Final Project Guidelines and Suggestions

Causality, Decision-Making, and Data Science
Fall 2025

Overview

The final project should be in groups, ideally of size three (smaller groups and singletons are okay as well). **You can choose your own groups.** For your project, you will identify a question you want to answer, a data set that might help you answer it, and you will use techniques from this class to try to answer that question with that data. Deliverables at the end of the quarter include a write-up (5 to 15 pages) and an in-class presentation. In more detail, here is what your project should do.

1. Your project should have a clear causal question that your project tries to answer (although you may revise or replace your question after exploring the data).
2. Your project should find (or create) a data set. *There is a list of suggested data set sources at the end of this document to get you started, and we are happy to chat about finding data sets for a particular topic.*
3. Your project should do at least one thing to that data set that we saw in class in order to answer your question (identify a natural/actual experiment, compute ATE/LATE, run a regression, ...).
4. You should write up what you did, and the conclusions that you drew from it, and give a short presentation the class. (More details below).

Note: Empirical work is hard! You may not be able to satisfactorily answer your question with your data. That's okay! We just want you to *think critically* about what you could do to answer your question given the data you have, and about what your results do and do not tell you.

Timeline

There is more information about each of these steps below.

- October 20: Project ideas due (11:59pm, on Gradescope). Your project idea will be a few paragraphs outlining what data set(s) you plan to use, and what questions you want to ask/answer.
- By October 30: We will give you feedback on your ideas.
- By November 10: You should have had at least one "check-in" appointment about your project. *If you can't make regularly scheduled office hours, please schedule your check-in a week or so before the due date to make sure that we can find a good time to meet.*
- By December 1 (5pm), upload your slides to a Google Drive folder (the link will be on Canvas).
- December 2 and 4: Final project presentations (during class; about 10 minutes each, depending on how many groups there are).
- By December 8: Turn in your final project write-up (11:59pm, on Gradescope). Your project write-up will be 5-15 pages (instructor permission required to exceed 15 pages), and will present your data set, questions, methods, and findings.

Project idea

What to hand in on Gradescope by 11:59pm on October 20:

The answers to the following questions (a Word document or pdf is fine):

- Who are the members of your group? (Note: each group should submit only once).
- What question do you want to answer?
- What datasets do you think you might want to use?
- (Optional) Based on the class so far, what techniques might you try? *Note: This question is optional, since there are plenty more techniques that we'll learn in class later in the quarter! One point of having the project idea due so early is that as we learn new techniques, you can think about how they might apply to your project.*

There is a list of possible data resources at the end of this document for inspiration.

We are happy to chat beforehand about possible ideas! Please come to office hours, or send an email to the staff email list. We would suggest first coming up with a question/topic; and then contacting us. Then we can brainstorm ideas for datasets and inspirational papers, or if needed give suggestions on how to tweak the question – the constraint in empirical research is typically the existence of data, not the question.

If you want our help finding a dataset or references, we encourage you to reach out as soon as possible.

Project Check-In

Before November 11, your group should have met *at least* once with a member of the teaching staff, either in office hours or by separate appointment.[1] You don't need to hand in anything from the check-in, just have at least one meeting. We encourage you to meet as often as you want with us throughout the quarter, but you must meet at least once.

Project Presentation

Final project presentations will be in Week 10, during class. We will post more information about the presentation format once we know how many groups there are, but expect 5-10 minute presentations. **Update: Please aim for 7 minutes.** Some tips:

- *Practice* ahead of time to make sure that you are within the time limit!
- Your presentation slides will be due (uploaded to a google drive link) by *Monday Dec 1* at 5pm. We will compile them into one slide deck before the day of the presentation.
- We will post a presentation schedule and the google drive link by the end of Week 9; if you have a hard conflict with the time you are assigned, please reach out ASAP.
- Be very selective in what you choose to include in your presentation! It's better to present less material clearly than more material rushed.[2]

[1] If needed we may schedule extra office hours so that all groups can meet in the week before the due date.

[2] Your goal is not to convince that you are smart, we know that already. Rather, your goal is to teach us, and your peers, something!

Project Write-Up

Project Write-ups are due December 8, 11:59pm, on Gradescope. For your write-up, you should address the following items.

- **Question:** Clearly state your question(s), and explain why it is interesting.
- **Data:** Describe the data set(s) you chose to use.
- **Methods:** Discuss which methods you used, and why they are a good fit to answer your questions.
- **Results:** Clearly describe and present your results, including any relevant tables and graphs.
- **Conclusions and Discussion:** Explain what your conclusions are: What do your quantitative results imply about your question? What are some possible issues? (e.g., confounding variables, etc). Can you rule out these issues, or confirm that they are really issues?
- **Future Directions:** If you aren't completely convinced one about the answer to your question, what would help you be more convinced? (That is, what would your dream data set be, what would you like to try if you had more time or computational power, etc). If you are completely convinced, what other questions would be interesting to ask/answer?

We suggest that you organize your paper into six sections according to the above bullet points, possibly with an abstract/introduction at the beginning that gives an overview.

Grading

The grade breakdown is as follows:

- 5% Project idea (graded for on-time completion)
- 5% Project check-in (graded for on-time completion)
- 5% Uploading presentation slides on time (graded for on-time completion)
- 35% Project presentation (see criteria below)
- 50% Project write-up (see criteria below)

Written report criteria

The report will be assessed on the following criteria:

- **Technical content:** does the report make technical sense and does it convey a good understanding of the material? Note that there is *not* a requirement to do any "fancy" technical stuff, just that any technical stuff you do be well-done.
- **Non-technical content:** One focus in this class is how to interpret data and the results of data analyses. Your report should contain a thoughtful discussion of how you interpret your results.
- **Lucidity of exposition:** We want to enjoy reading your paper! Is the question clear and well-motivated? Are your methods clearly explained? Are your findings clearly stated? Are your conclusions clear? Is the paper well-organized?

Presentation criteria

The presentation will be assessed on the following criteria:

- **Teach your peers something:** The main criteria for the presentation is that you teach your peers (and hopefully the teaching team) something!
- **Polish and practice:** Is the presentation well-prepared, taking into account the time limit?

Possible sources of data

Here are some ideas for sources of datasets:

- Integrated Public Use Microdata Series (`https://www.ipums.org/`). This website hosts many different kinds of data. In general, it hosts "census" type information, but it also hosts many other kinds of information.
- World Value Survey (`https://www.worldvaluessurvey.org/wvs.jsp`).
- Inter-university Consortium for Political and Social Research (`https://www.icpsr.umich.edu/web/pages/ICPSR/index.html`). This organization hosts many, many datasets from many, many studies on many topics.
- Panel Study of Income Dynamics (`https://psidonline.isr.umich.edu/`).
- UCI Machine Learning Repository (`https://archive.ics.uci.edu/datasets`). This repository has data sets about all sorts of things, from heart disease to a dataset made up of images of dried beans.
- Data is Plural (`https://www.data-is-plural.com`): Weekly newsletter with many"useful/curious" data sets ranging from disability claims processing to NYC evictions to trash balloons transported from North Korea to South Korea.
- American Economics Association Website (AEA) (`https://www.aeaweb.org/about-aea/committees/economic-statistics/data-resources`): This website contains links to a lot of relevant economic data sets.

Note!

The earlier you reach out to us, the more that we can help you!