Project: When Data Science Goes Bad

(Also available in [Pyret](#))

Students 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. *This project supports the learning goals of [Threats to Validity](#).*

Lesson Goals	Students will be able to… • consider how threats to validity can emerge when research is conducted • gain awareness of how data can be misconstrued
Student-facing Lesson Goals	• Let's be "bad data scientists"!
Prerequisites	• [Introduction to Data Science](#) • [Threats to Validity](#)
Materials	• [PDF of all Handouts and Page](#) • [Threats to Validity Project Template](#) • [Lesson Slides](#) • [Printable Lesson Plan](#) (a PDF of this web page)
Preparation	• Students will work with a partner for this project. You may assign partners, or allow students to choose. • Decide what sort of final product you would like students to create. A slide deck template is included in the lesson, but any format will work. • Sharing is an important piece of this project. Think about if you have time for each group of students to share their work, or if you would prefer to use a different format. • This project is designed to span approximately 2-3 class periods.

Threats to Validity *flexible*

Overview

Students develop a faulty research plan on a question of their choosing to demonstrate their understanding of four unique threats to validity.

Launch

Congratulations! You're a data scientist who has been granted funds to conduct research on a topic of your choosing! There's one complication…

You're a **BAD** data scientist!

For this project, you and a partner will develop a statistical question that you would be interested in exploring.

Note: We recommend printing and distributing the student-facing Threats to Validity Rubric to help students understand the scope of the project - and your expectations - at the outset. Teachers are welcome and encouraged to edit and adapt the rubric for their unique classroom context.

- With your partner, review the four threats to validity to ensure that the two of you understand what each one entails.
- With your partner, agree on a statistical question that you want to pretend to study.
- Record your question on Q1 of Our Research Plan.

As a class, invite students to share their questions and begin to consider how they might incorporate the threats to validity. You'll want to ensure that all students have developed statistical questions before moving onto the next portion of the project.

- With your partner, develop a brief research plan and record it on Q2 of Our Research Plan.
- Complete the rest of the page, describing in detail how you plan to misrepresent your data.

Investigate

It's time to do some analysis!

- Turn to Analysis of Research Plan Predicted Outcomes.
- For each threat, explain how the validity of your conclusions will be impacted.

- Describe what you could do to minimize each of the four threats.

Finally, students create a presentation highlighting the (very bad!) choices that they have made.

If you'd like, you may opt to have students conduct a peer review of the pages they have completed before proceeding to their final draft. The Threats to Validity Rubric is useful for this.

- Create a presentation of your plan!
- Check the Threats to Validity Rubric to ensure you've included all of the required components.
- Use this Threats to Validity Project Template to get started - but feel free to personalize it and make it your own.

Synthesize

- A fun way to conclude this project is to invite each pair to present their flawed research plan to the rest of the class. **If using the Threats to Validity Project Template, each group will present only their first three slides.** Then, invite the rest of the class to discuss and critique the (hopefully very flawed) research plan.

- Have fun with the sharing or this project - and encourage students to be silly as well! Maybe you want to encourage costumes! Maybe the class is a panel of very serious and critical scholars judging the work of each group.

- If you can't carve out time for every group to give a presentation but want to give students the opportunity to practice identifying threats, you could pair up groups. Each group would present their research plan, and then create the other group's research plan.

- Be sure to ask students if they notice any additional threats to validity that emerge. Can they think of any other reasons why we might mistrust data?