**Statistics 355 — Final Project**

Team: Camylle Arsenault, Viola Wu, Steven Huang, and Yizhe Zhang

A study was done in order to determine the factors leading to a higher probability of having a heart attack, and even predicting them. The study was done specifically to predict possible diameter narrowing, which could in turn predict heart attacks. The data was obtained from a set of hospitals around the world in 1988. The data set, containing 135 adults ranging from the age of 29 to 77, explores many factors that could either increase or decrease the probability of an individual having a heart attack. Knowing the possible predictors and their respective probability of causing a heart attack can be very important, since some of those variables are able to be prevented or treated, ultimately reducing the risk of a heart attack.

**Variables:**

Age (in years)

sex (1=male, 0=female)

cp—the type of chest pain (0=typical angina, 1=atypical angina, 2=non-anginal pain, 3=asymptomatic)

trestbp—the resting blood pressure (in mm Hg)

chol—the level of serum cholesterol (mg/dl)

fbs—the fasting blood sugar > 120 mg/dl (1=true, 0=false)

restecg—the resting electrocardiographic results (0=normal, 1=not normal)

thalach—the maximum heart rate achieved

exng—the presence of exercise-induced angina (1=yes, 0=no)

oldpeak—ST depression from exercise to rest

slp—the slope of the peak exercise ST segment (0=up-sloping, 1=flat, 2=down-sloping)

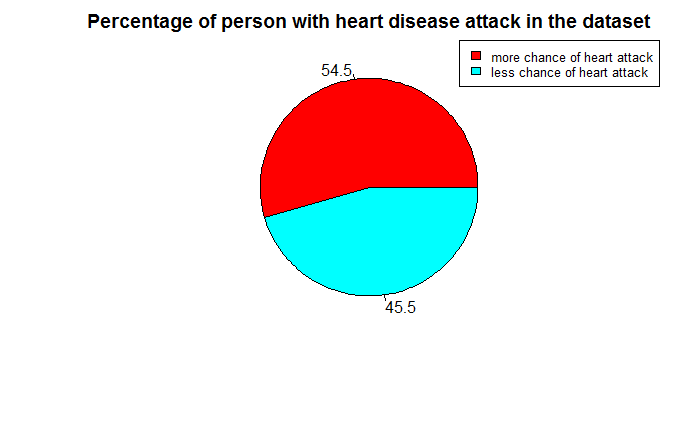
caa—the number of major vessels (0-4)

thall— thalassemia (1=normal, 2=fixed defect, 3=reversible defect)

num—the diagnosis of heart disease (angiographic disease status) (0=<50% diameter narrowing, 1=<50% diameter narrowing)

