IS6400 Final Project

Madyson Rowland

4/30/2021

## INTRODUCTION

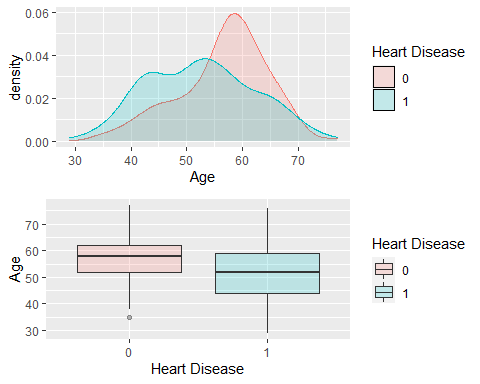
The world Health Organizations(WHO) announced that cardiovascular diseases is the top one killer over the world. Nearly 17.9 million people die from cardiovascular disease every year. Preventing cardiovascular disease could be better than a cure. If there was a way to evaluate the risk of every patient who could have a hear disease. This data set is real data that include important features of patients.

### Data Dictionary

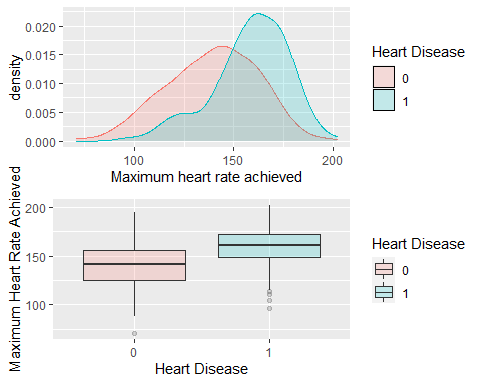
* **age:** Age of the patient in years
* **sex:** (1 = male; 0 = female)
* **cp:** chest pain type:
  + Value 1: typical angina
  + Value 2: atypical angina
  + Value 3: non-anginal pain
  + Value 4: asymptomatic
* **trestbps:** Resting blood pressure (in mm Hg on admission to the hospital)
* **chol:** serum cholestoral in mg/dl
* **fbs:** fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
* **restecg:** resting electrocardiographic results
  + Value 0: normal
  + Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  + Value 2: showing probable or definite left ventricular hypertrophy by Estes
* **thalach:** maximum heart rate achieved
* **exang:** exercise induced angina (1 = yes; 0 = no)
* **oldpeak:** ST depression induced by exercise relative to rest
* **slope:** the slope of the peak exercise ST segment:
  + Value 1: upsloping
  + Value 2: flat
  + Value 3: downsloping
* **ca:** number of major vessels (0-3) colored by flourosopy
* **thal:**
  + 3 = normal
  + 6 = fixed defect
  + 7 = reversable defect
* **target:** diagnosis of heart disease (angiographic disease status)
  + Value 0: < 50% diameter narrowing
  + Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

## Visualization

In the following visualizations we see the density of age to heart disease. With 0 meaning no disease and 1 having heart disease.  
**Variable Age**

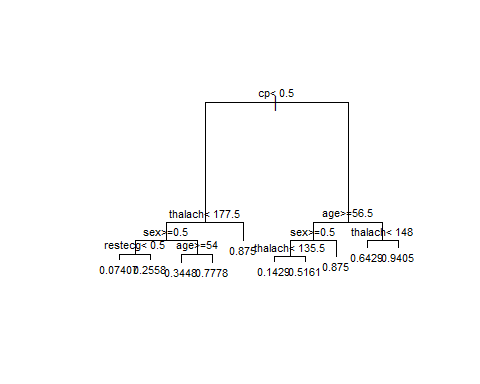


**Maximum heart rate achieved**



## Analysis

Applying more variables with higher correlations to check how the chances of heart disease vary:



Through the decision Tree we can observe that the highest probable chance for a person to get a heart disease is for females with the following conditions:  
Above the age of 57 with a chance of 87%  
Having a thalach(maximum heart rate) reading greater than 153 with a chance of 94%  
Any level of cp is indicative of a heart disease for females.

