

Learning goals:

- Project Checkpoint
- Regression Demo
- Inference Lab & Q&A

Discussion 5

COGS 108 Winter 2021

Project Checkpoint: Data

DATASET(S): What data will you use to answer your question? **Describe the dataset(s) in terms of number of observations, what kind of features it contains, etc.** (Typically students have datasets of ~1000 observations across their datasets. This is not a requirement, but is good to know around the scale of data we're expecting.) You are welcome (and in fact recommended) to find multiple datasets! If you do so, describe each one, and briefly explain how you will combine them together. Include the source of the dataset in the description here.

SETUP: Include packages used for analysis in cell provided

DATA CLEANING: Get each of your datasets into a usable (likely, tidy) format. Briefly explain what steps you had to take before you were able to use the datasets you chose to answer your question of interest.

- How 'clean' is the data?
- What did you have to do to get the data into a usable format? (If you did nothing, how did you determine there was nothing to do?)
- What pre-processing steps that were required for your methods (for example, checking data distributions and performing any transformations that may be required)

- Documentation: https://github.com/COGS108/Projects/blob/master/FinalProject_Guidelines.md#Checkpoints

Inference

Introduction to Regression (text):

https://www.textbook.ds100.org/ch/14/linear_tips.html

Notebook demo: TipsDemo.ipynb

Your Turn: Work on

D5_Inference