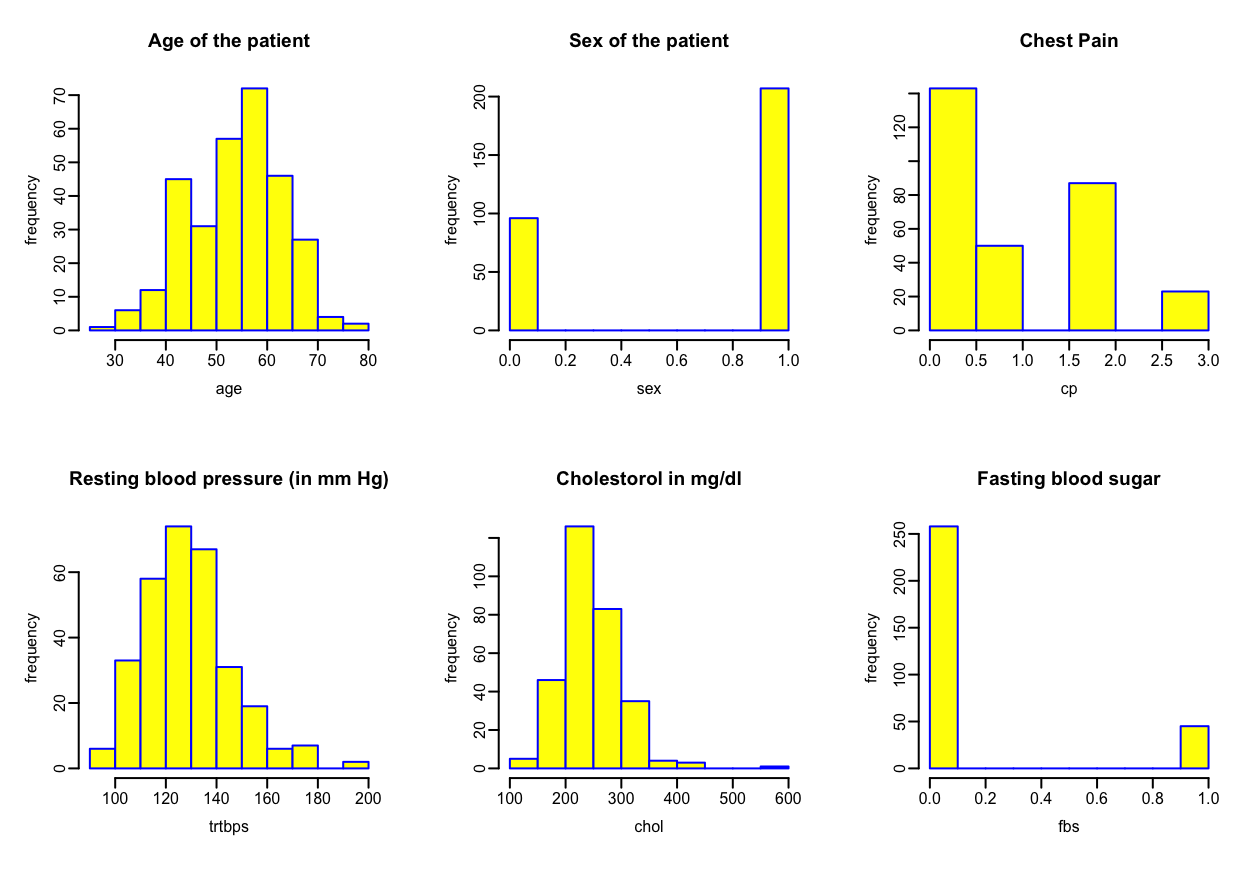
**Treatment of data**

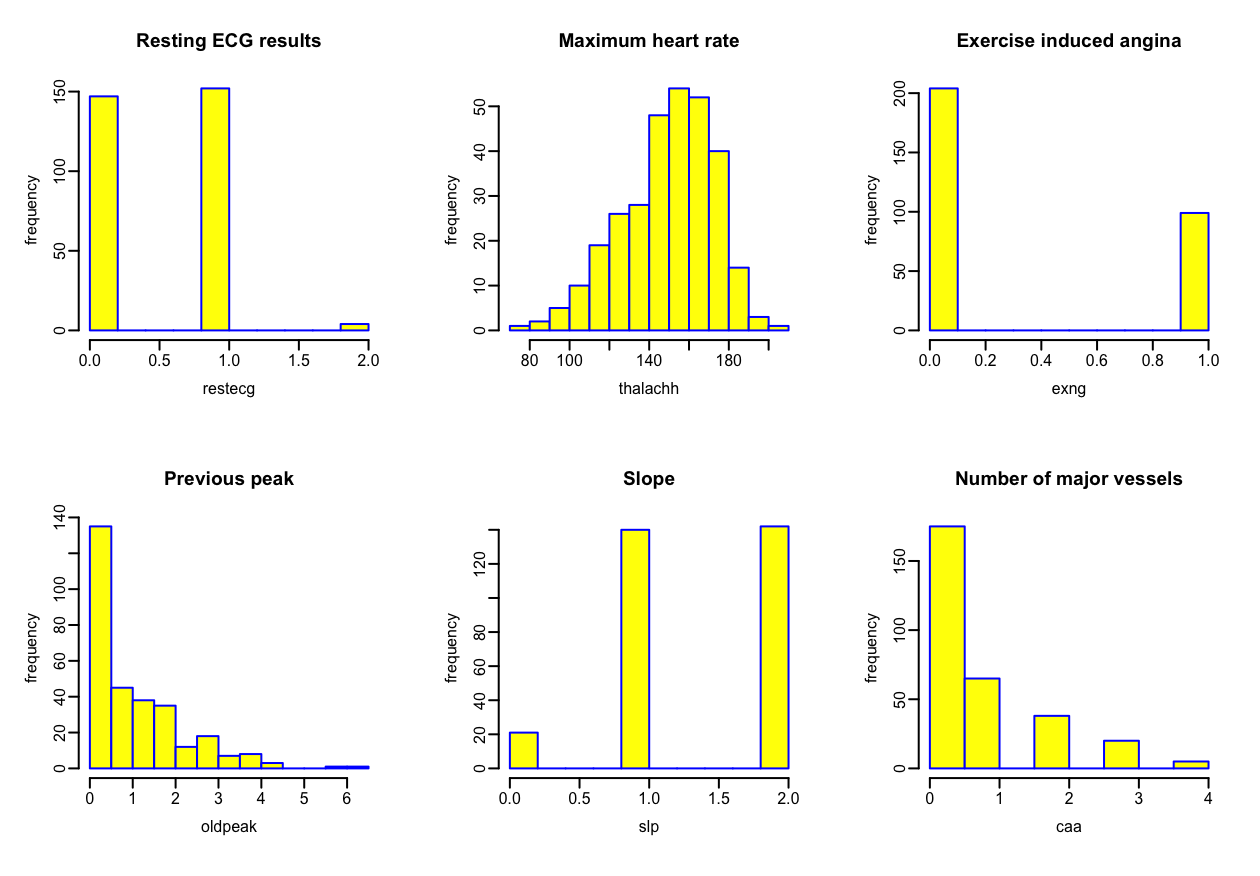
**Pie chart of data (Appendix 1.1)**



As the pie chart show above, 54% of the people in this study had heart attack, almost creating