ADM\_Project

Anastasios Vlaikidis

5/30/2021

# Libraries

#################### For data manipulation ###################  
library(tidyverse)  
  
#################### For graphics ############################  
library(ggplot2)  
library(gmodels)  
library(corrplot)  
library(factoextra)  
  
#################### Machine Learning Libraries ##############  
library(caret)  
library(glmnet)

# Load the dataset

divorce <- read.csv("divorce.csv",sep = ";")

# Sneak peak at our Data

head(divorce)

## Atr1 Atr2 Atr3 Atr4 Atr5 Atr6 Atr7 Atr8 Atr9 Atr10 Atr11 Atr12 Atr13 Atr14  
## 1 2 2 4 1 0 0 0 0 0 0 1 0 1 1  
## 2 4 4 4 4 4 0 0 4 4 4 4 3 4 0  
## 3 2 2 2 2 1 3 2 1 1 2 3 4 2 3  
## 4 3 2 3 2 3 3 3 3 3 3 4 3 3 4  
## 5 2 2 1 1 1 1 0 0 0 0 0 1 0 1  
## 6 0 0 1 0 0 2 0 0 0 1 0 2 1 0  
## Atr15 Atr16 Atr17 Atr18 Atr19 Atr20 Atr21 Atr22 Atr23 Atr24 Atr25 Atr26 Atr27  
## 1 0 1 0 0 0 1 0 0 0 0 0 0 0  
## 2 4 4 4 4 3 2 1 1 0 2 2 1 2  
## 3 3 3 3 3 3 2 1 0 1 2 2 2 2  
## 4 3 3 3 3 3 4 1 1 1 1 2 1 1  
## 5 1 1 1 1 2 1 1 0 0 0 0 2 1  
## 6 2 0 2 1 0 1 0 0 0 0 2 2 0  
## Atr28 Atr29 Atr30 Atr31 Atr32 Atr33 Atr34 Atr35 Atr36 Atr37 Atr38 Atr39 Atr40  
## 1 0 0 1 1 2 1 2 0 1 2 1 3 3  
## 2 0 1 1 0 4 2 3 0 2 3 4 2 4  
## 3 2 3 2 3 3 1 1 1 1 2 1 3 3  
## 4 1 1 3 2 3 2 2 1 1 3 3 4 4  
## 5 2 1 1 1 1 1 1 0 0 0 0 2 1  
## 6 0 0 0 4 1 1 1 1 1 1 2 0 2  
## Atr41 Atr42 Atr43 Atr44 Atr45 Atr46 Atr47 Atr48 Atr49 Atr50 Atr51 Atr52 Atr53  
## 1 2 1 1 2 3 2 1 3 3 3 2 3 2  
## 2 2 2 3 4 2 2 2 3 4 4 4 4 2  
## 3 3 3 2 3 2 3 2 3 1 1 1 2 2  
## 4 2 2 3 2 3 2 2 3 3 3 3 2 2  
## 5 0 2 3 0 2 2 1 2 3 2 2 2 1  
## 6 2 1 2 3 0 2 2 1 2 1 1 1 2  
## Atr54 Class  
## 1 1 1  
## 2 2 1  
## 3 2 1  
## 4 2 1  
## 5 0 1  
## 6 0 1

tail(divorce)

## Atr1 Atr2 Atr3 Atr4 Atr5 Atr6 Atr7 Atr8 Atr9 Atr10 Atr11 Atr12 Atr13 Atr14  
## 165 2 1 1 0 0 2 0 0 0 2 0 1 1 1  
## 166 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## 167 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## 168 1 1 0 0 0 0 0 0 0 1 0 1 1 0  
## 169 0 0 0 0 0 0 0 0 0 0 0 1 0 0  
## 170 0 0 0 0 0 0 0 1 0 0 0 1 1 1  
## Atr15 Atr16 Atr17 Atr18 Atr19 Atr20 Atr21 Atr22 Atr23 Atr24 Atr25 Atr26  
## 165 1 0 0 0 0 0 0 0 0 0 1 1  
## 166 0 0 0 0 0 0 0 0 0 0 0 0  
## 167 0 0 0 0 0 0 0 0 0 0 0 0  
## 168 0 1 0 0 0 1 0 0 0 0 1 1  
## 169 0 0 0 0 0 0 0 0 0 0 0 0  
## 170 0 0 0 0 0 1 1 0 1 1 0 0  
## Atr27 Atr28 Atr29 Atr30 Atr31 Atr32 Atr33 Atr34 Atr35 Atr36 Atr37 Atr38  
## 165 0 0 0 1 1 1 2 2 0 0 0 0  
## 166 0 0 0 0 4 4 3 4 0 0 4 0  
## 167 0 0 0 0 0 1 1 1 1 1 1 1  
## 168 1 0 0 1 1 1 0 1 0 0 1 1  
## 169 0 0 0 0 0 0 0 1 0 0 1 0  
## 170 0 0 0 0 1 3 0 0 0 0 0 0  
## Atr39 Atr40 Atr41 Atr42 Atr43 Atr44 Atr45 Atr46 Atr47 Atr48 Atr49 Atr50  
## 165 0 0 1 0 1 2 0 1 0 3 1 1  
## 166 1 0 1 0 0 0 0 1 0 4 1 1  
## 167 1 1 1 1 3 1 3 4 1 2 2 2  
## 168 1 2 1 3 3 0 2 3 0 2 0 1  
## 169 4 1 2 1 1 0 4 3 3 2 2 3  
## 170 0 0 0 2 2 0 1 3 4 4 0 1  
## Atr51 Atr52 Atr53 Atr54 Class  
## 165 3 1 1 1 0  
## 166 4 2 2 2 0  
## 167 2 3 2 2 0  
## 168 1 3 0 0 0  
## 169 2 4 3 1 0  
## 170 3 3 3 1 0

head(divorce[,53:55],n=2)

## Atr53 Atr54 Class  
## 1 2 1 1  
## 2 2 2 1

tail(divorce[,53:55],n=2)

## Atr53 Atr54 Class  
## 169 3 1 0  
## 170 3 1 0

# Exploring data by using summary statistics

## Type of the data

SLD <- divorce  
# command which reveals how the object's storage is implemented  
cat("How object's storage is implemented:",typeof(SLD),"\n")

## How object's storage is implemented: list

# what kind of object we have  
cat("What kind of object's we have:",class(SLD),"\n")

## What kind of object's we have: data.frame

# dimension of dataset  
cat("Dimensions of our object",dim(SLD),"\n")

## Dimensions of our object 170 55

## Structure of our data

str(SLD)

## 'data.frame': 170 obs. of 55 variables:  
## $ Atr1 : int 2 4 2 3 2 0 3 2 2 1 ...  
## $ Atr2 : int 2 4 2 2 2 0 3 1 2 1 ...  
## $ Atr3 : int 4 4 2 3 1 1 3 2 1 1 ...  
## $ Atr4 : int 1 4 2 2 1 0 2 2 0 1 ...  
## $ Atr5 : int 0 4 1 3 1 0 1 2 0 1 ...  
## $ Atr6 : int 0 0 3 3 1 2 3 1 4 2 ...  
## $ Atr7 : int 0 0 2 3 0 0 4 0 1 0 ...  
## $ Atr8 : int 0 4 1 3 0 0 3 3 3 2 ...  
## $ Atr9 : int 0 4 1 3 0 0 2 3 3 2 ...  
## $ Atr10: int 0 4 2 3 0 1 2 2 3 2 ...  
## $ Atr11: int 1 4 3 4 0 0 2 4 3 3 ...  
## $ Atr12: int 0 3 4 3 1 2 2 3 3 0 ...  
## $ Atr13: int 1 4 2 3 0 1 2 2 3 0 ...  
## $ Atr14: int 1 0 3 4 1 0 3 3 3 2 ...  
## $ Atr15: int 0 4 3 3 1 2 2 4 3 1 ...  
## $ Atr16: int 1 4 3 3 1 0 3 3 3 0 ...  
## $ Atr17: int 0 4 3 3 1 2 3 2 3 1 ...  
## $ Atr18: int 0 4 3 3 1 1 3 3 3 2 ...  
## $ Atr19: int 0 3 3 3 2 0 3 2 3 1 ...  
## $ Atr20: int 1 2 2 4 1 1 2 1 3 0 ...  
## $ Atr21: int 0 1 1 1 1 0 3 2 2 0 ...  
## $ Atr22: int 0 1 0 1 0 0 3 1 2 0 ...  
## $ Atr23: int 0 0 1 1 0 0 3 1 2 0 ...  
## $ Atr24: int 0 2 2 1 0 0 3 2 3 1 ...  
## $ Atr25: int 0 2 2 2 0 2 2 3 2 1 ...  
## $ Atr26: int 0 1 2 1 2 2 3 3 3 1 ...  
## $ Atr27: int 0 2 2 1 1 0 3 2 2 1 ...  
## $ Atr28: int 0 0 2 1 2 0 2 2 3 1 ...  
## $ Atr29: int 0 1 3 1 1 0 2 2 2 1 ...  
## $ Atr30: int 1 1 2 3 1 0 2 3 3 1 ...  
## $ Atr31: int 1 0 3 2 1 4 1 1 1 1 ...  
## $ Atr32: int 2 4 3 3 1 1 2 1 1 1 ...  
## $ Atr33: int 1 2 1 2 1 1 2 0 1 0 ...  
## $ Atr34: int 2 3 1 2 1 1 1 2 1 1 ...  
## $ Atr35: int 0 0 1 1 0 1 1 2 1 0 ...  
## $ Atr36: int 1 2 1 1 0 1 2 1 1 0 ...  
## $ Atr37: int 2 3 2 3 0 1 3 4 1 1 ...  
## $ Atr38: int 1 4 1 3 0 2 2 4 2 1 ...  
## $ Atr39: int 3 2 3 4 2 0 2 4 2 2 ...  
## $ Atr40: int 3 4 3 4 1 2 3 4 2 2 ...  
## $ Atr41: int 2 2 3 2 0 2 3 4 2 1 ...  
## $ Atr42: int 1 2 3 2 2 1 3 4 2 2 ...  
## $ Atr43: int 1 3 2 3 3 2 3 3 2 3 ...  
## $ Atr44: int 2 4 3 2 0 3 4 2 2 2 ...  
## $ Atr45: int 3 2 2 3 2 0 3 0 2 2 ...  
## $ Atr46: int 2 2 3 2 2 2 3 0 1 2 ...  
## $ Atr47: int 1 2 2 2 1 2 2 1 1 0 ...  
## $ Atr48: int 3 3 3 3 2 1 3 2 1 2 ...  
## $ Atr49: int 3 4 1 3 3 2 2 2 1 2 ...  
## $ Atr50: int 3 4 1 3 2 1 3 2 1 2 ...  
## $ Atr51: int 2 4 1 3 2 1 3 1 1 2 ...  
## $ Atr52: int 3 4 2 2 2 1 2 1 1 4 ...  
## $ Atr53: int 2 2 2 2 1 2 2 1 1 3 ...  
## $ Atr54: int 1 2 2 2 0 0 2 0 1 3 ...  
## $ Class: int 1 1 1 1 1 1 1 1 1 1 ...