For this project, we created a training and test set using the 70% vs 30% of observations

##   
## Call:  
## glm(formula = goodmodel, family = binomial(link = logit), data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4194 -0.4008 0.1756 0.5663 2.6888   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.075249 2.165165 1.420 0.155512   
## slope 0.637519 0.418099 1.525 0.127308   
## thalach 0.015961 0.011879 1.344 0.179077   
## exang -0.708709 0.481025 -1.473 0.140662   
## chol -0.008026 0.004119 -1.949 0.051326 .   
## oldpeak -0.510552 0.249239 -2.048 0.040516 \*   
## thal -0.957078 0.344900 -2.775 0.005521 \*\*   
## ca -1.199536 0.261416 -4.589 4.46e-06 \*\*\*  
## cp 0.841599 0.229148 3.673 0.000240 \*\*\*  
## sex -1.893213 0.540837 -3.501 0.000464 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 292.36 on 211 degrees of freedom  
## Residual deviance: 149.60 on 202 degrees of freedom  
## AIC: 169.6  
##   
## Number of Fisher Scoring iterations: 6

## Warning in cbind(df$target, yhat): number of rows of result is not a multiple of  
## vector length (arg 2)

## yhat  
## [1,] 1 0.337757461  
## [2,] 1 0.003736319  
## [3,] 1 0.072959952  
## [4,] 1 0.026118097  
## [5,] 1 0.746078238  
## [6,] 1 0.037185182

we see that Sex, cpm Ca, thal, exang, and thalach all have significance.

## result1  
## test\_disease 0 1  
## 0 33 8  
## 1 8 42

### Conclusion

As we mentioned earlier the following criteria are most likely to have heart disease:  
Above the age of 57 with a chance of 87%  
Having a thalachmaximum heart rate) reading greater than 153 with a chance of 94%  
Any level of cp is indicative of a heart disease for females.

## age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal  
## 1 58 0 3 150 283 1 0 162 0 1.0 2 0 2  
## 2 58 0 2 120 340 0 1 172 0 0.0 2 0 2  
## 3 65 0 2 140 417 1 0 157 0 0.8 2 1 2  
## 4 63 0 2 135 252 0 0 172 0 0.0 2 0 2  
## 5 60 0 2 102 318 0 1 160 0 0.0 2 1 2  
## 6 67 0 2 115 564 0 0 160 0 1.6 1 0 3  
## 7 67 0 2 152 277 0 1 172 0 0.0 2 1 2  
## 8 60 0 3 150 240 0 1 171 0 0.9 2 0 2  
## target  
## 1 1  
## 2 1  
## 3 1  
## 4 1  
## 5 1  
## 6 1  
## 7 1  
## 8 1

Cardiovascular diseases is the top one killer for many years. We think that the reason for it being the top killer would our lack of knowledge about heart disease an life habits. According to the models and our analysis, we know which features that we can to self-examinations regularly.

We think the most obvious symptom is chest paint. But there are three types of chest pain, but only atypical anguna is strongly related to heart disease.

In, addition, everyone should always keep an eye your resting blood pressure. The ideal resting blood preassure is lower than 120mmHg, but if your blood pressure is much lower than 120mmhg, it means that you are under hgh ris of heart disease. Besides, the problem will not be only heart when the blood pressure is higher than 150mmHg.

Investing in an electronic device like any newer model smart watches can measure your heart rate, so you can measure it on your own.

No matter how healthy we are, we must do annual examination because another features can not be taken care of by ourselves. My future mother in law had a quadulapbe bypass summer of 2019 and when she went back in for her check up this year 4 of the 5 bypasses failed. She changed her habits and stayed active, doing those things earlier in life would prevent heart disease. She also had the gene of heart disease from her father so. If she would have started a healthier life style before the age of 50 she might have prevented her heart disease.