

STA217: Introduction to Data Science

Final Project: Data Analysis and Storytelling

Format: Report + Presentation

Report requirements: Font size 12, at most 10 pages including figures. Group report is welcome.

Final projects will require you to use the skills you learned across the entire course. We want you to develop an interesting research question that can be answered using the data analysis and visualization you have learned in this class.

Note - the items listed below should be considered as guidelines and not just boxes to check. Quality of the work, degree of effort shown, and interpretability and clarity all should be given substantial weight.

Requirements:

- **Collect data:**
 - Option 1: You can choose one of the datasets provided in the zipped file.
 - Option 2: You are encouraged to collect your favorite data.
- **Clean, organize, and process the data** using Python into a nice analyzable format.
- **Conduct analysis.** This could include a series of exploratory data analysis, descriptive statistics and analytical methods learnt from the class.
- **Create data visualizations**, such as scatter plots, bar charts, line graphs, pie charts, etc., to show the correlations between variables you are interested in.
- **Describe your research question and results**, its importance, your analysis, and your findings in a short write-up.

Project team sign up: put the name, student ID of your team members in the online doc: <https://docs.qq.com/sheet/DQmF2ZWfzYlhhdXBr> by Dec 10.

The final projects will have the following deliverables: a PDF report, all the code used to generate results for the project submitted as PDF or a Jupyter Notebook, and an in-class presentation.