Module 2 – Final Project Due Nov. 8, 2018

Visualization in R, Fall 2020, Instructor Dave Brown

# Dataset

Select a dataset that contains some data diversity, including both categorical and continuous variables. At a very bare minimum, one categorical and two continuous.

# Demonstrate ability to use ggplot (15 pts)

Visualize your data using all of the major plots discussed in class and utilized in homework, including: boxplot, histogram (many different versions), faceting, scatterplot, and a map.

Be sure to demonstrate your understanding of the use of color and shape to represent variables.

Utilize annotation techniques.

If you have two or more categorical variables, then demonstrate the use of contingency tables, mosaic plots and association plots.

# Interpret data visualizations (5 pts)

The point of visualizing data it to identify outliers (or other data problems), examine variable distributions and explore relationships. As a researcher, be sure to use this exercise to gain some insight into your dataset.