

Final
MATH 110
Due Dec 17.

Find and analyze a dataset. In your analysis, please describe your dataset in enough detail so I can understand what you are doing. (Do not include your dataset in your submission if it is over 1MB.) You should specify the question(s) you are seeking to answer. Your goal is to use R to do some interesting analysis.

Your submission should be composed of R scripts, **containing only functions**; a .Rmd file, in which you use your functions to present your analysis; and an output pdf file created from your .Rmd file.

You must use ggplot at least once in your analysis.

Submit your pdf file directly to the dropbox as usual, but also submit your full analysis, with all .R and .Rmd files, through github as in the previous two homeworks. Fork my github repo SLeviyang/Final. Be sure to place your work in a subdirectory so there are no collisions between different repo merges. Be aware that your disk quota on github is 100MB. Don't push your dataset to your repo if it is large, otherwise you will run into problems.

You may discuss the final and help each other, but everyone should have a different dataset and be doing their own analysis. I am also happy to answer any questions. The website Kaggle is a particularly good resource for datasets.