# Threats to Validity

Students pretend to be terrible data scientists who develop and support claims based on faulty sampling techniques (selection bias, bias in the study design, poor choice of summary data, and confounding variables).

## **The Issue**

It seems like “Fake News” is on the rise today. People share conclusions in the media – including social media – that may be a misrepresentation of the data. Graphic representations of data can often be misleading. As consumers of information, we must carefully consider whether the data is being accurately represented. This project will help us to become aware of how data can be misconstrued.

## **Objectives**

Students will investigate four types of threats to validity by pretending to be “bad data scientists” who fail to consider the impact of selection bias, bias in the study design, poor choice of summary data, and confounding variables.

## **Phases of the Project**

### **1. Decide on a question.**

You and a partner will choose a statistical question that you would be interested in exploring. (Note, you will not actually be investigating this question in full. Rather, you will be developing a faulty plan to answer the question.) Be sure that your question is one that could be answered by closely analyzing data, which also lends itself to many threats needing to be addressed.

### **2. Develop a faulty research plan.**

You and your partner will develop a faulty plan to research your statistical question. Remember, your goal here is to use data to misconstrue and mislead. Be sure to describe in detail how you will incorporate each of the following threats to validity.

* Selection bias. What is your plan to gather information from a non-representative sample of the population?
* Bias in the study design. What sorts of “loaded” questions will you include in order to misrepresent true opinions?
* Poor choice of summary data. What sorts of extreme outliers will you include in your dataset so as not to represent the population as a whole?
* Confounding variables. How will you ensure that you overlook factors that might influence a relationship?

### **3. Analyze.**

Explain how the validity of your conclusions will be impacted based on the threats you allowed in your sampling. Explain how you could change your study to minimize these threats

And finally, share your results! You will develop a presentation (poster, Google slides, etc) that outlines: your research question; your faulty research plan (including a discussion of each of the four threats); your analysis of the validity of your conclusions; and a discussion of how to minimize threats.

Have fun with this. Pretend to be a terrible data scientist - costumes encouraged!

(Developed by Joy Straub.)

*These materials were developed partly through support of the National Science Foundation, (awards 1042210, 1535276, 1648684, and 1738598).  Bootstrap by the* [*Bootstrap Community*](https://bootstrapworld.org/community/index.shtml) *is licensed under a* [*Creative Commons 4.0 Unported License*](https://creativecommons.org/licenses/by-nc-nd/4.0/)*. This license does not grant permission to run training or professional development. Offering training or professional development with materials substantially derived from Bootstrap must be approved in writing by a Bootstrap Director. Permissions beyond the scope of this license, such as to run training, may be available by contacting* [*contact@BootstrapWorld.org*](https://www.bootstrapworld.org/materials/spring2021/en-us/courses/reactive/mailto:contact@BootstrapWorld.org)*.*

Last updated 2021-12-09 14:35:50 -0500