Heart attack

This are the libraries required for this analysis

library(ggbiplot)

## Loading required package: ggplot2

## Loading required package: plyr

## Loading required package: scales

## Loading required package: grid

library(tidyverse)

library(RColorBrewer)

As I stated in my other posts, I always prefer file.choose(), sometimes I use read.csv directly if the working directory is set.

Heart<-read.csv(file.choose(),header=T)

CHecking what the data set looks like…. Sneak Peek

head(Heart)

## sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal  
## 1 male 3 145 233 1 0 150 0 2.3 0 0 1  
## 2 male 2 130 250 0 1 187 0 3.5 0 0 2  
## 3 female 1 130 204 0 0 172 0 1.4 2 0 2  
## 4 male 1 120 236 0 1 178 0 0.8 2 0 2  
## 5 female 0 120 354 0 1 163 1 0.6 2 0 2  
## 6 male 0 140 192 0 1 148 0 0.4 1 0 1  
## target age  
## 1 1 63  
## 2 1 37  
## 3 1 41  
## 4 1 56  
## 5 1 57  
## 6 1 57

str(Heart)

## 'data.frame': 303 obs. of 14 variables:  
## $ sex : chr "male" "male" "female" "male" ...  
## $ cp : int 3 2 1 1 0 0 1 1 2 2 ...  
## $ trestbps: int 145 130 130 120 120 140 140 120 172 150 ...  
## $ chol : int 233 250 204 236 354 192 294 263 199 168 ...  
## $ fbs : int 1 0 0 0 0 0 0 0 1 0 ...  
## $ restecg : int 0 1 0 1 1 1 0 1 1 1 ...  
## $ thalach : int 150 187 172 178 163 148 153 173 162 174 ...  
## $ exang : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ oldpeak : num 2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...  
## $ slope : int 0 0 2 2 2 1 1 2 2 2 ...  
## $ ca : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ thal : int 1 2 2 2 2 1 2 3 3 2 ...  
## $ target : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ age : int 63 37 41 56 57 57 56 44 52 57 ...

Heart$sex <-as.factor(Heart$sex)

Time to do some quick visuals and peep the data, we want to see how the data is spread by variable.

Heart\_long<-  
 gather(data=Heart, key=Variable, value=Value, -sex)

Time to visualize our data

library(ggridges)  
  
ggplot(Heart\_long)+  
 geom\_density\_ridges(aes(y=sex, x=Value, fill=sex))+  
 facet\_wrap(~Variable, scales="free")+theme\_bw()

A picture containing chart

Description automatically generated

library(randomForest)

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':  
##   
## combine

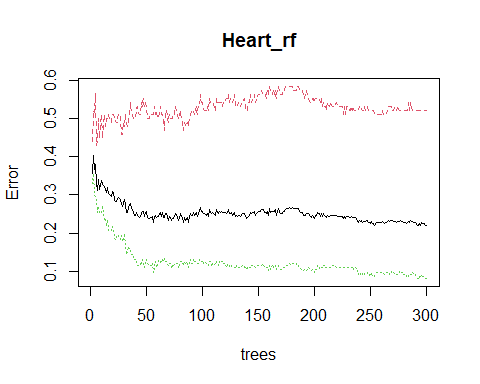
## The following object is masked from 'package:ggplot2':  
##   
## margin

Heart\_rf<-randomForest(factor(sex) ~ ., data=Heart, mtry=3, ntree=300, importance=TRUE, proximity=TRUE)  
  
Heart\_rf

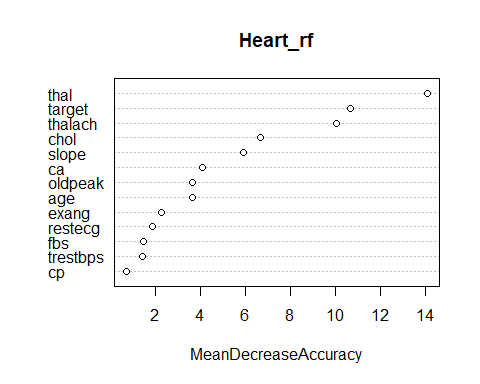
##   
## Call:  
## randomForest(formula = factor(sex) ~ ., data = Heart, mtry = 3, ntree = 300, importance = TRUE, proximity = TRUE)   
## Type of random forest: classification  
## Number of trees: 300  
## No. of variables tried at each split: 3  
##   
## OOB estimate of error rate: 22.11%  
## Confusion matrix:  
## female male class.error  
## female 46 50 0.5208333  
## male 17 190 0.0821256

Lets find out what’s pushing that OOB error rate up

plot(Heart\_rf)

 Time to plot variable by importance

varImpPlot(Heart\_rf, type=1)



According to the dataset metadata the thal, and target represents the AHD; ALveolar Hydatid disease. Thus our model suggests that this disease is largely responsible for Heart attacks.

importance(Heart\_rf)