str(SLD[,53:55])

## 'data.frame': 170 obs. of 3 variables:  
## $ Atr53: int 2 2 2 2 1 2 2 1 1 3 ...  
## $ Atr54: int 1 2 2 2 0 0 2 0 1 3 ...  
## $ Class: int 1 1 1 1 1 1 1 1 1 1 ...

## Summary of our data

summary(SLD)

## Atr1 Atr2 Atr3 Atr4   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
## Median :2.000 Median :2.000 Median :2.000 Median :1.000   
## Mean :1.776 Mean :1.653 Mean :1.765 Mean :1.482   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr5 Atr6 Atr7 Atr8   
## Min. :0.000 Min. :0.0000 Min. :0.0000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.000   
## Median :1.000 Median :0.0000 Median :0.0000 Median :1.000   
## Mean :1.541 Mean :0.7471 Mean :0.4941 Mean :1.453   
## 3rd Qu.:3.000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:3.000   
## Max. :4.000 Max. :4.0000 Max. :4.0000 Max. :4.000   
## Atr9 Atr10 Atr11 Atr12   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
## Median :1.000 Median :2.000 Median :1.000 Median :1.500   
## Mean :1.459 Mean :1.576 Mean :1.688 Mean :1.653   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr13 Atr14 Atr15 Atr16   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
## Median :2.000 Median :1.000 Median :1.000 Median :1.000   
## Mean :1.835 Mean :1.571 Mean :1.571 Mean :1.476   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr17 Atr18 Atr19 Atr20   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
## Median :1.000 Median :1.000 Median :1.000 Median :1.000   
## Mean :1.653 Mean :1.518 Mean :1.641 Mean :1.459   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr21 Atr22 Atr23 Atr24   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
## Median :1.000 Median :0.000 Median :0.000 Median :1.000   
## Mean :1.388 Mean :1.247 Mean :1.412 Mean :1.512   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr25 Atr26 Atr27 Atr28 Atr29   
## Min. :0.000 Min. :0.000 Min. :0.0 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.0 1st Qu.:0.000 1st Qu.:0.000   
## Median :1.000 Median :1.000 Median :1.0 Median :0.500 Median :1.000   
## Mean :1.629 Mean :1.488 Mean :1.4 Mean :1.306 Mean :1.494   
## 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:3.0 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :4.000 Max. :4.000 Max. :4.0 Max. :4.000 Max. :4.000   
## Atr30 Atr31 Atr32 Atr33 Atr34   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.0   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.0   
## Median :1.000 Median :2.000 Median :2.000 Median :1.000 Median :1.0   
## Mean :1.494 Mean :2.124 Mean :2.059 Mean :1.806 Mean :1.9   
## 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.0   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.0   
## Atr35 Atr36 Atr37 Atr38   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
## Median :0.500 Median :0.000 Median :2.000 Median :1.000   
## Mean :1.671 Mean :1.606 Mean :2.088 Mean :1.859   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr39 Atr40 Atr41 Atr42   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000 1st Qu.:0.000   
## Median :2.000 Median :1.500 Median :2.000 Median :2.000   
## Mean :2.088 Mean :1.871 Mean :1.994 Mean :2.159   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr43 Atr44 Atr45 Atr46   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:2.000 1st Qu.:0.000 1st Qu.:1.000 1st Qu.:2.000   
## Median :3.000 Median :2.000 Median :3.000 Median :3.000   
## Mean :2.706 Mean :1.941 Mean :2.459 Mean :2.553   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr47 Atr48 Atr49 Atr50   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :2.000 Median :3.000 Median :3.000 Median :2.000   
## Mean :2.271 Mean :2.741 Mean :2.382 Mean :2.429   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Atr51 Atr52 Atr53 Atr54   
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000   
## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:0.000   
## Median :3.000 Median :3.000 Median :2.000 Median :2.000   
## Mean :2.476 Mean :2.518 Mean :2.241 Mean :2.012   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000   
## Class   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.4941   
## 3rd Qu.:1.0000   
## Max. :1.0000

summary(SLD[,53:55])

## Atr53 Atr54 Class   
## Min. :0.000 Min. :0.000 Min. :0.0000   
## 1st Qu.:1.000 1st Qu.:0.000 1st Qu.:0.0000   
## Median :2.000 Median :2.000 Median :0.0000   
## Mean :2.241 Mean :2.012 Mean :0.4941   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:1.0000   
## Max. :4.000 Max. :4.000 Max. :1.0000

## Which of our variables are categorical

outcome <- "Q1821" # random string not the same as colnames  
vars <- setdiff(colnames(SLD), outcome)  
is\_categorical <- vapply(SLD[, vars],  
 function(v) !is.numeric(v),  
 logical(1))  
  
is\_categorical

## Atr1 Atr2 Atr3 Atr4 Atr5 Atr6 Atr7 Atr8 Atr9 Atr10 Atr11 Atr12 Atr13   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## Atr14 Atr15 Atr16 Atr17 Atr18 Atr19 Atr20 Atr21 Atr22 Atr23 Atr24 Atr25 Atr26   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## Atr27 Atr28 Atr29 Atr30 Atr31 Atr32 Atr33 Atr34 Atr35 Atr36 Atr37 Atr38 Atr39   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## Atr40 Atr41 Atr42 Atr43 Atr44 Atr45 Atr46 Atr47 Atr48 Atr49 Atr50 Atr51 Atr52   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## Atr53 Atr54 Class   
## FALSE FALSE FALSE

### Summary of categorical data

summary(is\_categorical)

## Mode FALSE   
## logical 55

### Count the levels of categorical variables

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
##

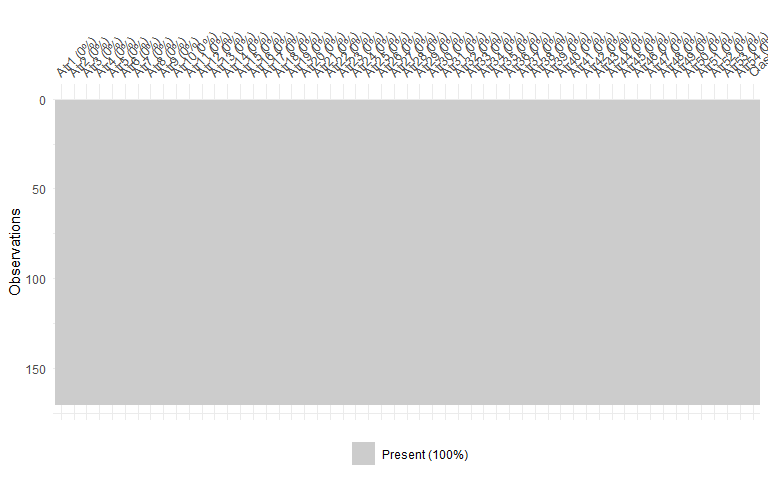
## Number of missing data

## Atr1 Atr2 Atr3 Atr4 Atr5 Atr6 Atr7 Atr8 Atr9 Atr10 Atr11 Atr12 Atr13   
## 0 0 0 0 0 0 0 0 0 0 0 0 0   
## Atr14 Atr15 Atr16 Atr17 Atr18 Atr19 Atr20 Atr21 Atr22 Atr23 Atr24 Atr25 Atr26   
## 0 0 0 0 0 0 0 0 0 0 0 0 0   
## Atr27 Atr28 Atr29 Atr30 Atr31 Atr32 Atr33 Atr34 Atr35 Atr36 Atr37 Atr38 Atr39   
## 0 0 0 0 0 0 0 0 0 0 0 0 0   
## Atr40 Atr41 Atr42 Atr43 Atr44 Atr45 Atr46 Atr47 Atr48 Atr49 Atr50 Atr51 Atr52   
## 0 0 0 0 0 0 0 0 0 0 0 0 0   
## Atr53 Atr54 Class   
## 0 0 0

