

Final project structure

The goal of this project is to provide a rare opportunity to apply a new technique to your research or a topic of research (using open-source data) you are interested in. Opportunities for analytic exploration are few and far between so I encourage you to capitalize on this to ask research questions of your data that you may not be able to ask otherwise. Be creative! I am here to work with you along the way!

Your final project should include the following:

- Descriptions of data processing, cleansing steps and verification that final data tables are tidy compliant.
 - If data is unable to be processed within a jupyter notebook due to size limitations (e.g., fMRI) or software-specific constraints (e.g., SPM) please describe the steps with enough detail to follow what transformations are happening to the data.
- Please use at least 1 of the data analysis methods covered in class in a way that is outside what is typically used within your research domain (for example factor analysis, linear regression, classification, kNN, bootstrapping).
 - There is a lot of variability in how applicable this could be so I want you to use your best subjective judgment on what that would mean in your context. Otherwise, I am happy to chat through this with you.
- Clear visualizations of the results critical to your hypotheses using *ggplot2*.

Final notebook

The final project will be turned in and presented as a Jupyter notebook. Please start a new notebook file titled `NAME_DSPN_S26_FinalProject.ipynb` that sits in your Github repository for the class. Here, NAME is your last name.

Use the markdown formatting and Latex equations (if applicable) to show all models.

Your notebook should have the following subsections (with header titles) and content:

1. Title: Up at the top, a clear & creative project title. Can be a working title for now.
2. Background: Provide a paragraph motivating the project. Why should someone outside your field care about this research?
3. Variables: Specify all variables, their data type, how they were collected, and number of observations.
4. Hypotheses: Explicitly state your hypothesis in terms of the structure of the models that you will be evaluating (e.g., What is the form of $Y = f(X)$ for each hypothesis? Will you be using linear regression or classifier models?)
5. Data Organization: Provide descriptions of Data Architecture (e.g., how is the data organized), Data Cleansing & Tidying, and an example of the final of the Data Table(s) that will go into your analysis.
6. Analysis: Show both your data visualizations and summarize the results from your models.
7. Conclusions: A short (1 paragraph) conclusion with respect to the models you have run.

A link to the notebook with your data plan (in your class Github repository) needs to be submitted to Canvas by no later than 5pm on May 1st, 2026.