setwd("C:/Users/ko7855/Documents/R Final Project")

library(tree)

library(ISLR)

library("readxl")

data<-read\_xlsx("Final Data\_Report 3.xlsx")

set.seed(1)

train = sample(1:nrow(data), nrow(data)/2)

colnames(data, do.NULL = FALSE)

colnames(data) <- c("TypeA","SourceA","Discharged","LengthStay","Age","RiskMortality","IllnessSev","FRisk","Beds")

data$FRisk<-as.factor(data$FRisk)

tree.data=tree(data$FRisk~.,data,subset=train)

summary(tree.data)

tree.data

plot(tree.data)

text(tree.data)

cv.data=cv.tree(tree.data)

plot(cv.data$size,cv.data$dev,type='b')

prune.data=prune.tree(tree.data,best=6)

plot(prune.data)

text(prune.data,pretty=0)

#the yhat plot is not functioning not sure if its necessaryl, xy lengths differ

yhat=as.numeric(predict(tree.data,newdata=data[-train,]))

data.test=data[-train,"FRisk"]

correlation mat

auc

roc