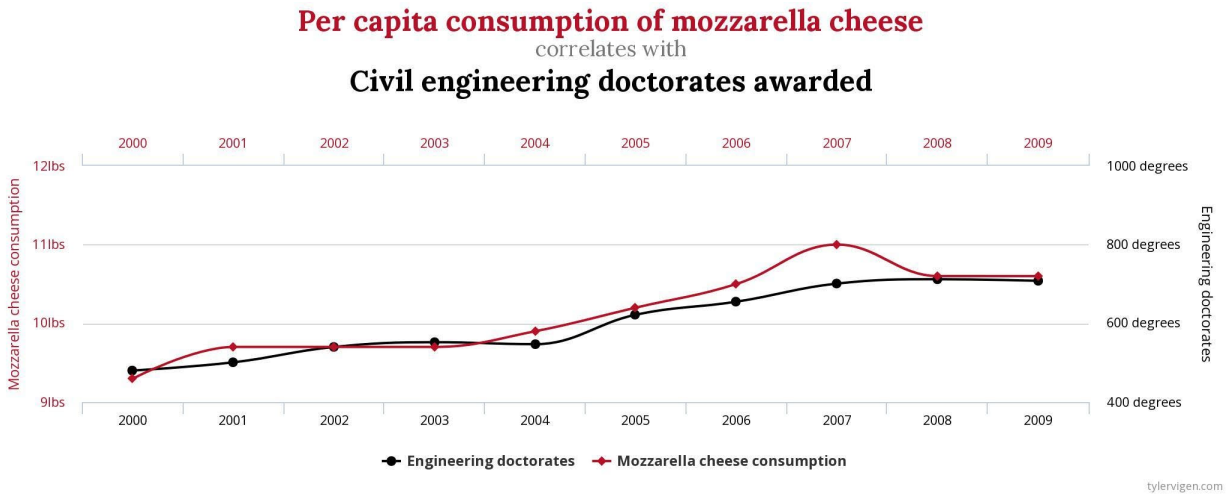
Correlation NOT Causation

Estimated Time: 30 minutes

SUMMARY: The amount of mozzarella cheese consumed in the United States is correlated to the number of Civil Engineering degrees awarded. Does that mean that the more string cheese you eat the more likely it is that universities will award more degrees in civil engineering? Try this activity to learn more about correlation and causation and how advertisers hope you will forget to think critically!



WHAT YOU'LL LEARN

- Recognize the difference between positive correlation, negative/inverse correlation, and no correlation.
- Recognize that causation is different from correlation.
- Look at data sets and determine if further variables need to be identified.

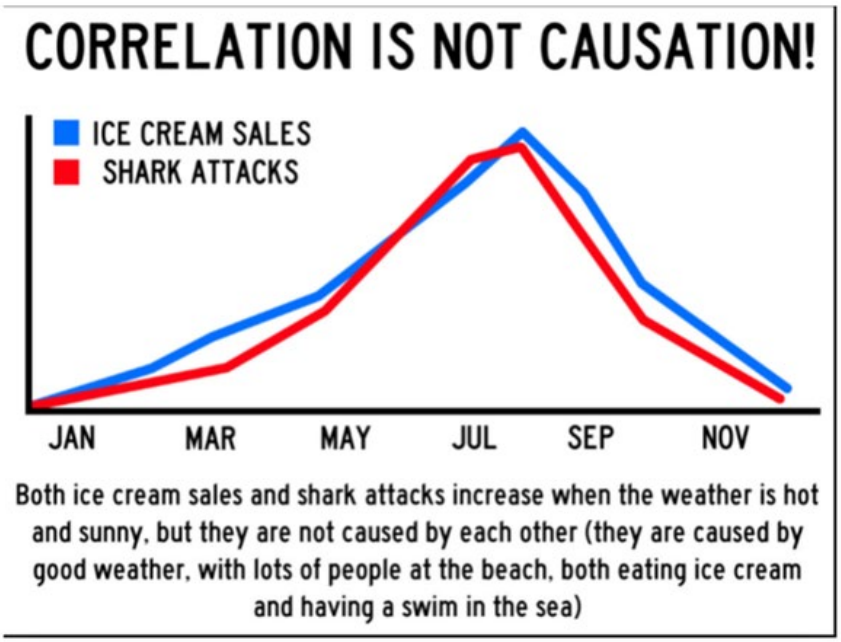
Materials Used

- Ruler
- Graph paper or plain paper
- Pencil with eraser
- Data

WHAT TO DO

1. In the above graph from http://tylervigen.com/spurious-correlations, you can see two data sets that have a positive correlation. This is a relationship between two variables that increase and decrease together. In other data sets, there might be a negative or inverse correlation so that when one variable increases the other decreases. Is saying things are correlated the same as saying that one thing causes the other? Careful, you don't want to commit a type 1 error!
2. Correlation strength is measured as a coefficient "r", with a range of -1.0 to 1.0. When comparing data sets, there are three possible explanations when there is a correlation:

1) dataset a is affecting dataset b in a positive way; 2) a negative way, or 3) not at all. Sometimes there is a hidden variable that affects both datasets the same way. Another example:



In this graph you see a correlation between ice cream sales and shark attacks. Does this mean that ice cream makes you taste more enticing to sharks? Or is there a third hidden variable that impacts both ice cream sales and shark attacks? (Hint, it has to do with the weather!)

3. Gather your own set of data to analyze. First, ask for adult permission to get out what your family has on hand for snack food. This can be chips, popcorn, candy, etc. Sort the snacks into three piles, salty, sweet, and other/both. Tally for each category and put the snacks away.
4. On your graph paper create a bar graph with the x axis as the snack category and the y axis as the number of snacks.
5. Now, ask everyone in your house their most common choice for snack. It has to fit one of the categories salty, sweet, or other/both. Be prepared to give examples and explain you are collecting data.
6. Graph your results just like you did the snacks, with the categories on the x axis and the number of responses on the y axis.
7. If possible, contact a friend or family member and ask them to do a similar pantry assessment. The more data you can collect, the better your graph will be. Be consistent in your categories so the data are comparable!
8. Compare your two datasets. Look for trends between them, like the more salty snacks in the pantry the more people in your household like sweet snacks.

9. If you notice a correlation, is it positive, negative/inverse, or no correlation? What are the other variables that might have an impact on your data? Can you design an experiment to test another variable? (Age, gender, activity level, etc.?)
10. Remember that correlation does not always equal causation.

TIPS

- For some additional spurious correlations, check out the website listed above. Be aware that some content may be of a mature nature.
- Conduct your snack data collection over a period of time and see what trend if any you can identify.
- Look at the news headlines, particularly the ones dealing with public health. Can you find examples of exaggerated causation?