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0 0 0 0 0 0

## Visualize missing values

visdat::vis\_miss(SLD, cluster = T)



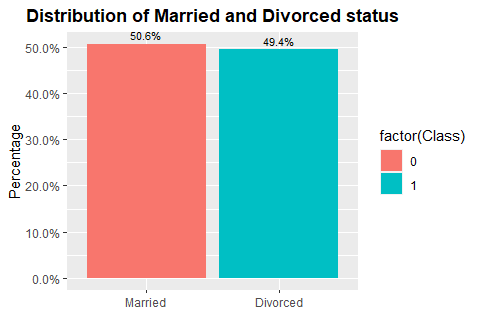
# Data visualization

## Target variable distribution

SLD$Class <- factor(SLD$Class)  
CrossTable(SLD$Class)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 170   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 86 | 84 |   
## | 0.506 | 0.494 |   
## |-----------|-----------|  
##   
##   
##   
##

common\_theme <- theme(plot.title =   
 element\_text(hjust = 0.5,   
 face = "bold"))  
  
ggplot(data = SLD,   
 aes(x = factor(Class),   
 y = prop.table(stat(count)),   
 fill = factor(Class),  
 label = scales::percent(prop.table(stat(count)))))+  
   
 xlab(NULL)+  
  
 ylab("Percentage")+  
   
 geom\_bar(position = "dodge")+   
   
 geom\_text(stat = 'count',  
 position = position\_dodge(.9),   
 vjust = -0.5,   
 size = 3)+  
   
