IST 652 – Final Project Description
Spring / 2021

The final project for IST652 involves locating an open data set of interest, formulating an inquiry or set of inquiries that could be addressed with the data, importing the data set into Jupyter Notebook, and conducting some analyses on the data to illuminate the inquiry. The project focuses on open data in order to ensure that your chain of transformations and analyses is reproducible.
The objective of the final project is to demonstrate your ability to write Python scripts to access and amass data in one or more of the three types of data studied in the course and to prepare and use data to produce data summaries, lists and other structures. You must process the data wholly with Python code in a way that requires no manual intervention (or use of Excel, or other data programs).

Steps and Milestones

1. Select a dataset or a combination of datasets for your project Many data sets are available at sites such as the World Bank (http://data.worldbank.org), the U.S. Federal Government (http://www.data.gov), ICPSR at the University of Michigan (http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp) - other potential sites for data sets will be provided by the instructor but it is recommended that you search for open data sets too on your own.
2. You can work on this project in a team of at most 4 people. You must mention the names of the members of the group in your proposal and every subsequent document/report. Only one document needs to be submitted per group when required.
3. On **March 31**, submit a project proposal (about two to three pages plus a cover page) describing your data analysis objectives, the data set you have chosen, what you have learned about it so far, and the questions you plan to ask and that you wish to answer with the data.
4. On **April 5**, you will **add to your proposal document** a data analysis plan. Spelling out how you plan to analyze the data, including the cleaning, transformations, missing data mitigation, analysis, and visualization aspects. As before, get feedback from as many people as possible on this document both before and after submission.
5. On **April 21**, submit a progress report that describes a data analysis plan. Spelling out how have been planning or executing the analysis of the data, including the cleaning, transformations, missing data mitigation, analysis of the data.
6. Project presentations will take place on **May 5 and May 12**. The schedule of presentations will be determined by the instructor.
7. On **May 12**, your final project submission is due. This final submission includes a term project written report that provides a description of your data analysis and the meaning and significance of the final results of your analysis. You must also include all of the code you used to transform the data. The report should not be more than 20 pages long.