



## Simpson's Paradox: Lesson Plan

Topic	
Simpson's paradox is a statistical phenomenon that occurs when a correlation appears in several subsets of data but disappears or reverses when the subsets are combined.	
Possible subjects/classes	Time needed
Statistics, Psychology, Politics, Government	45-60 mins
Video link:	
<a href="https://academy4sc.org/topic/simpsons-paradox-when-correlation-does-not-equal-correlation/">https://academy4sc.org/topic/simpsons-paradox-when-correlation-does-not-equal-correlation/</a>	
Objective: <i>What will students know/be able to do at the end of class?</i>	
Students will be able to... <ul style="list-style-type: none"><li>• Explain Simpson's paradox.</li><li>• Recognize contradictory patterns in partitioned and aggregated data.</li></ul>	
Key Concepts & Vocabulary	
Correlation, partitioned data, aggregated data	
Materials Needed	
Worksheet, calculator (for the Worksheet), white/blackboard or projector	
Before you watch	
Recreate this graph on the board, without the lines intersecting the data points: <a href="https://upload.wikimedia.org/wikipedia/commons/thumb/f/fb/Simpsons_paradox_-_animation.gif/220px-Simpsons_paradox_-_animation.gif">https://upload.wikimedia.org/wikipedia/commons/thumb/f/fb/Simpsons_paradox_-_animation.gif/220px-Simpsons_paradox_-_animation.gif</a> . The graph should show separate subgroups with a positive correlation, but an overall negative correlation. Do not draw any correlation lines. Ask students what they think the correct relationship in the data is. Play devil's advocate, pointing out the apparent contradiction between the positively correlated subgroups and the negatively	



correlated data set.

### While you watch

1. What is Simpson's paradox?
2. Who were two preeminent scholars on Simpson's paradox?
3. In what types of research are Simpson's paradox most likely to occur?

### After you watch/discussion questions

1. Can you think of other real life examples of Simpson's paradox?
2. In what type of situation would using partitioned data make more sense than aggregated data?
3. In what type of situation would using aggregated data make more sense than partitioned data?

### Activity Ideas

- Imagine you are one of the researchers in the UC Berkeley gender bias study. You overhear that the university's administration is about to enact a policy based on incorrect statistics that would increase gender bias in graduate school admissions. How would you explain to the administration that their hasty actions are based on Simpson's paradox and convince them not to enact the new admissions policy? Write down your explanation.
- Create two lists of hypothetical or real-life scenarios in which using the correlations from partitioned data is favorable to aggregated data, and vice versa. Are there any noticeable similarities among situations where one type of data is preferred? Are there any noticeable differences between the two?
- Complete the Worksheet and then switch papers with a partner. Evaluate your partner's data and explain why it is or isn't a good example of the Simpson's Paradox.

### Sources/places to learn more

1. Carlson, B. (2017, January 7). Simpson's paradox. *Encyclopædia Britannica*. <https://www.britannica.com/topic/Simpsons-paradox>.
2. Malinas, G. & Bigelow, J. (2016). "Simpson's Paradox", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.) <https://plato.stanford.edu/archives/fall2016/entries/paradox-simpson/>.



3. minutephysics. (2017, Oct 24). *Simpson's Paradox*. Retrieved from <https://www.youtube.com/watch?v=ebEkn-BiW5k>.
4. Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2), 238-241.
5. Wang, B., Wu, P., Kwan, B., Tu, X. M., & Feng, C. (2018). Simpson's Paradox: Examples. *Shanghai archives of psychiatry*, 30(2), 139-143. <https://doi.org/10.11919/j.issn.1002-0829.218026>.