DS 2001
Final Project Guidelines

Project Teams - Due October 20th

If you would like to work with a team of 2-3 people on the final project, you will submit a short document listing the members of the team (two or three people) and the planned division of labor among you. The best teams will include people from different disciplinary backgrounds so you can leverage each other's specialized knowledge.

Project Proposal - Due October 27th

You will submit a 1-page document detailing a preliminary plan for your final project. You should include the following in your project proposal:

1. Identify a general question related to the social sciences that you plan to address in your final project. You should outline why this is an interesting or important question and describe why computational methods are necessary and/or helpful in exploring this question. If possible, explain how others have answered/attempted to answer this question using different methods.

2. Identify the data or collection of material you will use to explore this question, and briefly describe why the data/material is appropriate. Additionally describe how you will collect the data/material and whether or not it will need to be cleaned prior to analysis. If possible, import your data and provide a glimpse of its format.

3. Describe the techniques you expect, or would like, to use to analyze the data/material and explore your question. Why these techniques and not others? What kind of evidence will these techniques produce, and how will this help you answer the question and persuade others of your answer? If you already have some preliminary analyses, include these as well.

4. Briefly discuss any data visualization or interpretive techniques you will or would like to use to present your findings and convince others of your interpretation.

Project Notebook - Due December 8th

You will submit a Jupyter notebook detailing your final project. Project notebooks combine written analysis with data analysis and visualization in Python. You should expound on the points in your project proposal, present your results, and synthesize your findings. Successful projects will likely include the following:

1. A general question related to the social sciences that you are addressing in your final project. You should motivate your project by describing why this is an interesting or important question and why computational methods are useful in exploring this question. If appropriate for your project, you could mention how others have approached this question using different methods.

2. A description of the data you use to explore your question and the reasons this data is useful for doing so. You should explain how you collected the data and any important choices you made in the process of cleaning and preparing the data. Please include an overview of the dataset itself including such things as what constitutes a record, the number of records, and a discussion of the relevant variables. If appropriate for your project, you could note any data or measurement issues to keep in mind.

3. Analysis of the data using computational techniques in Python and the results of this analysis. You should describe the techniques you are using and the results these techniques are producing in the context of the question you are exploring. Please include descriptive (summary) statistics for key variables. You will likely want to analyze relevant relationships between variables or differences across relevant subgroups. If appropriate for your project, you could identify outliers and discuss their significance or employ statistical or machine learning models (e.g., linear regression, random forests) and explain their outputs.

4. Interpretation that relates specific results from your analysis to the broader question you are exploring. You should discuss the takeaways from your analysis and how your findings contribute towards an understanding of your topic. Where visualizations aid in interpretation, please use effective annotations - axis labels, titles etc. If appropriate for your project, you could discuss the limitations of your work and what might be done in the future.

Final Presentation - TBD

You will present your final project to the class. Project presentations will be 10 minutes long. This can be done in-person or on a zoom video. One of your project responsibilities is to attend the presentations (remote or in-person) You should motivate your project, describe your analysis, present your results, and synthesize your findings. Please do not include code.