Final Project Info 250

Sanobar Lala

8/26/2020

##Welcome Physicians!

Through analyzing this dataset, I came across interesting relationships between different variables in this dataset that you can explore. The first visual you will see is a correlation heatmap with shows correlations for all the variables in this dataset. The second visual shows the relationship between Rest ECG and the likelihood of developing heart disease. The final visual shows you a distribution of the relationship between trestbps and heart disease for males and females (individually). I hope you can take away valuable insights from these visualizations and use it for preventative medicine!

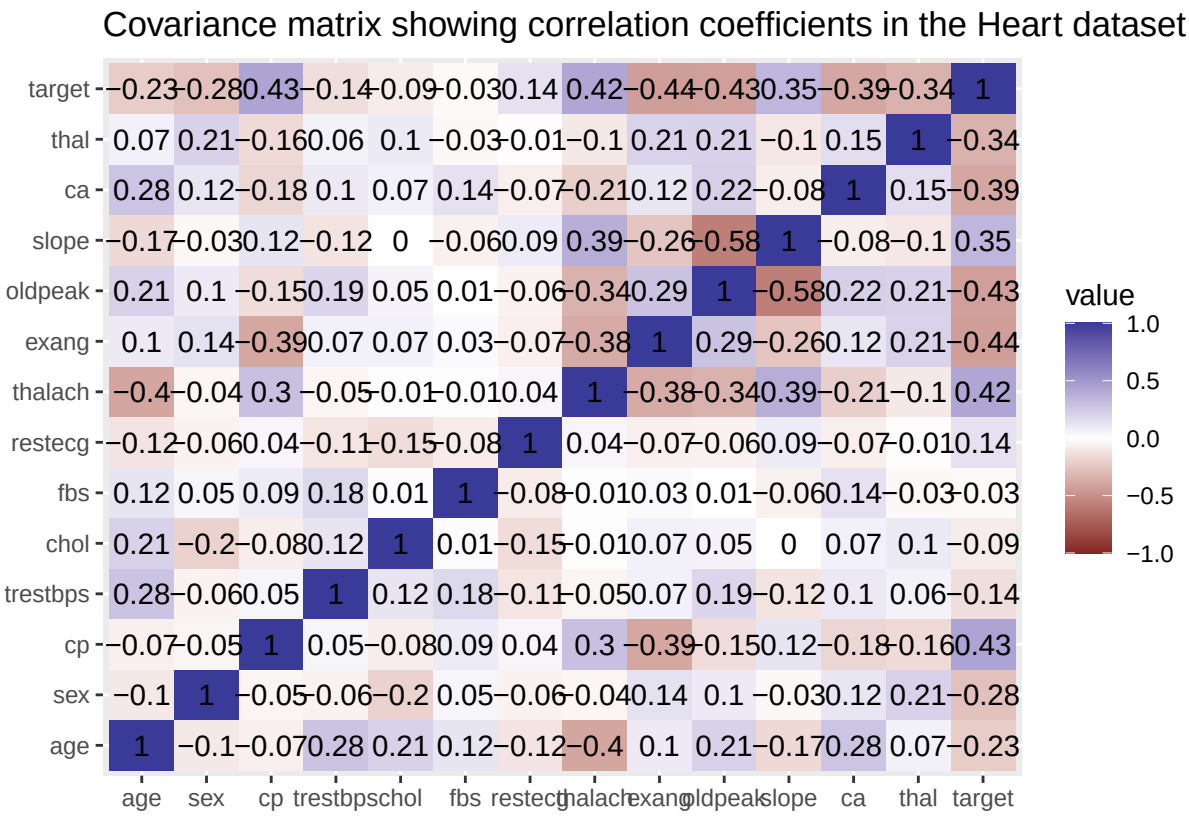
First, to provide you with the summary statistics and column names of the Data

```
##       age             sex               cp           trestbps    
##  Min.   :29.00   Min.   :0.0000   Min.   :0.000   Min.   : 94.0  
##  1st Qu.:47.50   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:120.0  
##  Median :55.00   Median :1.0000   Median :1.000   Median :130.0  
##  Mean   :54.37   Mean   :0.6832   Mean   :0.967   Mean   :131.6  
##  3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:140.0  
##  Max.   :77.00   Max.   :1.0000   Max.   :3.000   Max.   :200.0  
##       chol            fbs            restecg          thalach     
##  Min.   :126.0   Min.   :0.0000   Min.   :0.0000   Min.   : 71.0  
##  1st Qu.:211.0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:133.5  
##  Median :240.0   Median :0.0000   Median :1.0000   Median :153.0  
##  Mean   :246.3   Mean   :0.1485   Mean   :0.5281   Mean   :149.6  
##  3rd Qu.:274.5   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:166.0  
##  Max.   :564.0   Max.   :1.0000   Max.   :2.0000   Max.   :202.0  
##      exang           oldpeak         slope             ca        
##  Min.   :0.0000   Min.   :0.00   Min.   :0.000   Min.   :0.0000  
##  1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:1.000   1st Qu.:0.0000  
##  Median :0.0000   Median :0.80   Median :1.000   Median :0.0000  
##  Mean   :0.3267   Mean   :1.04   Mean   :1.399   Mean   :0.7294  
##  3rd Qu.:1.0000   3rd Qu.:1.60   3rd Qu.:2.000   3rd Qu.:1.0000  
##  Max.   :1.0000   Max.   :6.20   Max.   :2.000   Max.   :4.0000  
##       thal           target      
##  Min.   :0.000   Min.   :0.0000  
##  1st Qu.:2.000   1st Qu.:0.0000  
##  Median :2.000   Median :1.0000  
##  Mean   :2.314   Mean   :0.5446  
##  3rd Qu.:3.000   3rd Qu.:1.0000  
##  Max.   :3.000   Max.   :1.0000
```

```
colnames(data)
```

