Title: Leading Factors of Heart Disease

Group Members: Em Greene, Gabriel Rodriguez, Noor Zaki

In this project we look at several factors related to heart disease and evaluate their relationship to age and the significance of their relationship to heart disease. We pulled the dataset from Kaggle. It is called “Predicting Heart Disease using Clinical Variables” there are 270 patients evaluated on 13 features of heart disease. Credit for this dataset’s availability goes to Robert Hoyt, MD.

In our analysis we took out a few items as their relationship to heart disease was not explained in the dataset’s description, and we could not confidently make an analysis of these items without the additional information. These were FBS over 120, Slope of ST, and Thallium, and we also removed the index row as that was not necessary. The data had no duplicates and no N/A values, so it was a very clean dataset.

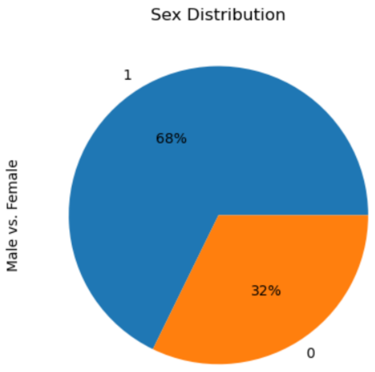
After cleaning the data, we looked at a few of the items using summary tables of the mean, median, variance, standard deviation, and standard error of measurement for Age, Blood Pressure (BP), Cholesterol, Max HR, and ST Depression. For each of these summary tables, the group with Heart Disease present had a higher mean, with the exception of Max HR, which had a lower average value in the group with heart disease present.

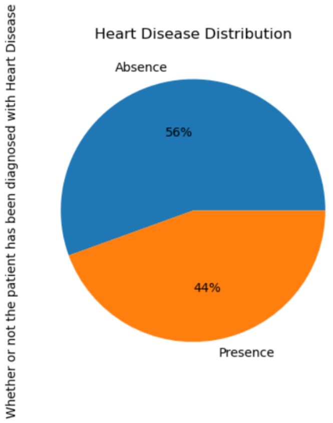
For the next block of code, we created some visualizations of the distribution of Sex, Exercise Angina, Chest Pain Type, EKG results, and Heart Disease presence and absence using pie charts so we could visualize the distribution of each of these factors. In these visualizations we found that the majority of the population of the study was male. Heart Disease was present in a little less than half of the 270 patients. EKG results were positive for atypical heart rhythm for about half of the population. About half of the population of the study experienced some sort of chest pain with exercise, with 33% experiencing severe pain, or exercise induced angina.

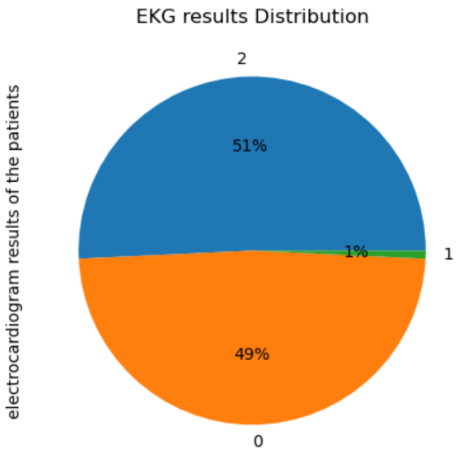
Moreover, the sex distribution chart simply divides study into two categories. (1) male and (2) female. With that, you can clearly understand the pie chart. Secondly, the chest pain distribution chart has 4 values. (1) typical angina, (2) atypical angina, (3) non-anginal pain and (4) asymptomatic. Value 4 has the highest percentage vs. Value 1,2 and 3. Thirdly, the Exercise angina chart has two values. Value 1 is yes and Value 2 is no. 33% of individuals had exercise induced angina and 67% didn't have exercise induced angina. Chart 4, “Heart Disease Distribution”, 56% of patients were NOT diagnosed with heart disease but 44% were diagnosed with it. Lastly, EKG results distribution chart had 3 values. Value (0): normal, Value (1) having

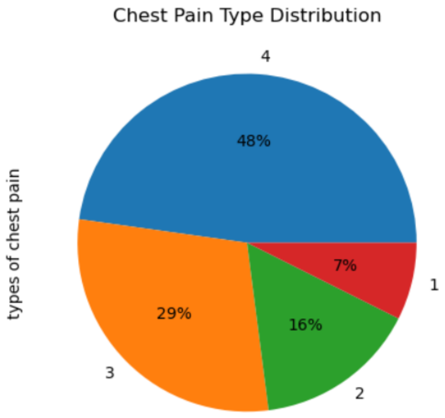
ST-T wave abnormality and Value (2) showing probable or definite left ventricular hypertrophy by Estes criteria. Value (1) had 1% of patients with ST-T wave abnormality, Value 2 had 51% of patients showing probably left ventricular hypertrophy and Value 0 had 49% of patients results as normal.