 scale\_x\_discrete(labels = c("Married","Divorced"))+  
   
 scale\_y\_continuous(labels = scales::percent)+  
   
 ggtitle("Distribution of Married and Divorced status")+  
 common\_theme



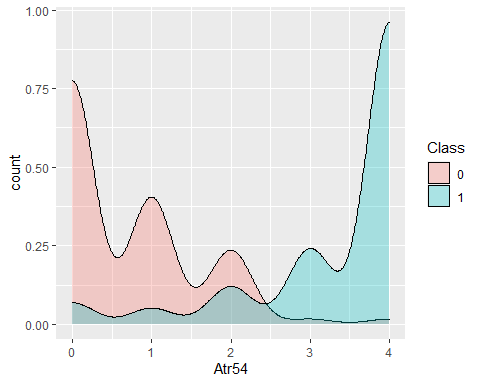
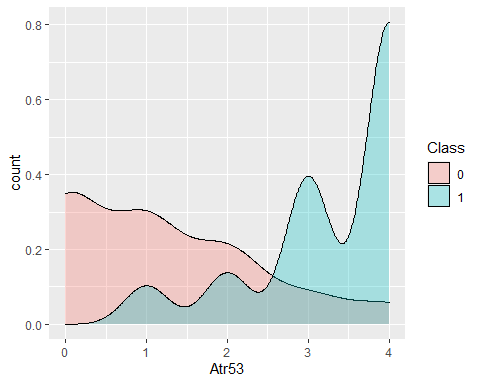
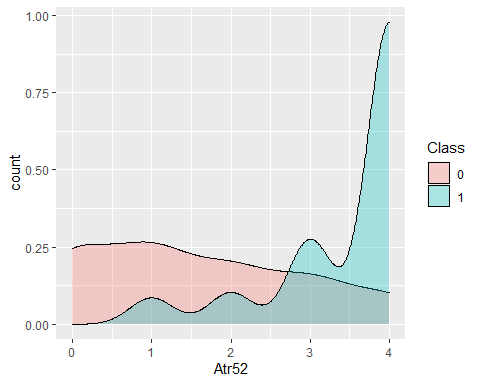
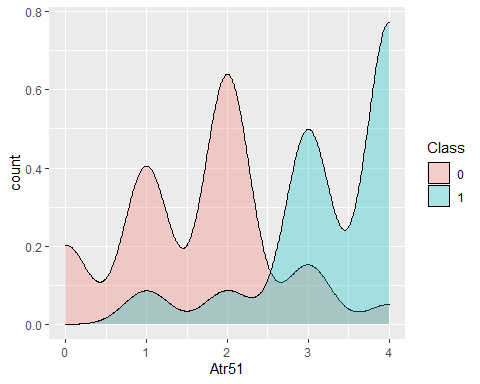
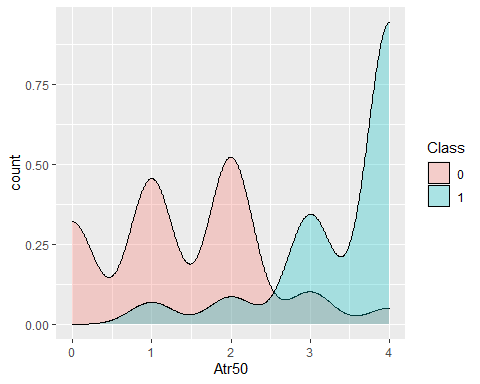
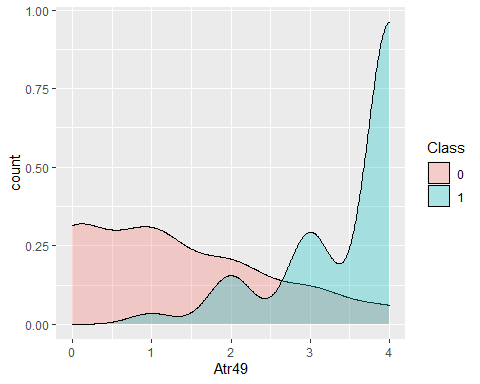
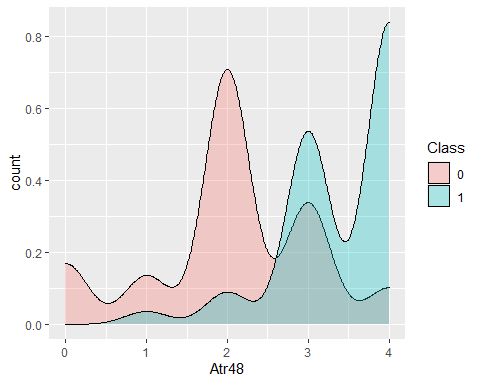
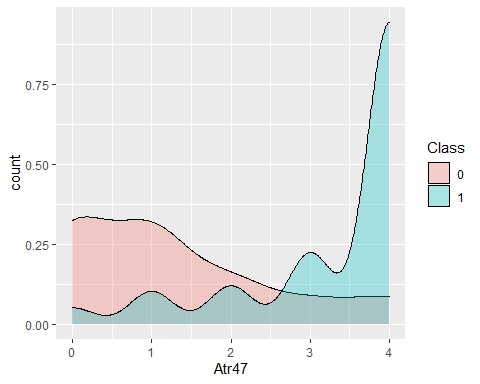
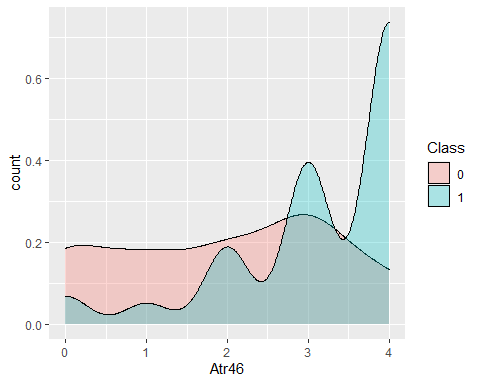
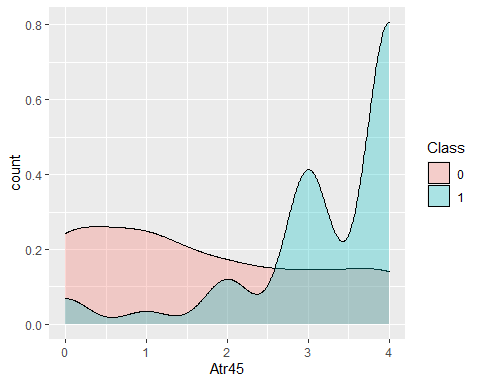
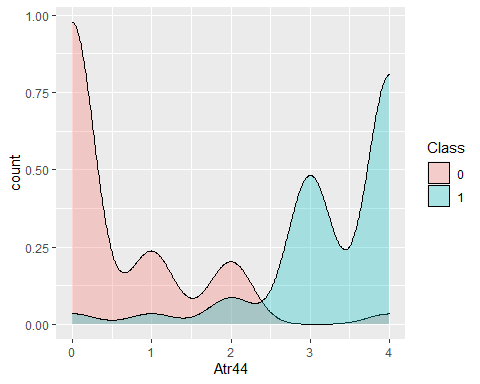
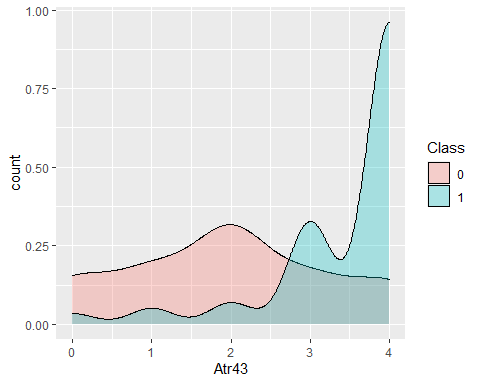
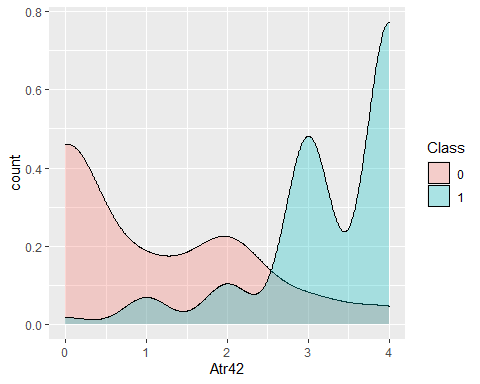
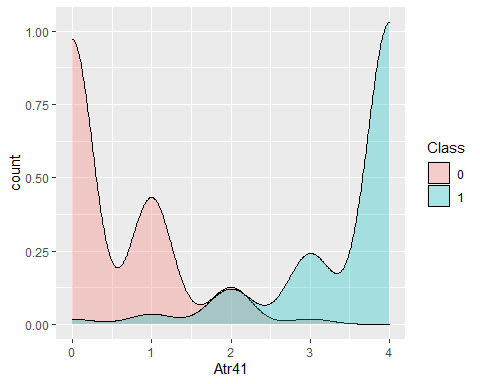
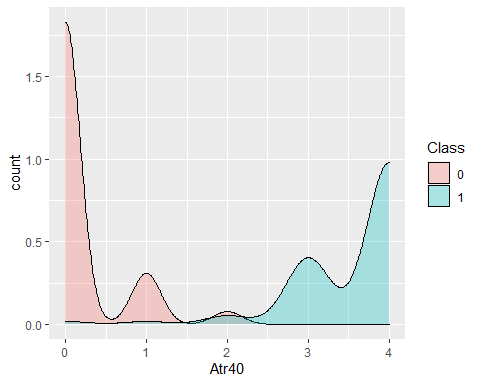