a 1:1 ratio between the people who had a heart attack and those who did not. (Appendix 1.1)

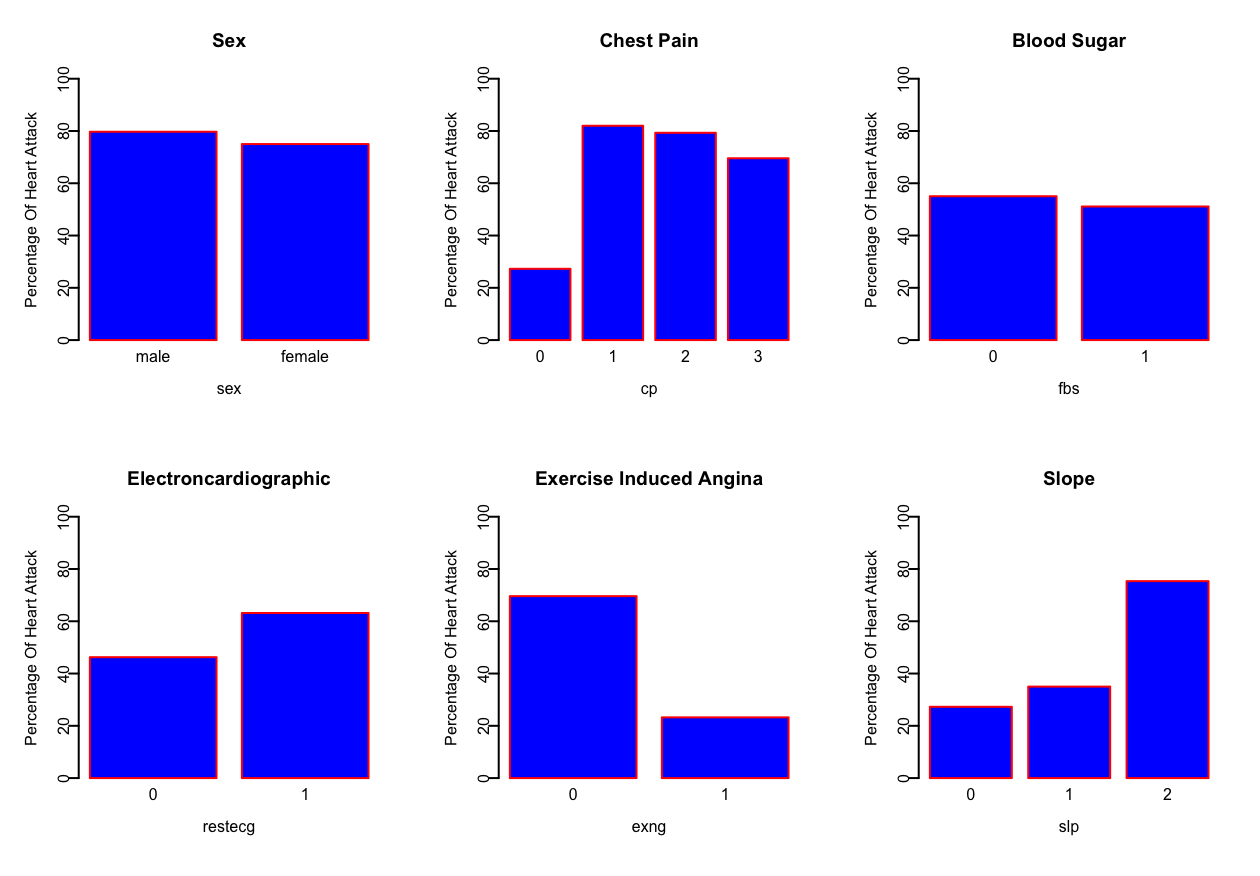
**Series of histograms summarizing the variables and their frequency in the data set: (Appendix 1.2)**