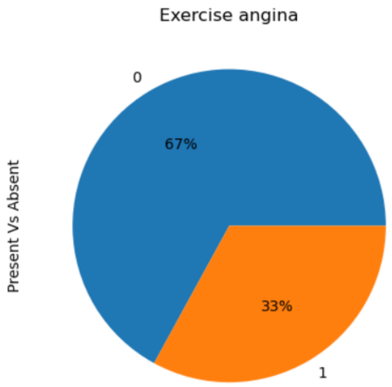
See the below pie charts for further information:





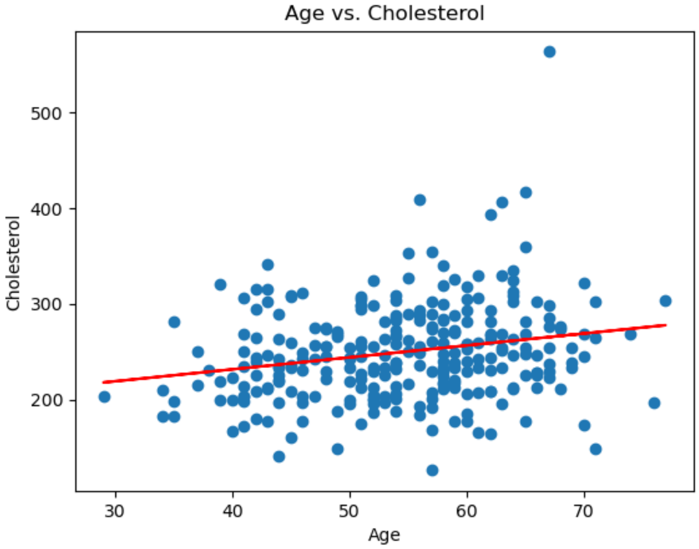


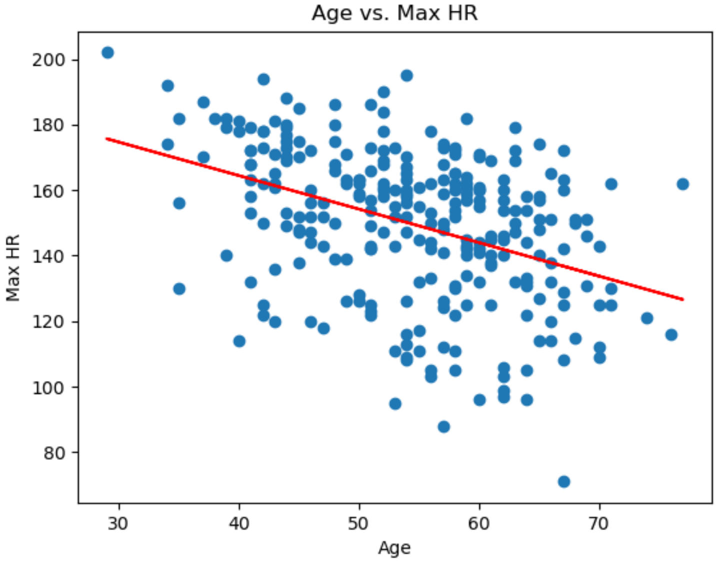


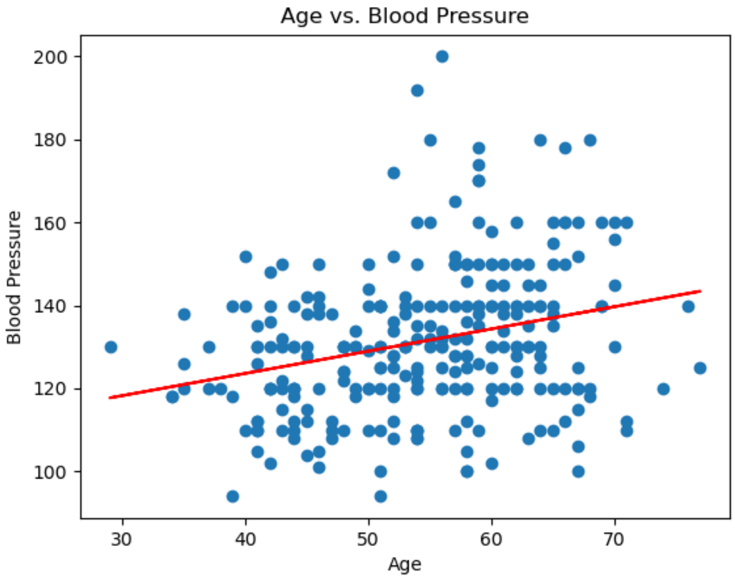


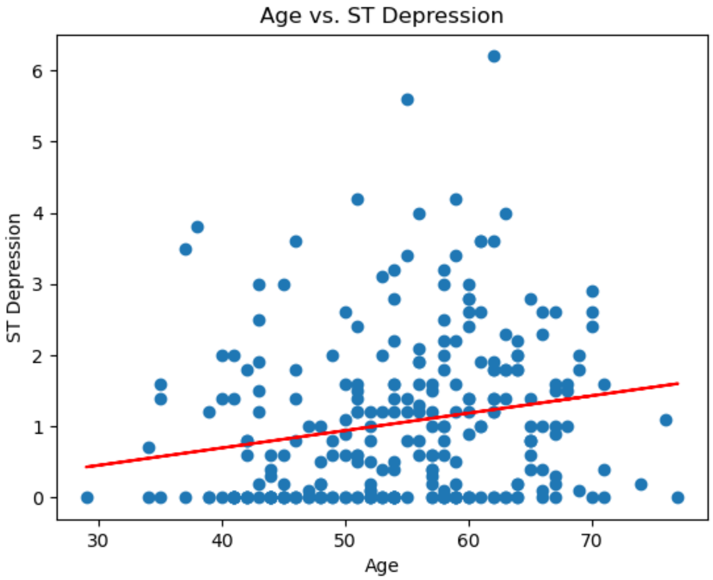
We continued to find correlations between age and various cardiovascular factors, namely cholesterol levels, maximum heart rate (Max HR), blood pressure, and ST depression. We generated scatter plots and regression lines to visually represent these correlations to highlight distinct age groups and patterns, there is a steady range between 200 to 300 in cholesterol levels. Cholesterol findings indicate a rise in levels due to aging. We concurred that outliers in correlated data could result in genetics, lifestyle, and pre-existing health conditions. Moving to the next cardiovascular data, maximum heart rate, there was evidence displayed by the trend line that Max HR decreases over time.The average Max HR zone was 180 and declining over time from that data point. We noticed the low spectrum and high spectrum of Max HR data, we read though as possible factors that result in physical activity over time, and other health conditions that take part in the output performance of heart rate.

Blood pressure showed a general increase as age increases. Most of the findings with blood pressure fell between 120 and 140 in the age ranges of 40 to 70. We understood how analysis in other factors can have an effect on blood pressure so we take consideration of that when interpreting the outliers of higher or lower pressure that fall away