I ran basic data for all variables; mean, mode, variance, standard deviation. Since we picked 5 variables, and had 5 different testing assignments to run, I chose to run one test per variable (with a couple of exception). First I ran a histogram test on cholesterol, one for those with heart disease and one for those without. The graphs showed that those without heart disease did have lower highest number. It appeared that the graphs might be in favor of those without heart disease. So, I ran a mean test of those with and without heart disease and the average was actually higher for those without heart disease.

Next I ran the cdf plot on resting blood pressure. The resting blood pressure actually appeared lower in those with heart disease. Next I attempted to do an analytical distribution with fasting blood sugar. This may not have turned out like it could due to the fact that both heart disease and fasting blood sugar only have 0 and 1 as variables. I ran scatterplots on the patient’s ages, along with a correlation test. Graph showed pretty uniform and correlation stated it wasn’t corrected. Ran the DiffMeansPermute from the book on maximum heart rate achieved. According to the p-values ran, it was not statistically significant. Lastly, I attempted to do a fitline scatterplot with several of the factors (resting blood pressure vs age, resting blood pressure vs heart disease, and fasting blood sugar vs heart disease).

Much more tests would need to be run on this dataset to see which variables are predictive of heart disease and which aren’t. Also, a larger dataset might be needed to get a better conclusion since only 300 people participated in this study. None of these seem to be predictors on the tests run. Also, it’s difficult to tell if the younger people in the study who don’t have heart disease, might one day get it. Meaning, they could be skewing the data if they have some of the predictors of heart disease when they don’t have it (because they will get heart disease someday). Also, not sure if the age given for those with heart disease is the age in which they were diagnosed or the age that they took the survey because that is also an important factor. The only assumptions I can make from this is that either the dataset wasn’t big enough, the data taken was biased, or the variables tested were just not predictors of heart disease. It’s possible that heart disease is widely due to your genes, which was not a variable in this study. I also admit that I did not fully understand the fitline scatterplot or the cdf.