Correlation vs Causation: Lesson Plan

Topic
Correlation does not equal causation. Correlation is a measurement of the strength and direction of the relationship between two or more variables. Causation indicates a similar but different relationship between variables, namely that one variable produces an effect on another variable or causes it. Data can be strongly or weakly correlated, which shows the strength of the link between them, as well as the strength of their predictive power. Positively correlated data means both variables increase or decrease together while negatively correlated data means that one variable increases while the other decreases. Just because data correlates does not mean there's a causal link between them: the variables could be influenced by a third, unknown variable or just randomly happen to correlate.

Possible subjects/classes	Time needed
Science, English, History, Philosophy, Psychology, Sociology, Government	30-45 minutes

Video link:
https://academy4sc.org/topic/correlation-vs-causation-the-missing-link/

Objective: *What will students know/be able to do at the end of class?*
Students will be able to... • Explain the four descriptors for how data correlates. • Identify how correlation can indicate about the relationship between two variables. • Explain the difference between correlation and causation.

Key Concepts & Vocabulary
Data point, relationship

Materials Needed
Worksheets, either print outs or a projection screen for scatter plot examples

Before you watch

Think-pair-share: Show students an example of a scatter plot that shows strong negative correlation and another that shows weak positive correlation. Have students group up and ask them to describe the relationship between the two variables for the two examples.

While you watch

1. Define causation.
2. What are scatter plots used for?
3. List four words that describe correlation.

After you watch/discussion questions

1. Why do you think people readily assume correlation indicates causation?
2. How can you prevent yourself from jumping to such conclusions?
3. Why is it important to differentiate between correlation and causation? What fields can be affected by the mistaken assumption that they're the same?

Activity Ideas

- Find some examples of correlating data (try Spurious Correlations). Have students work either independently or in groups to come up with possible explanations for why the data may correlate. It's okay if they get stuck - sometimes two unrelated data sets will match up! The purpose is to show students that correlating data doesn't necessarily indicate that the two variables directly affect each other, or even have anything in common.
- Look for popular articles that mention correlation between two variables. Does the article presume that one variable causes the other because of this correlation? Why? How could the article be changed to better reflect the truth about correlation?

Sources/places to learn more

1. Aldrich, John. "Correlations Genuine and Spurious in Pearson and Yule." *Statistical Science,* vol 10, no 4, 1995, pp 364-376. DOI: 10.1214/ss/1177009870.

2. Allen, John-Mark A, Jonathan Barrett, Dominic C. Horsman, Ciarán M. Lee, and Robert W. Spekkens. “Quantum common causes and quantum causal models.” *Physical Review X,* vol 7, issue 3, July 2017. DOI: 10.1103/PhysRevX.7.031021.
3. Altman, Naomi and Martin Krzywinski. “Association, correlation, and causation.” *Nature Methods*, vol 12, Sept 2015, pp 899-900. DOI: 10.1038/nmeth.3587.
4. Pearl, Judea. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 2000. ISBN: 9780521773621.
5. Wood, Christopher J and Robert W Spekkens. “The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning.” *New Journal of Physics*, vol 17, Mar 2015.