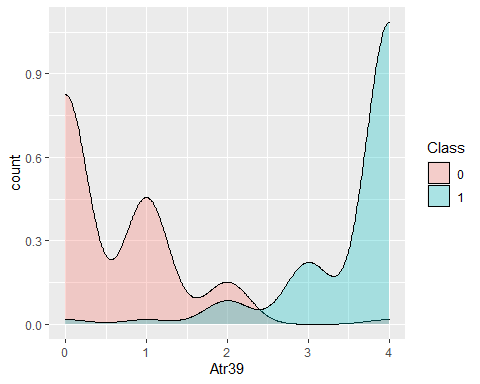
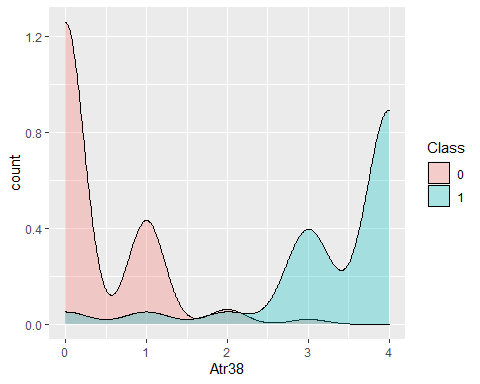
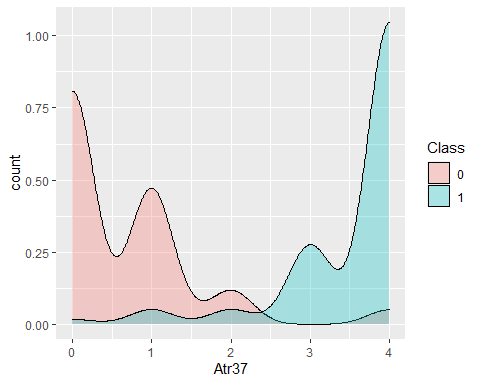
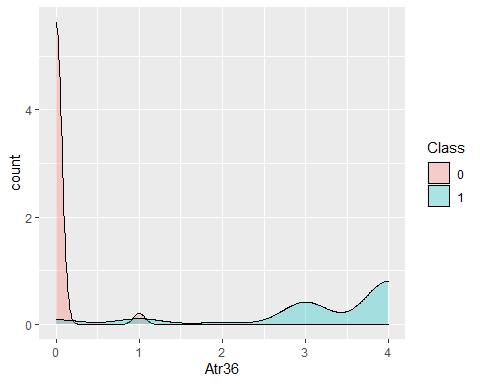
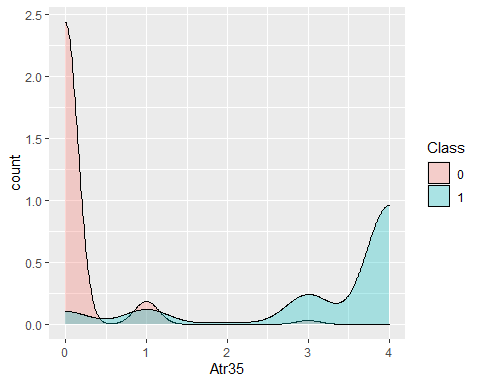
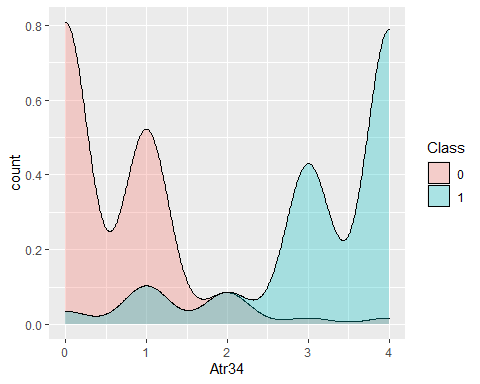
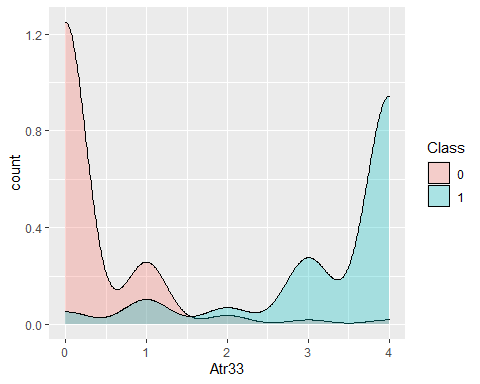
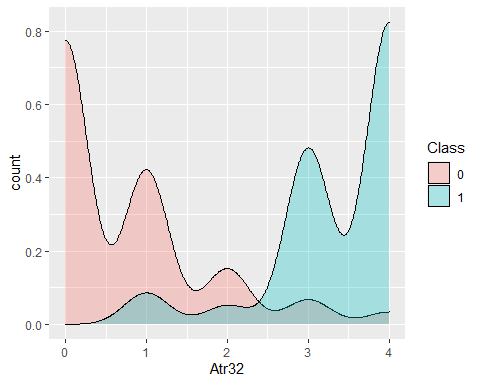
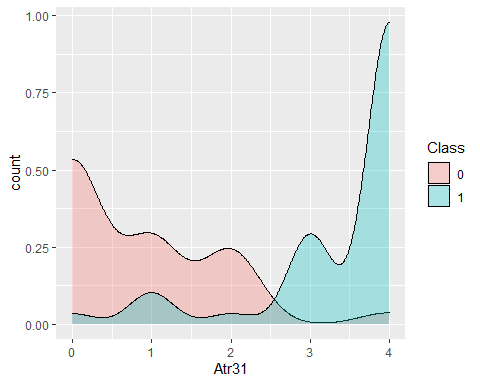
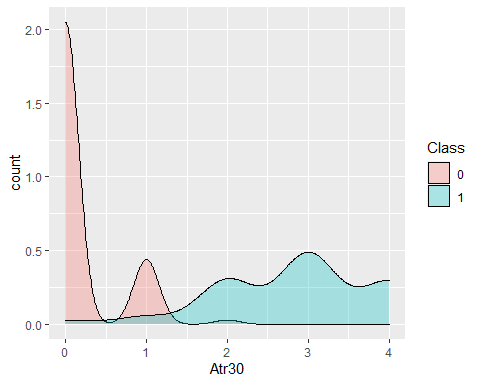
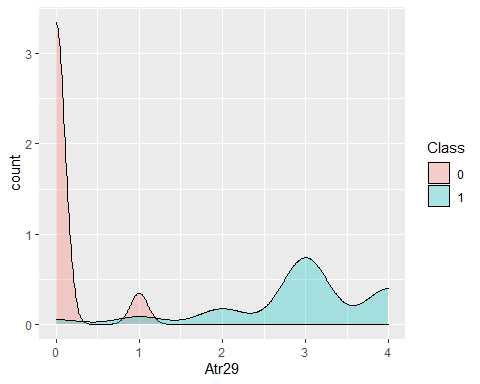
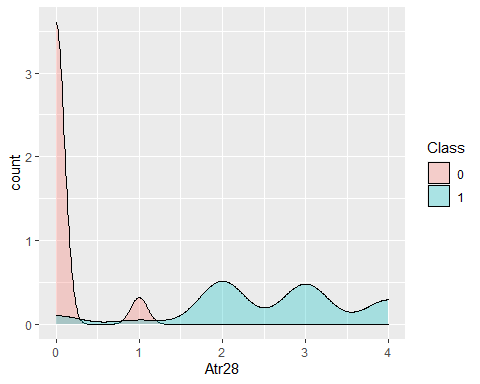
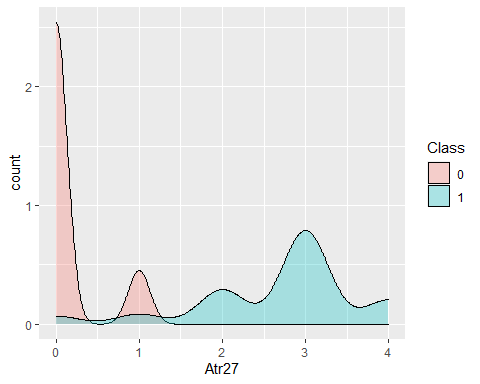
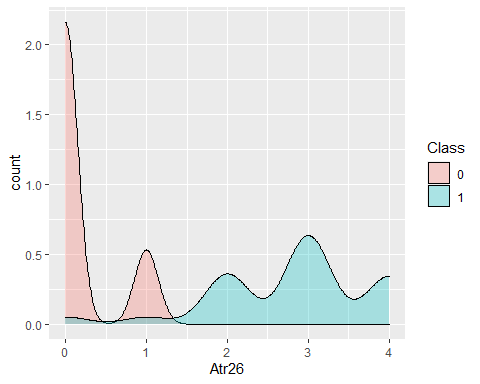
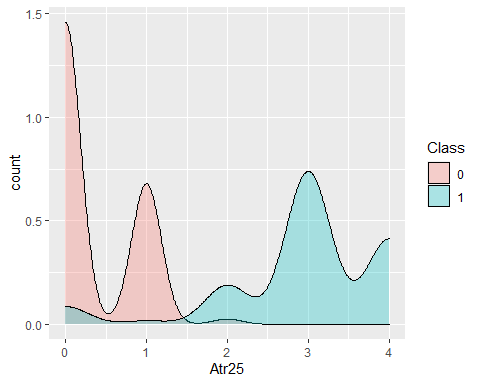
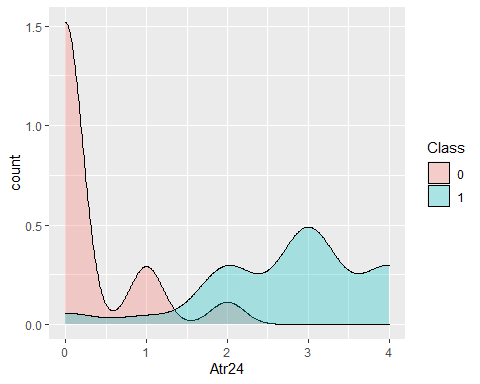
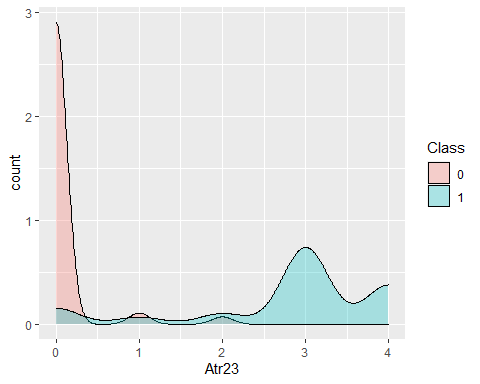
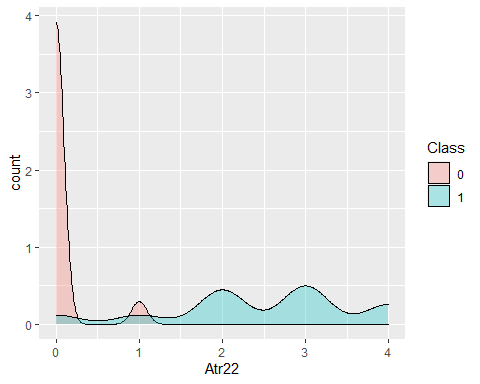
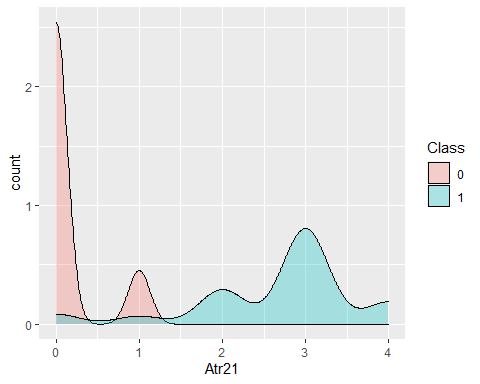
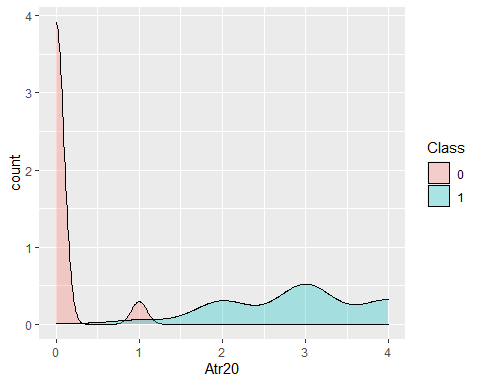
