Data 2020: Midterm Project

The midterm project will be a group exercise designed to flex your statistical learning skills. The main components of the project will be exploring a data set, creating compelling visualizations, analyzing a few scientific questions, and communicating your results. We are interested in seeing you use the data and your analysis to tell a story.

Data

Each group will use the New York City census data provided on Canvas. This data is taken from the 2015 American Community Survey, and information about this data set can be found in the accompanying pdf. Your project should use this data to create good models for median household income and income per capita in NYC and the surrounding area and comment on the differences and similarities between the two models.

Note that the corresponding census tract location file provides the latitude and longitude for each census tract. This will allow you to geographically display information about the data and/or your model.

Deadlines

- **March 6th**: Form a team (2-3 people) and submit your team information through Canvas. If you are having trouble finding teammates, you may email Alice.
- **April 3rd**: Project is due. There will be two deliverables, a website and accompanying analysis file, detailed below. You should submit the latter as a pdf (with the corresponding website address clearly marked) through Canvas.

Deliverables and Grading

- **Website**: Create a public website communicating your insights and findings through visualizations and written analysis. You do not need to include any code chunks. The idea is that someone could stumble across this website and understand the data, your objectives, your approach, and your results, if not all the technical details. You may use Github Pages, Google Sites, or any other website provider. Below are some examples of website layouts and data visualizations taken from Harvard's CS 109 Data Science class.
  - https://scholar.harvard.edu/rajiv/classes/materials/data-science-final-project
  - https://cs109hubway.github.io/classp/
  - https://hamelsmu.github.io/AirbnbScrape/
- **Analysis Document**: This file should include the code you used for the above analysis and provide more insight into your overall approach. For example, you might show how you used regression diagnostics to analyze a potential model.
- **Grading**: The project grade will be based 30% on your communication skills and 70% on the quality and complexity of your overall analysis.