Math 222, Fall 2020
Statistical Analysis Using R

Final Project

Due Dates:

Data: Wednesday, November 25 by 11:59pm.

Project: Friday, December 4 by 11:59pm.

Data Requirements: For this project, you may use any data set of your choosing, subject to my approval. You must email your data, or a link to the source of your data, by the due date listed above.

Project Requirements: For this project, you will be demonstrating your mastery of the entire data analysis workflow from the past semester. Your submission will consist of 3 parts, each consisting of figures and brief written explanation. Specific requirements for each part are listed below.

Part 1 Exploration

a. SIX figures showing various aspects and relationships within your data set. Among them all, you must use at least two distinct geometries. Each individual figure must make use of at least 3 aesthetics.

b. At last 3 of your figures must be clearly linked attempts at "focusing" your attention on some particular aspect of the data.

c. At least 1 figure must specifically make use of arranging and/or filtering to focus in on some subset of the originally presented data.

d. At least 1 figure must specifically have a modified scale and/or include zooming.

e. At least 1 figure must make use of mutation to indicate information derived from, but not initially present in, the original data set.

f. Each figure should be accompanied by a maximum of 300 words discussing what "interesting" thing you see.

Part 2 Modeling

a. TWO distinct models describing some of the "interesting" behavior you discovered above.

b. This could be two different attempts at explaining the same single behavior. Or, one attempt at explaining each of two different behaviors. Or, one model applied to two different parts of the data to highlight differences between the behavior within those parts. In any case, all models should be "good" or you must clearly explain why you could not do any better.

c. For each model, you must include both a visualization of the model's explicit behavior as well as presentation and discussion of the residuals.

d. Each model should be accompanied by a maximum of 200 words discussing your model's behavior and quality.

Part 3 Communication

a. ONE "Best Figure Ever" to show off your favorite "interesting" observation.

b. This may be accompanied by a maximum of 100 words explaining your result.

The intent of this project is that you could display only your answer to Part 3 and people would understand what you're showing, and *why*.

The diminishing word-count requirements are intended to force you to focus on making the figures themselves have elaborate and informative labels, titles, annotations, and so on.

Note: There is no specific requirement that you properly Wrangle the data into a Tidy format. Obviously, you will still need to do this.

Let me know if you have any questions!