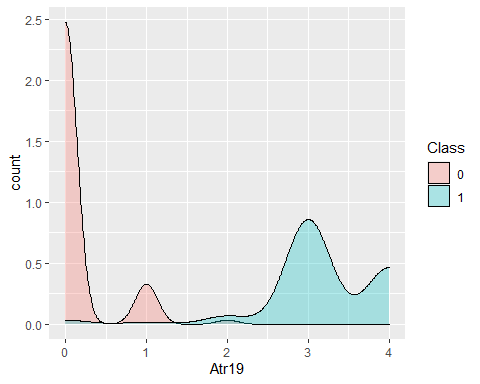
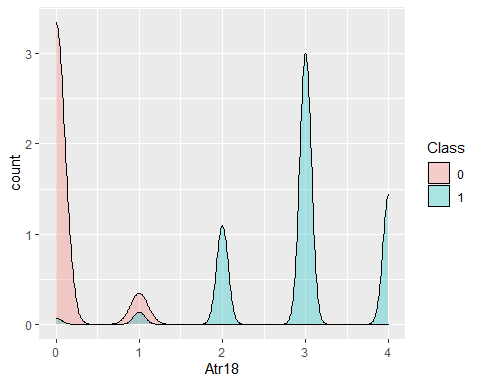
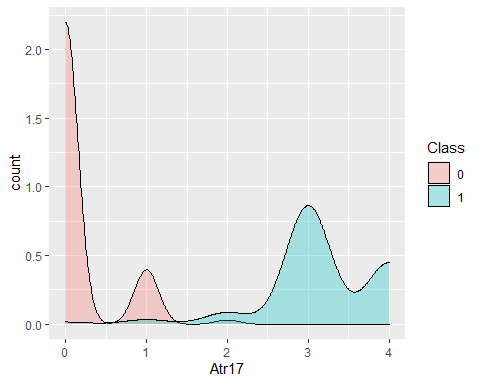
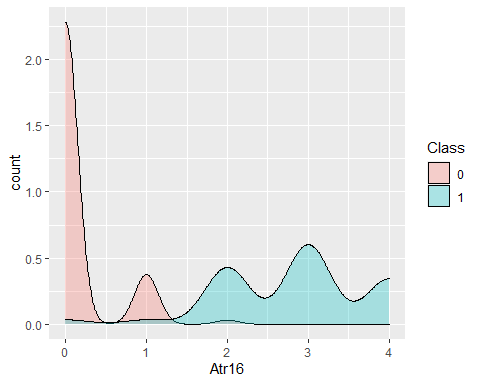
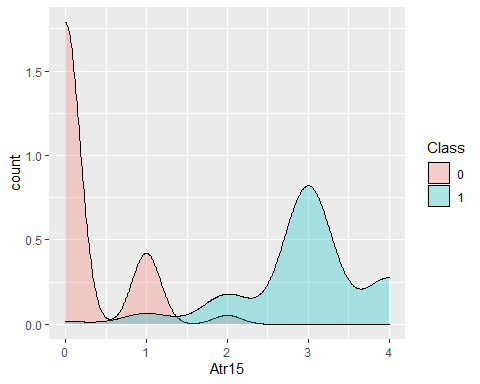
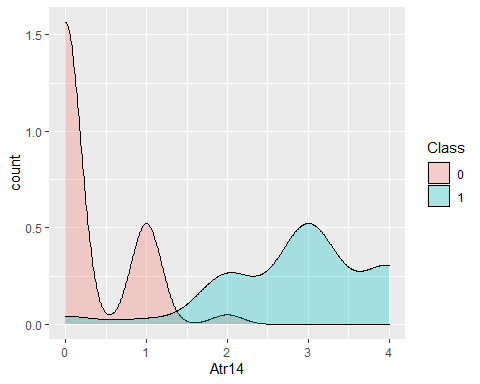
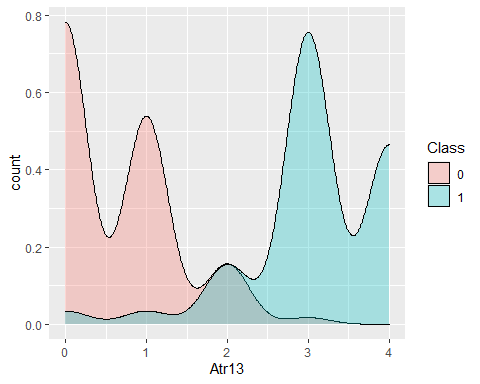
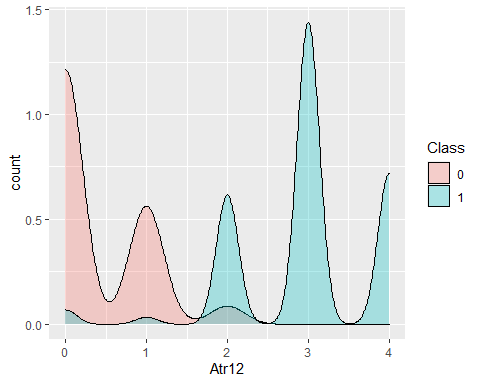
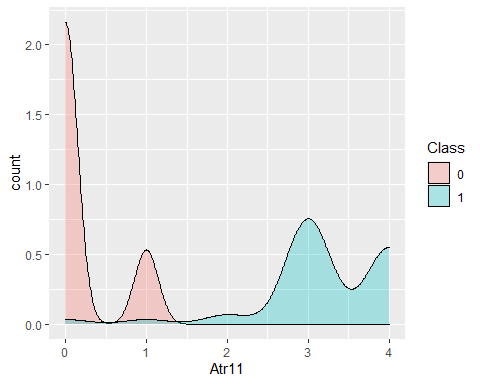
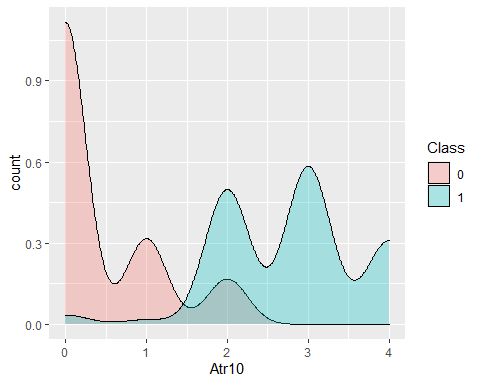
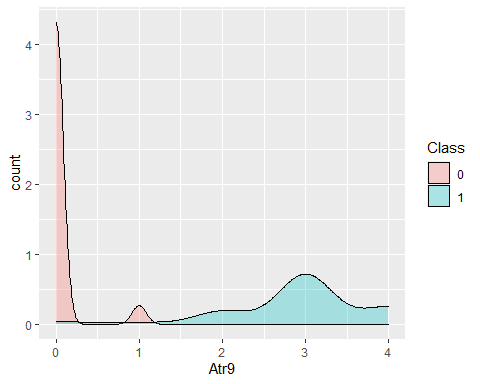
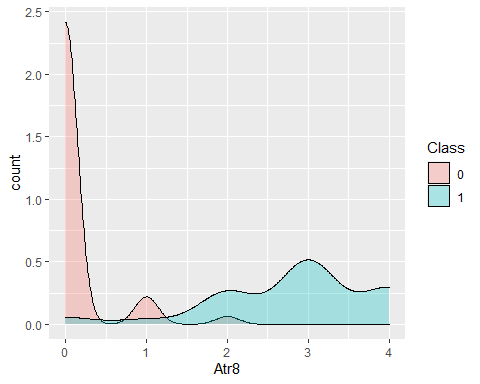
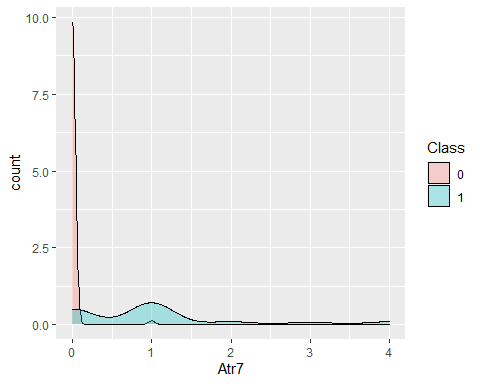
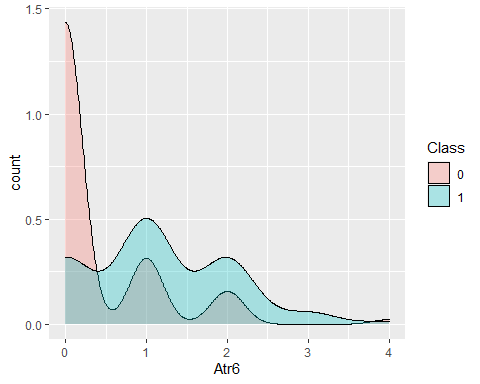
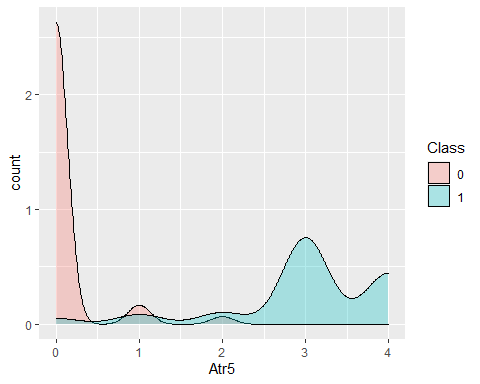
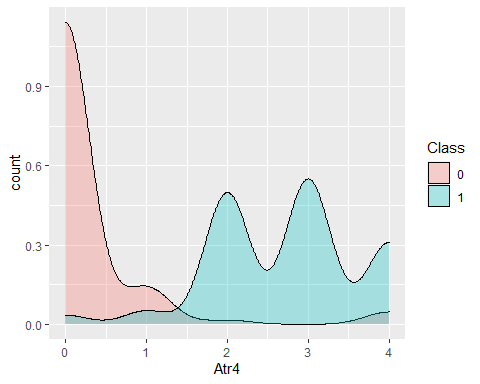
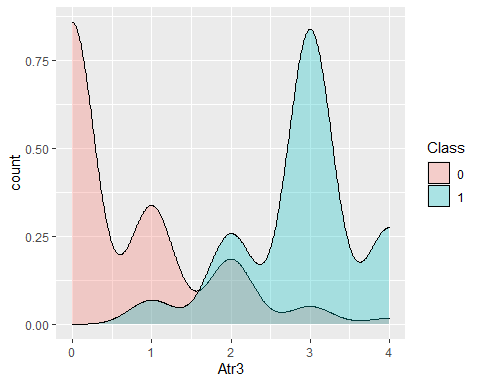
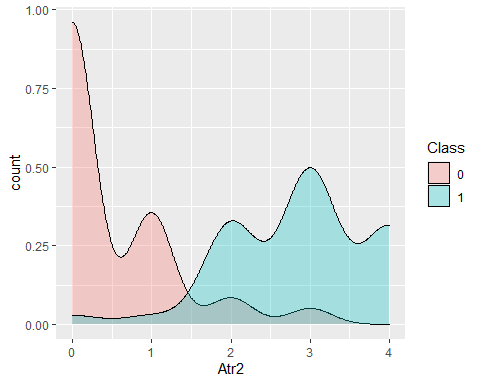
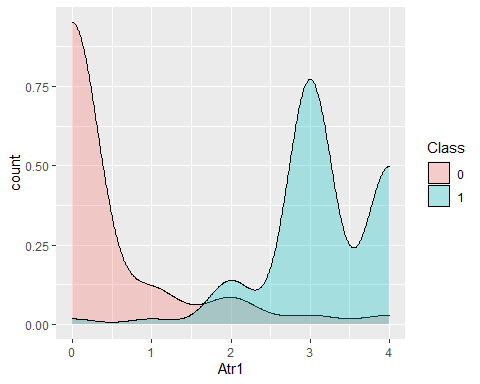
## Crosstables

