

Final Project Grading Rubric

Data 102: Data, Inference, and Decisions, Fall 2023

Due Date: Monday December 11, 2023 at 11:59 PM

The final project is worth 15% of your overall grade. Each section will be graded on a 3-point scale, where the scale for each section is described on the next page. In general, the scale for each section will be based on how well you answer the questions posed in the project specification.

Each section of the project will be weighted as described in the project specification. In general, a high-scoring project will achieve all of the following:

- Critically evaluate data and discuss limitations of the dataset and what conclusions can be drawn from it
- Correctly apply the methods learned in class
- Demonstrate understanding of assumptions are being made by the methods used, and how to evaluate whether those assumptions are true
- Correctly interpret and summarize results
- Explain and justify any “bad” or “surprising” results. For example, we won’t take off points for poor prediction accuracy, a finding of no causality, or failure to reject null hypotheses, as long as you clearly explain why it happened.

An additional note on assumptions: if the assumptions made aren’t quite right but there is no readily available public data to meet them, and none of the methods taught in class can overcome this, then you should clearly state this and explain how that will affect the results (see “Assumptions” section of rubric). For example:

- Suppose you choose to use instrumental variables but your instrumental variable has a small direct effect on the outcome. If there is no better choice of method based on the available data (including additional publicly available data sources), then as long as you discuss this and how it might affect your results, you won’t be penalized.
- Suppose you choose to assume there is a linear relationship between two variables in your data, but the relationship is clearly nonlinear, and we’ve learned methods in class for dealing with this kind of nonlinear relationship. In this case, you will be penalized, since you could have used the more appropriate method.

Section	Exemplary (3)	Acceptable (2)	Inadequate (1)
Data overview	Clearly explains the context of the data (sample vs census, granularity, consent, etc.). Describes any biases or problems with the dataset and how those will affect the conclusions.	Provides some context. May have some errors in data description, or limited discussion of how the dataset circumstances will affect conclusions.	Does not provide any relevant context or discuss any limitations of the data.
EDA	Visualizations capture important trends and are relevant to research questions. Correct visualization types are used, axes are properly labeled, and color/size are used appropriately. Text clearly explains important discoveries or anything unusual in the data.	Several small issues are present in the visualizations, or one major error relevant to the research questions.	Multiple major errors. Visualizations are nonexistent or not relevant to the research questions.
Research question			
Question, algorithm, and modeling choices	Methods section clearly describes method used, along with any design choices (as described in project guidelines). Explains why the particular methods are appropriate for this question and this data.	Missing one major justification or method description, or one major error in justification.	Methods are not well-described or justified at all, or are not appropriate to answer the question with the data available.
Assumptions	Methods section clearly describes any relevant assumptions being made. Assumptions are appropriate for the data, and whenever they might not be (e.g., non-independence of null p -values, nonlinearity, missing confounder, etc.), clearly explains how and why.	Several small or one major assumption(s) is/are missing or incorrectly stated.	Assumptions are incorrect, missing, or not discussed.
Implementation and statement of results	Method is correctly implemented and results are stated correctly. All relevant results are in the report, and (where relevant) all statements of uncertainty are provided.	Several small or one major error(s) in implementation, or results incorrectly stated.	Method implemented incorrectly, results are copy-pasted with no context.
Interpretation of results	Clearly explains the results. Explains any “bad” or negative results (low prediction accuracy, no causal effect, etc.), or any “surprisingly good” results (prediction accuracy close to 100%, etc.). Where relevant, describes additional relevant data that would be useful. Describes limitations of results.	Draws some incorrect or unjustified conclusions, or has some major lack of clarity when explaining results.	Major misunderstanding of what the results mean.
Conclusion	Describes and justifies any real-world actions (or lack thereof) that are justified by the results. Summarizes results and describes limitations. Proposes at least one interesting direction of future work. Makes a call to action based on the answers to both research questions.	Makes some unjustified conclusions, or missing future work.	Discussion is irrelevant to real-world context.