

# EDA Homework 9

Due: Friday, April 7, 5pm.

We will analyze a dataset on heart disease, which you can download at <http://jfukuyama.github.io/teaching/stat670/assignments/heart.csv>.

The dataset has information about some predictors of heart disease. The columns that we will look at are:

- HeartDisease: 0/1 for heart disease or not. The variable we would like to explain.
- ST\_Slope: Up, down, or flat. A feature related to an electrocardiogram measurement of the heart.
- ExerciseAngina: Y/N for yes/no. A type of chest pain that occurs during exercise.
- RestingECG: LVH/Normal/ST. Describes whether the resting electrocardiogram measurements were normal (Normal), indicated left ventricular hypertrophy (LVH), or indicated an STT abnormality (ST).

Assignment:

1. Make mosaic plots of HeartDisease in combination with each of the other three variables (so three plots total). Which of the variables seem useful for predicting HeartDisease? Which do not?
2. Make mosaic plots to investigate the relationships among the three potential predictor variables (ST\_Slope, ExerciseAngina, and RestingECG). Are there any pairs that seem like they might be redundant?
3. Based on the relationship between ST\_Slope and ExerciseAngina, we might wonder if the two together have more information about HeartDisease than either one individually. Investigate this using a cotabplot.