for (i in 1:54){  
 print(names(SLD)[i])  
 CrossTable(SLD[,i],  
 SLD$Class,   
 prop.c = F,  
 format ="SPSS",  
 prop.t = F,  
 digits = 2,  
 prop.chisq = F)   
}

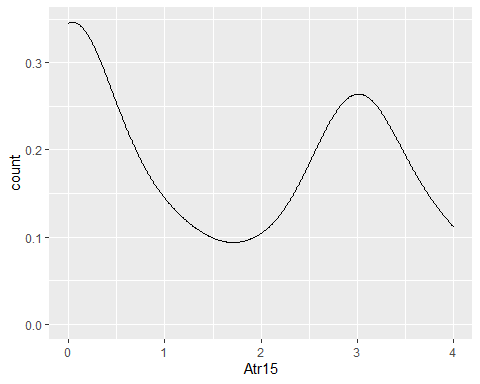
## [1] "Atr1"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 68 | 1 | 69 |   
## | 98.55% | 1.45% | 40.59% |   
## -------------|-----------|-----------|-----------|  
## 1 | 8 | 1 | 9 |   
## | 88.89% | 11.11% | 5.29% |   
## -------------|-----------|-----------|-----------|  
## 2 | 6 | 8 | 14 |   
## | 42.86% | 57.14% | 8.24% |   
## -------------|-----------|-----------|-----------|  
## 3 | 2 | 45 | 47 |   
## | 4.26% | 95.74% | 27.65% |   
## -------------|-----------|-----------|-----------|  
## 4 | 2 | 29 | 31 |   
## | 6.45% | 93.55% | 18.24% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr2"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 57 | 2 | 59 |   
## | 96.61% | 3.39% | 34.71% |   
## -------------|-----------|-----------|-----------|  
## 1 | 21 | 2 | 23 |   
## | 91.30% | 8.70% | 13.53% |   
## -------------|-----------|-----------|-----------|  
## 2 | 5 | 23 | 28 |   
## | 17.86% | 82.14% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 3 | 3 | 35 | 38 |   
## | 7.89% | 92.11% | 22.35% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 22 | 22 |   
## | 0.00% | 100.00% | 12.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr3"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 51 | 0 | 51 |   
## | 100.00% | 0.00% | 30.00% |   
## -------------|-----------|-----------|-----------|  
## 1 | 20 | 4 | 24 |   
## | 83.33% | 16.67% | 14.12% |   
## -------------|-----------|-----------|-----------|  
## 2 | 11 | 15 | 26 |   
## | 42.31% | 57.69% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 3 | 49 | 52 |   
## | 5.77% | 94.23% | 30.59% |   
## -------------|-----------|-----------|-----------|  
## 4 | 1 | 16 | 17 |   
## | 5.88% | 94.12% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr4"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 73 | 2 | 75 |   
## | 97.33% | 2.67% | 44.12% |   
## -------------|-----------|-----------|-----------|  
## 1 | 9 | 3 | 12 |   
## | 75.00% | 25.00% | 7.06% |   
## -------------|-----------|-----------|-----------|  
## 2 | 1 | 29 | 30 |   
## | 3.33% | 96.67% | 17.65% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 32 | 32 |   
## | 0.00% | 100.00% | 18.82% |   
## -------------|-----------|-----------|-----------|  
## 4 | 3 | 18 | 21 |   
## | 14.29% | 85.71% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr5"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 79 | 3 | 82 |   
## | 96.34% | 3.66% | 48.24% |   
## -------------|-----------|-----------|-----------|  
## 1 | 5 | 5 | 10 |   
## | 50.00% | 50.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 2 | 6 | 8 |   
## | 25.00% | 75.00% | 4.71% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 44 | 44 |   
## | 0.00% | 100.00% | 25.88% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 26 | 26 |   
## | 0.00% | 100.00% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr6"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 64 | 22 | 86 |   
## | 74.42% | 25.58% | 50.59% |   
## -------------|-----------|-----------|-----------|  
## 1 | 14 | 35 | 49 |   
## | 28.57% | 71.43% | 28.82% |   
## -------------|-----------|-----------|-----------|  
## 2 | 7 | 22 | 29 |   
## | 24.14% | 75.86% | 17.06% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 4 | 4 |   
## | 0.00% | 100.00% | 2.35% |   
## -------------|-----------|-----------|-----------|  
## 4 | 1 | 1 | 2 |   
## | 50.00% | 50.00% | 1.18% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr7"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 85 | 29 | 114 |   
## | 74.56% | 25.44% | 67.06% |   
## -------------|-----------|-----------|-----------|  
## 1 | 1 | 41 | 42 |   
## | 2.38% | 97.62% | 24.71% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 5 | 5 |   
## | 0.00% | 100.00% | 2.94% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 4 | 4 |   
## | 0.00% | 100.00% | 2.35% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 5 | 5 |   
## | 0.00% | 100.00% | 2.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr8"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 77 | 4 | 81 |   
## | 95.06% | 4.94% | 47.65% |   
## -------------|-----------|-----------|-----------|  
## 1 | 7 | 3 | 10 |   
## | 70.00% | 30.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 2 | 19 | 21 |   
## | 9.52% | 90.48% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 37 | 37 |   
## | 0.00% | 100.00% | 21.76% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 21 | 21 |   
## | 0.00% | 100.00% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr9"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 81 | 3 | 84 |   
## | 96.43% | 3.57% | 49.41% |   
## -------------|-----------|-----------|-----------|  
## 1 | 5 | 2 | 7 |   
## | 71.43% | 28.57% | 4.12% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 13 | 13 |   
## | 0.00% | 100.00% | 7.65% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 49 | 49 |   
## | 0.00% | 100.00% | 28.82% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 17 | 17 |   
## | 0.00% | 100.00% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr10"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 60 | 2 | 62 |   
## | 96.77% | 3.23% | 36.47% |   
## -------------|-----------|-----------|-----------|  
## 1 | 17 | 1 | 18 |   
## | 94.44% | 5.56% | 10.59% |   
## -------------|-----------|-----------|-----------|  
## 2 | 9 | 29 | 38 |   
## | 23.68% | 76.32% | 22.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 34 | 34 |   
## | 0.00% | 100.00% | 20.00% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 18 | 18 |   
## | 0.00% | 100.00% | 10.59% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr11"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 69 | 2 | 71 |   
## | 97.18% | 2.82% | 41.76% |   
## -------------|-----------|-----------|-----------|  
## 1 | 17 | 2 | 19 |   
## | 89.47% | 10.53% | 11.18% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 4 | 4 |   
## | 0.00% | 100.00% | 2.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 44 | 44 |   
## | 0.00% | 100.00% | 25.88% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 32 | 32 |   
## | 0.00% | 100.00% | 18.82% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr12"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 56 | 2 | 58 |   
## | 96.55% | 3.45% | 34.12% |   
## -------------|-----------|-----------|-----------|  
## 1 | 26 | 1 | 27 |   
## | 96.30% | 3.70% | 15.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 4 | 18 | 22 |   
## | 18.18% | 81.82% | 12.94% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 42 | 42 |   
## | 0.00% | 100.00% | 24.71% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 21 | 21 |   
## | 0.00% | 100.00% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr13"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 45 | 2 | 47 |   
## | 95.74% | 4.26% | 27.65% |   
## -------------|-----------|-----------|-----------|  
## 1 | 31 | 2 | 33 |   
## | 93.94% | 6.06% | 19.41% |   
## -------------|-----------|-----------|-----------|  
## 2 | 9 | 9 | 18 |   
## | 50.00% | 50.00% | 10.59% |   
## -------------|-----------|-----------|-----------|  
## 3 | 1 | 44 | 45 |   
## | 2.22% | 97.78% | 26.47% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 27 | 27 |   
## | 0.00% | 100.00% | 15.88% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr14"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 63 | 3 | 66 |   
## | 95.45% | 4.55% | 38.82% |   
## -------------|-----------|-----------|-----------|  
## 1 | 21 | 2 | 23 |   
## | 91.30% | 8.70% | 13.53% |   
## -------------|-----------|-----------|-----------|  
## 2 | 2 | 19 | 21 |   
## | 9.52% | 90.48% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 38 | 38 |   
## | 0.00% | 100.00% | 22.35% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 22 | 22 |   
## | 0.00% | 100.00% | 12.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr15"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 68 | 1 | 69 |   
## | 98.55% | 1.45% | 40.59% |   
## -------------|-----------|-----------|-----------|  
## 1 | 16 | 4 | 20 |   
## | 80.00% | 20.00% | 11.76% |   
## -------------|-----------|-----------|-----------|  
## 2 | 2 | 11 | 13 |   
## | 15.38% | 84.62% | 7.65% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 51 | 51 |   
## | 0.00% | 100.00% | 30.00% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 17 | 17 |   
## | 0.00% | 100.00% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr16"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 73 | 2 | 75 |   
## | 97.33% | 2.67% | 44.12% |   
## -------------|-----------|-----------|-----------|  
## 1 | 12 | 2 | 14 |   
## | 85.71% | 14.29% | 8.24% |   
## -------------|-----------|-----------|-----------|  
## 2 | 1 | 25 | 26 |   
## | 3.85% | 96.15% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 35 | 35 |   
## | 0.00% | 100.00% | 20.59% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 20 | 20 |   
## | 0.00% | 100.00% | 11.76% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr17"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 72 | 1 | 73 |   
## | 98.63% | 1.37% | 42.94% |   
## -------------|-----------|-----------|-----------|  
## 1 | 13 | 2 | 15 |   
## | 86.67% | 13.33% | 8.82% |   
## -------------|-----------|-----------|-----------|  
## 2 | 1 | 5 | 6 |   
## | 16.67% | 83.33% | 3.53% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 50 | 50 |   
## | 0.00% | 100.00% | 29.41% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 26 | 26 |   
## | 0.00% | 100.00% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr18"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 78 | 1 | 79 |   
## | 98.73% | 1.27% | 46.47% |   
## -------------|-----------|-----------|-----------|  
## 1 | 8 | 2 | 10 |   
## | 80.00% | 20.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 16 | 16 |   
## | 0.00% | 100.00% | 9.41% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 44 | 44 |   
## | 0.00% | 100.00% | 25.88% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 21 | 21 |   
## | 0.00% | 100.00% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr19"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 75 | 2 | 77 |   
## | 97.40% | 2.60% | 45.29% |   
## -------------|-----------|-----------|-----------|  
## 1 | 10 | 1 | 11 |   
## | 90.91% | 9.09% | 6.47% |   
## -------------|-----------|-----------|-----------|  
## 2 | 1 | 4 | 5 |   
## | 20.00% | 80.00% | 2.94% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 50 | 50 |   
## | 0.00% | 100.00% | 29.41% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 27 | 27 |   
## | 0.00% | 100.00% | 15.88% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr20"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 80 | 1 | 81 |   
## | 98.77% | 1.23% | 47.65% |   
## -------------|-----------|-----------|-----------|  
## 1 | 6 | 4 | 10 |   
## | 60.00% | 40.