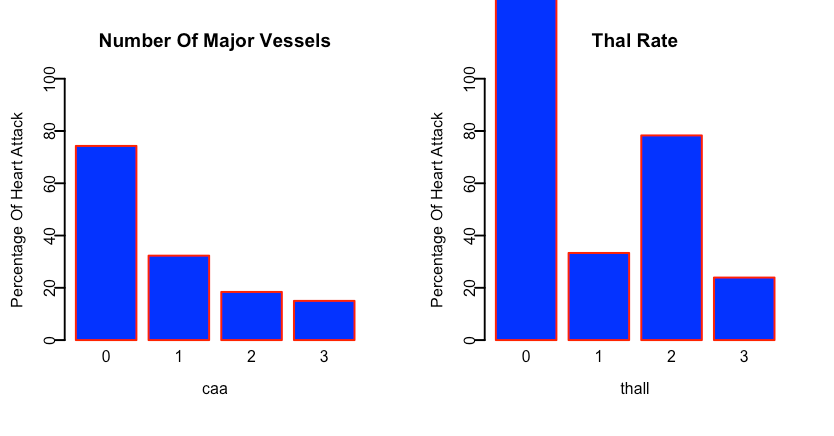




As the plots shown above, the highest frequencies are reserved for men (1) between the ages of 50 and 60 having less chest pain, a blood pressure between 120 and 140, cholesterol between 200 and 300, a blood sugar less than 120, and a heart rate between 150 and 175.

**Series of boxplots summarizing the probability suffering from a heart attack based on the variables: (Appendix 1.3)**





In our data, there is a higher chance of a male suffering from a heart attack than a woman, but the percentages are too close to make a statement that men are more likely to have heart attacks.

People having chest pains are more probable to suffer from a heart attack.

Blood sugar doesn’t have that much impact on heart attack, i.e., the person having heart attack is independent of whether person has high blood sugar or not.

People suffering from exercise-induced angina are less likely to suffer from a heart attack than people who do not.

**Logistic Regression Model: (Appendix 1.4)**

Upon assessing the predictor variables and the response fbs in a univariate model, only the p-value of slp (the slope) is less than 0.1. This means that slp is the only predictor variable associated to the fbs (fasting blood sugar), the other ones are not. Similarly, in a multivariate fashion, the age is significant to fbs (p<0.05).

**Conclusion:**

By analyzing the different predictor variables that raise the probability of having a heart attack, we were able to determine which factors play a large role in determining if a person is prone to a heart attack or not. We determined that there is not a large gap between the probability of a man and the probability of a woman having a heart attack, that gender does not play a drastic role in the probability of having a heart attack. We also determined that a person having typical anginal pains has a very low chance of having a heart attack, that it’s when it becomes atypical that it becomes more dangerous, i.e., a higher chance of having a heart attack. We also determined that a person’s blood sugar does not matter all that much when determining the chances of having a heart attack; however, we also learned that a fasting blood sugar is related to only one predictor variable, the slope of the peak exercise ST segment. The resting electrocardiographic results also proved that a person with abnormal results is more prone to heart attacks than a person with normal results. The presence of exercise-induced angina showed a decrease in the probability of having a heart attack, indicating that a person that does not have exercise-induced angina is more likely to have a heart attack. A down-sloping slope of the peak exercise ST segment shows a greater probability of having a heart attack. The presence of zero major vessels also shows a much greater chance of having a heart attack. By analyzing those different predictor variables, we were able to determine who is the most likely to have a heart attack, and perhaps be able to predict the occurrence of a heart attack in a specific individual, and prevent it.