from the trend line. Next, ST depression was a good challenge of analysis compared to previous conditions.The ST depression scatter plot was quite mixed in that there was consistency in no change over the years of aging and there were a multitude of spikes in ST depression levels over aging. A conclusion led to levels being slightly increased as individuals aged whilst taking strong consideration of lifestyle factors and pre-existing health conditions that can come into play. Observe scatter plot visuals beneath:


Age vs. Cholesterol
Cholesterol
500
400
300
200
30
40
50
60
70
Age


Age vs. Max HR
Max HR
200
180
160
140
120
100
80
30
40
50
60
70
Age





Finally, we took a look at each of the factors in each group (with and without heart disease) and ran an independent t-test on the factors to see if there is a significant difference between the groups' scores. In this analysis, we found significance at a p-value of less than 0.05 for all of the factors. The top 3 factors that are most significantly different between the groups were: "Number of Vessels fluro" which are the numbers of vessels that show in a

Fluoroscopy, and this is higher in the group with Heart Disease (t-statistic: 7.98 p-value: 1.3x10^ -13) , Chest pain type experienced during exercise is significantly higher for the Heart Disease group (t-statistic: 7.67 p-value: 3.3x10^ -13), and Max HR is significantly lower in the group with heart disease (t-statistic: -7.39 p-value: 2.6x10^ -12). These are the top 3 factors that we can say are definitely related to heart disease, and may even be predictors of heart disease.