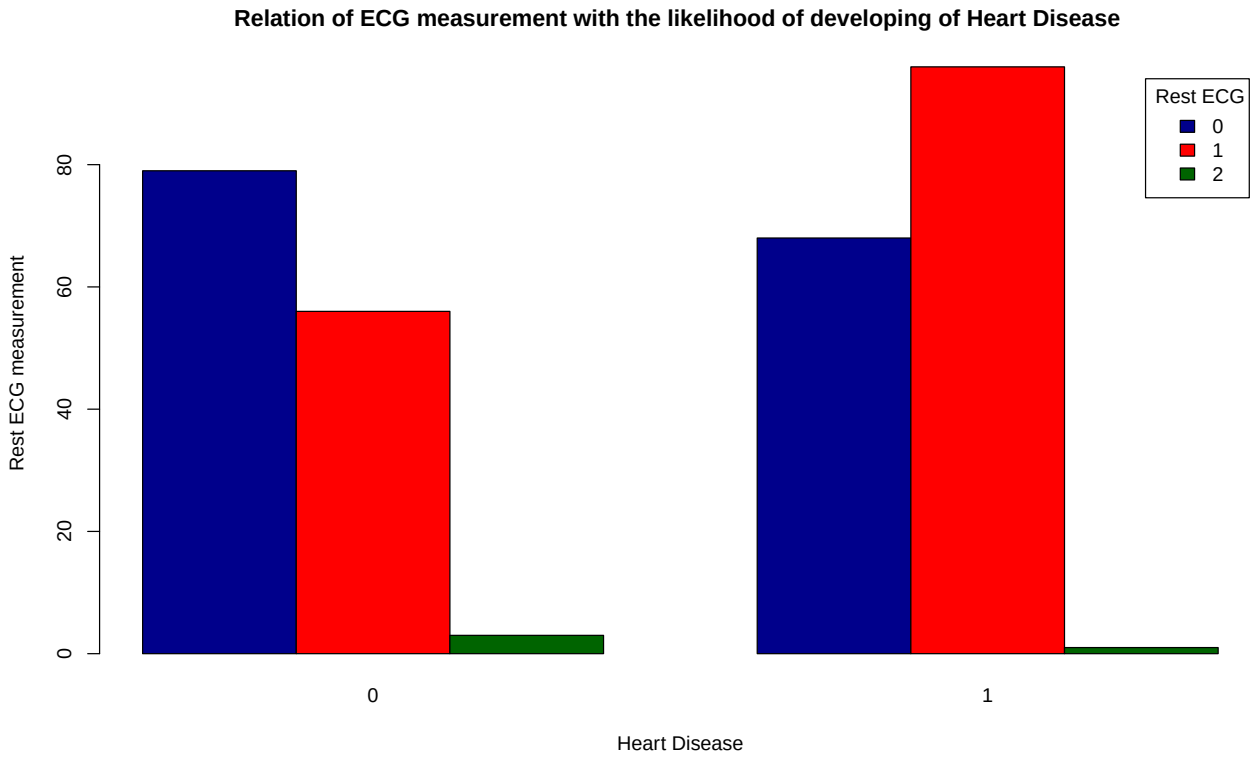
```
##  [1] "age"      "sex"      "cp"       "trestbps" "chol"     "fbs"
##  [7] "restecg"  "thalach"  "exang"    "oldpeak"  "slope"    "ca"
## [13] "thal"     "target"
```

