library(ggplot2)

library(cowplot)

library(reshape2)

library(caret)

library(Amelia)

library(e1071)

library(tidyverse)

data <- read.csv("data/data.csv")

label <- select(data,diagnosis)

cormat <- cor(subset(subset( data, select = - X ),select = - diagnosis))

melted\_cormat <- melt(cormat)

heatmap\_Plot <- ggplot(melted\_cormat, aes(x=Var1, y=Var2, fill=value))+

geom\_tile()+

scale\_fill\_gradientn(colours = c("black", "#5c0042", "#990159", "red", "#fee1b3", "#fef0be"), values = c(0,0.1,0.2,0.4,0.8,1), name = "")+

theme(axis.text.x=element\_text(angle=90,hjust=0.5,vjust=0.5))+

labs(x = "", y = "", title = "Heatmap")

ggsave("heatmap.png", width = 4000, height = 2500, units = "px")

############

data <- subset( data, select = - id )

data <- subset( data, select = - X )

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

############

set.seed(1)

partial\_data <- createDataPartition(y=data$diagnosis,p=0.8,list = FALSE)

train\_data <- data[partial\_data,]

test\_data <- data[-partial\_data,]

model <- glm(diagnosis~.,data = train\_data,family = "binomial")

summary(model)

############

prob <- predict(model , newdata = test\_data)

pred <- ifelse(prob>0 , "M", "B")

mean(test\_data$diagnosis == pred)

confusionMatrix(test\_data$diagnosis,as.factor(pred))

############

pca <- prcomp(data[3:31],center = TRUE, scale = TRUE)

summary(pca)

pca\_for\_heatmap <- as.data.frame(pca$x[,1:6])

pca\_cormat <- cor(pca\_for\_heatmap)

meltec\_pca\_cormat <- melt(pca\_cormat)

pca\_heatmap\_Plot <- ggplot(meltec\_pca\_cormat, aes(x=Var1, y=Var2, fill=value))+

geom\_tile()+

scale\_fill\_gradientn(colours = c("black", "#5c0042", "#990159", "red", "#fee1b3", "#fef0be"),values = c(0,0.1,0.2,0.4,0.8,1), name = "")+

theme(axis.text.x=element\_text(angle=90,hjust=0.5,vjust=0.5))+

labs(x = "", y = "", title = "PCAHeatmap")

ggsave("PCAHeatmap.png", width = 4000, height = 2500, units = "px")

############

set.seed(2)

pca\_for\_heatmap$diagnosis <- label$diagnosis

partial\_pca\_data <- createDataPartition(y=pca\_for\_heatmap$diagnosis,p=0.8,list = FALSE)

pca\_train\_data <- pca\_for\_heatmap[partial\_pca\_data,]

pca\_test\_data <- pca\_for\_heatmap[-partial\_pca\_data,]

pca\_model <- glm(diagnosis~.,data = pca\_train\_data,family = "binomial")

summary(pca\_model)

############

pca\_prob <- predict(pca\_model , newdata = pca\_test\_data)

pca\_pred <- ifelse(pca\_prob>0 , "M", "B")

mean(pca\_test\_data$diagnosis == pca\_pred)

confusionMatrix(pca\_test\_data$diagnosis,as.factor(pca\_pred))

############