

Project

Overview

The goal of the project is to explore a dataset, identify a story revealed by the data, and prepare a report that tells that story compellingly, including effective data visualizations.

The target audience is other people in the class, or members of the general public with a comparable level of statistics knowledge.

Format

The goal of the project is to identify a problem – something that needs to be answered. You'll then source some data, do your analysis, and interpret your results. A good way to phase this (borrowed from Allen Downey – the author of the ThinkStats book):

- Q: motivating question; what is the purpose of the experiment?
- M: methodology; how did you implement the experiment (at an appropriate level of detail)?
- R: results; what happened when you ran the experiment?
- I: interpretation; how do you interpret the result as an answer to the question?

Note: The intent of this project is to be a realistic application of the concepts we covered this term. The exact set of tools and techniques that each of you use will be different – you'll need to select the correct approach for your topic and your data, figure out how to apply it, and interpret the results.

Topics and Data

You may choose almost anything that suits your fancy. There are a few things to keep in mind when selecting a data source:

- There should be an interesting question you're answering. By interesting, I mean something that someone would talk about, write an article about, or use that information for some real purpose.
- You want to demonstrate your mastery of the material – a question that is answered by simply taking an average or generating a histogram doesn't.

Deliverables

You must show your idea and engage in a brief discussion. This is to ensure that you don't choose something that wastes your time.

You will turn in a Jupyter notebook file, containing:

- Your contains the code, comments and explanatory text.

- Commentary, in the text cells, that presents your question, methodology, results, and interpretation.
- Someone should be able to read your notebook page start to end, and it should make sense and communicate everything.

Grading

The grading on this project will be somewhat more demanding than previous, smaller assignments. The expectation is that you generate a cohesive, impactful, and well-presented product.

Category	Weight	Description
Topic and Question	10%	Have you chosen something that is interesting and useful?
Data – Sourcing and Processing	10%	Have you chosen suitable data, loaded and preprocessed it correctly in your program to the point that your analysis can begin?
Analysis	40%	Is your analysis of the data: <ul style="list-style-type: none"> • Suitable. • Sufficient. • Correctly applied. • In pursuit of an answer to the question.
Presentation and Writeup	20%	Is your notebook well-presented and does your written material do a good job of clearly explaining the goal, what you've done, and the result
Overall Quality and Ambition	20%	A subjective evaluation of how impactful your overall project is. Did you do a good job of communicating something interesting to me.