Correlation HeatMap



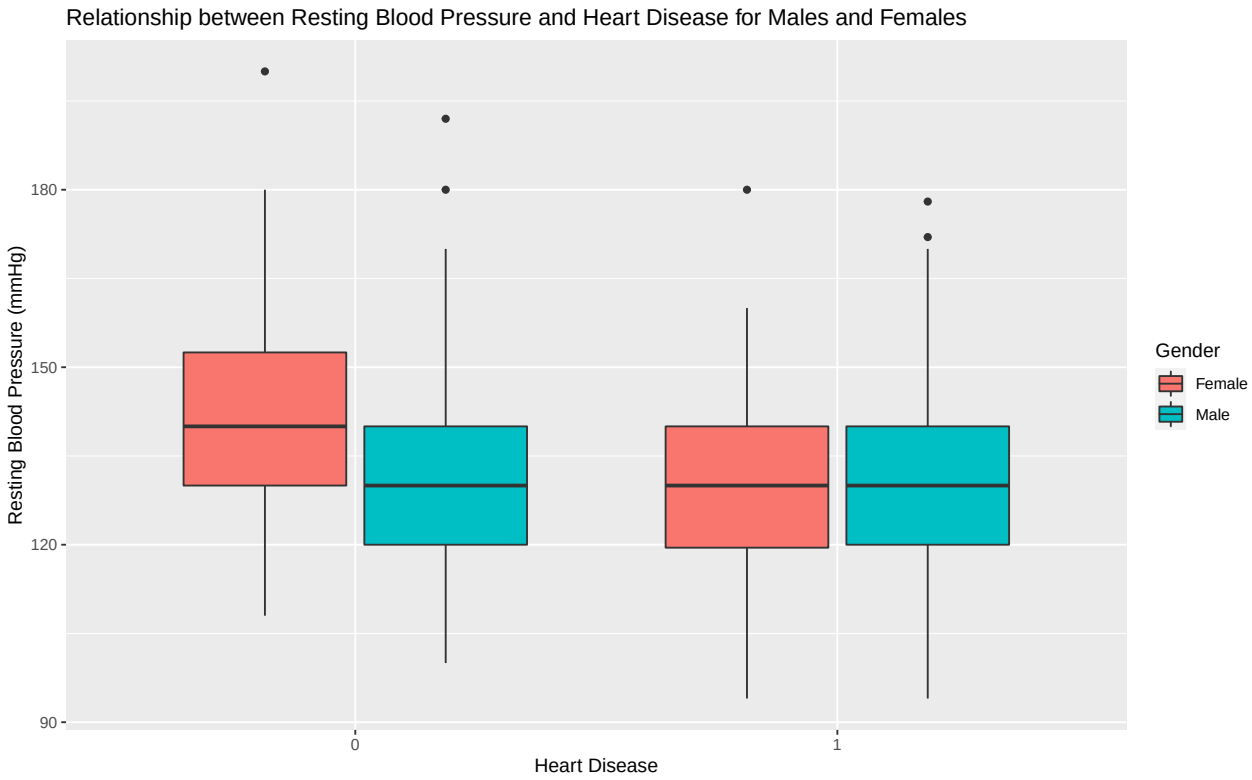
Analysis of the visual: As we can clearly see from the legend, the bluer the value, the higher the correlation and the closer to brown, the negative or lower the correlation. Looking at the correlation with the target variable, cp (chest pain), thalach (max heart rate) and slope (peak exercise) have the highest correlations and rest ECG and cholesterol have the lowest correlations.

Barchart



Analysis: As we can recall, the value meaning of restecg are: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria). The above plot shows that a high number of patients not likely to suffer from heart diseases have restecg value 0 (normal) whereas a greater number of people have restecg value 1 in case of more likelihood of suffering from a heart disease

##Boxplot



Analysis: In the above Box plot between Target and tresbps and Gender, shows that Women have higher tresbps than men when we are looking at target variable 0, which is the likelihood of not developing heart disease, whereas men and women have almost equal tresbps in case of suffering from a heart diseases. Also, when target variable is 1 (getting heart disease), patients have a slightly lower tresbps in comparison to the patients who are not suffering from heart diseases (when target variable is 0).