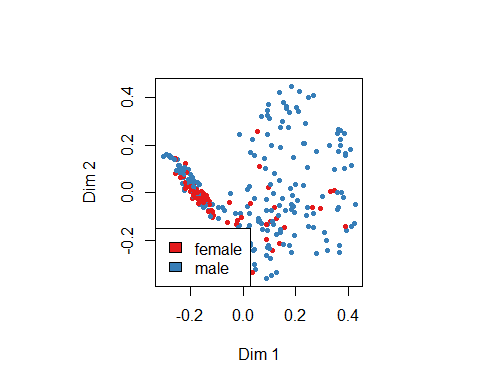
## female male MeanDecreaseAccuracy MeanDecreaseGini  
## cp -0.6610314 1.4267014 0.6845634 5.188772  
## trestbps 1.0541351 0.7545965 1.3923510 14.784688  
## chol 5.8569419 4.2976498 6.6673817 22.187160  
## fbs 1.4855495 0.7124783 1.4376188 2.427010  
## restecg -0.4036266 2.4049416 1.8686601 4.696343  
## thalach -1.5953821 12.3776084 10.0188377 18.553458  
## exang 2.7866481 0.6081735 2.2633732 2.794094  
## oldpeak 3.3036413 1.9688799 3.6220278 11.040999  
## slope 0.6910339 6.5022677 5.8717584 5.185078  
## ca 2.3016994 3.2123607 4.0642667 4.907264  
## thal 20.8149525 2.0971038 14.0808542 12.392322  
## target 14.7862890 1.3083545 10.6493201 7.343611  
## age 1.9068367 2.9443376 3.6189570 16.592915

For both males and females the thal, target and thalach are strong predictors.

MDSplot(Heart\_rf, factor(Heart$sex))

## Warning in RColorBrewer::brewer.pal(nlevs, "Set1"): minimal value for n is 3, returning requested palette with 3 different levels

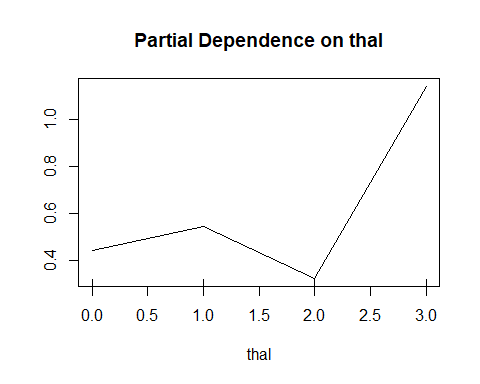
legend("bottomleft", legend=levels(Heart\_rf$predicted), fill=brewer.pal(3, "Set1")) #add a legend using base R



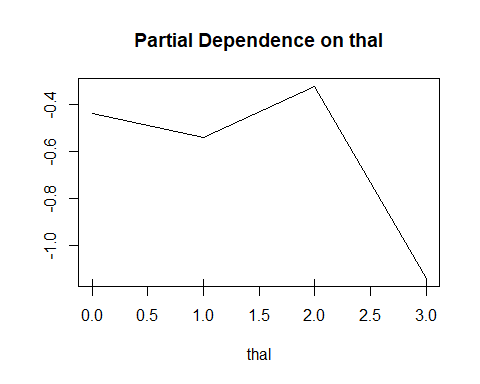
The female data seems to be a big cause of the error.

Now let’s view the effect of thal, target, and thalach

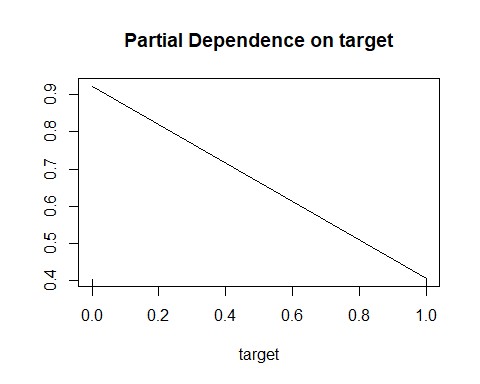
partialPlot(Heart\_rf, Heart[,2:14], thal, "male")



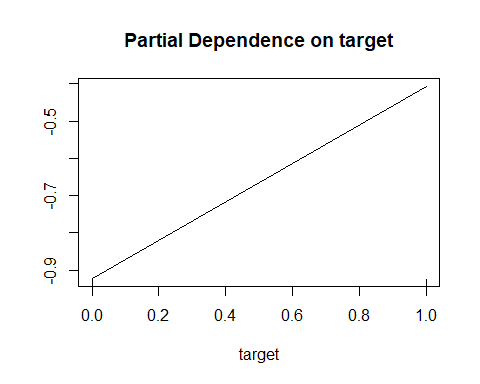
partialPlot(Heart\_rf, Heart[,2:14], thal, "female")



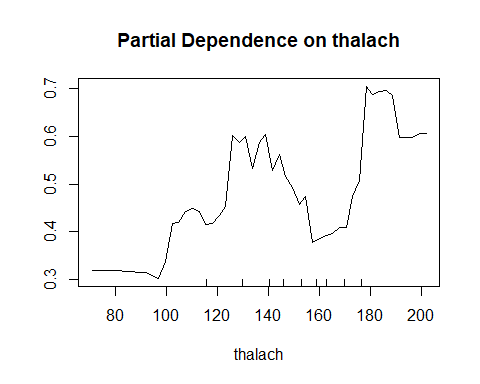
partialPlot(Heart\_rf, Heart[,2:14], target, "male")



partialPlot(Heart\_rf, Heart[,2:14], target, "female")



partialPlot(Heart\_rf, Heart[,2:14], thalach, "male")



partialPlot(Heart\_rf, Heart[,2:14], thalach, "female")