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 21 | 21 |   
## | 0.00% | 100.00% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 36 | 36 |   
## | 0.00% | 100.00% | 21.18% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 22 | 22 |   
## | 0.00% | 100.00% | 12.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr21"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 73 | 5 | 78 |   
## | 93.59% | 6.41% | 45.88% |   
## -------------|-----------|-----------|-----------|  
## 1 | 13 | 4 | 17 |   
## | 76.47% | 23.53% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 17 | 17 |   
## | 0.00% | 100.00% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 47 | 47 |   
## | 0.00% | 100.00% | 27.65% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 11 | 11 |   
## | 0.00% | 100.00% | 6.47% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr22"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 80 | 7 | 87 |   
## | 91.95% | 8.05% | 51.18% |   
## -------------|-----------|-----------|-----------|  
## 1 | 6 | 7 | 13 |   
## | 46.15% | 53.85% | 7.65% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 26 | 26 |   
## | 0.00% | 100.00% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 29 | 29 |   
## | 0.00% | 100.00% | 17.06% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 15 | 15 |   
## | 0.00% | 100.00% | 8.82% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr23"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 81 | 9 | 90 |   
## | 90.00% | 10.00% | 52.94% |   
## -------------|-----------|-----------|-----------|  
## 1 | 3 | 4 | 7 |   
## | 42.86% | 57.14% | 4.12% |   
## -------------|-----------|-----------|-----------|  
## 2 | 2 | 6 | 8 |   
## | 25.00% | 75.00% | 4.71% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 43 | 43 |   
## | 0.00% | 100.00% | 25.29% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 22 | 22 |   
## | 0.00% | 100.00% | 12.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr24"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 68 | 4 | 72 |   
## | 94.44% | 5.56% | 42.35% |   
## -------------|-----------|-----------|-----------|  
## 1 | 13 | 3 | 16 |   
## | 81.25% | 18.75% | 9.41% |   
## -------------|-----------|-----------|-----------|  
## 2 | 5 | 21 | 26 |   
## | 19.23% | 80.77% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 35 | 35 |   
## | 0.00% | 100.00% | 20.59% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 21 | 21 |   
## | 0.00% | 100.00% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr25"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 58 | 5 | 63 |   
## | 92.06% | 7.94% | 37.06% |   
## -------------|-----------|-----------|-----------|  
## 1 | 27 | 1 | 28 |   
## | 96.43% | 3.57% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 2 | 1 | 11 | 12 |   
## | 8.33% | 91.67% | 7.06% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 43 | 43 |   
## | 0.00% | 100.00% | 25.29% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 24 | 24 |   
## | 0.00% | 100.00% | 14.12% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr26"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 69 | 3 | 72 |   
## | 95.83% | 4.17% | 42.35% |   
## -------------|-----------|-----------|-----------|  
## 1 | 17 | 3 | 20 |   
## | 85.00% | 15.00% | 11.76% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 21 | 21 |   
## | 0.00% | 100.00% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 37 | 37 |   
## | 0.00% | 100.00% | 21.76% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 20 | 20 |   
## | 0.00% | 100.00% | 11.76% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr27"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 73 | 4 | 77 |   
## | 94.81% | 5.19% | 45.29% |   
## -------------|-----------|-----------|-----------|  
## 1 | 13 | 5 | 18 |   
## | 72.22% | 27.78% | 10.59% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 17 | 17 |   
## | 0.00% | 100.00% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 46 | 46 |   
## | 0.00% | 100.00% | 27.06% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 12 | 12 |   
## | 0.00% | 100.00% | 7.06% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr28"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 79 | 6 | 85 |   
## | 92.94% | 7.06% | 50.00% |   
## -------------|-----------|-----------|-----------|  
## 1 | 7 | 3 | 10 |   
## | 70.00% | 30.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 30 | 30 |   
## | 0.00% | 100.00% | 17.65% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 28 | 28 |   
## | 0.00% | 100.00% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 17 | 17 |   
## | 0.00% | 100.00% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr29"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 78 | 3 | 81 |   
## | 96.30% | 3.70% | 47.65% |   
## -------------|-----------|-----------|-----------|  
## 1 | 8 | 5 | 13 |   
## | 61.54% | 38.46% | 7.65% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 10 | 10 |   
## | 0.00% | 100.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 43 | 43 |   
## | 0.00% | 100.00% | 25.29% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 23 | 23 |   
## | 0.00% | 100.00% | 13.53% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr30"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 70 | 2 | 72 |   
## | 97.22% | 2.78% | 42.35% |   
## -------------|-----------|-----------|-----------|  
## 1 | 15 | 4 | 19 |   
## | 78.95% | 21.05% | 11.18% |   
## -------------|-----------|-----------|-----------|  
## 2 | 1 | 22 | 23 |   
## | 4.35% | 95.65% | 13.53% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 35 | 35 |   
## | 0.00% | 100.00% | 20.59% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 21 | 21 |   
## | 0.00% | 100.00% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr31"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 42 | 2 | 44 |   
## | 95.45% | 4.55% | 25.88% |   
## -------------|-----------|-----------|-----------|  
## 1 | 22 | 6 | 28 |   
## | 78.57% | 21.43% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 2 | 19 | 2 | 21 |   
## | 90.48% | 9.52% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 17 | 17 |   
## | 0.00% | 100.00% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## 4 | 3 | 57 | 60 |   
## | 5.00% | 95.00% | 35.29% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr32"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 46 | 0 | 46 |   
## | 100.00% | 0.00% | 27.06% |   
## -------------|-----------|-----------|-----------|  
## 1 | 25 | 5 | 30 |   
## | 83.33% | 16.67% | 17.65% |   
## -------------|-----------|-----------|-----------|  
## 2 | 9 | 3 | 12 |   
## | 75.00% | 25.00% | 7.06% |   
## -------------|-----------|-----------|-----------|  
## 3 | 4 | 28 | 32 |   
## | 12.50% | 87.50% | 18.82% |   
## -------------|-----------|-----------|-----------|  
## 4 | 2 | 48 | 50 |   
## | 4.00% | 96.00% | 29.41% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr33"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 68 | 3 | 71 |   
## | 95.77% | 4.23% | 41.76% |   
## -------------|-----------|-----------|-----------|  
## 1 | 14 | 6 | 20 |   
## | 70.00% | 30.00% | 11.76% |   
## -------------|-----------|-----------|-----------|  
## 2 | 2 | 4 | 6 |   
## | 33.33% | 66.67% | 3.53% |   
## -------------|-----------|-----------|-----------|  
## 3 | 1 | 16 | 17 |   
## | 5.88% | 94.12% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## 4 | 1 | 55 | 56 |   
## | 1.79% | 98.21% | 32.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr34"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 48 | 2 | 50 |   
## | 96.00% | 4.00% | 29.41% |   
## -------------|-----------|-----------|-----------|  
## 1 | 31 | 6 | 37 |   
## | 83.78% | 16.22% | 21.76% |   
## -------------|-----------|-----------|-----------|  
## 2 | 5 | 5 | 10 |   
## | 50.00% | 50.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 3 | 1 | 25 | 26 |   
## | 3.85% | 96.15% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 4 | 1 | 46 | 47 |   
## | 2.13% | 97.87% | 27.65% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr35"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 79 | 6 | 85 |   
## | 92.94% | 7.06% | 50.00% |   
## -------------|-----------|-----------|-----------|  
## 1 | 6 | 7 | 13 |   
## | 46.15% | 53.85% | 7.65% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 1 | 1 |   
## | 0.00% | 100.00% | 0.59% |   
## -------------|-----------|-----------|-----------|  
## 3 | 1 | 14 | 15 |   
## | 6.67% | 93.33% | 8.82% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 56 | 56 |   
## | 0.00% | 100.00% | 32.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr36"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 83 | 5 | 88 |   
## | 94.32% | 5.68% | 51.76% |   
## -------------|-----------|-----------|-----------|  
## 1 | 3 | 6 | 9 |   
## | 33.33% | 66.67% | 5.29% |   
## -------------|-----------|-----------|-----------|  
## 2 | 0 | 2 | 2 |   
## | 0.00% | 100.00% | 1.18% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 24 | 24 |   
## | 0.00% | 100.00% | 14.12% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 47 | 47 |   
## | 0.00% | 100.00% | 27.65% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr37"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 48 | 1 | 49 |   
## | 97.96% | 2.04% | 28.82% |   
## -------------|-----------|-----------|-----------|  
## 1 | 28 | 3 | 31 |   
## | 90.32% | 9.68% | 18.24% |   
## -------------|-----------|-----------|-----------|  
## 2 | 7 | 3 | 10 |   
## | 70.00% | 30.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 16 | 16 |   
## | 0.00% | 100.00% | 9.41% |   
## -------------|-----------|-----------|-----------|  
## 4 | 3 | 61 | 64 |   
## | 4.69% | 95.31% | 37.65% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr38"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 61 | 3 | 64 |   
## | 95.31% | 4.69% | 37.65% |   
## -------------|-----------|-----------|-----------|  
## 1 | 21 | 3 | 24 |   
## | 87.50% | 12.50% | 14.12% |   
## -------------|-----------|-----------|-----------|  
## 2 | 3 | 3 | 6 |   
## | 50.00% | 50.00% | 3.53% |   
## -------------|-----------|-----------|-----------|  
## 3 | 1 | 23 | 24 |   
## | 4.17% | 95.83% | 14.12% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 52 | 52 |   
## | 0.00% | 100.00% | 30.59% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr39"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 49 | 1 | 50 |   
## | 98.00% | 2.00% | 29.41% |   
## -------------|-----------|-----------|-----------|  
## 1 | 27 | 1 | 28 |   
## | 96.43% | 3.57% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 2 | 9 | 5 | 14 |   
## | 64.29% | 35.71% | 8.24% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 13 | 13 |   
## | 0.00% | 100.00% | 7.65% |   
## -------------|-----------|-----------|-----------|  
## 4 | 1 | 64 | 65 |   
## | 1.54% | 98.46% | 38.24% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr40"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 71 | 1 | 72 |   
## | 98.61% | 1.39% | 42.35% |   
## -------------|-----------|-----------|-----------|  
## 1 | 12 | 1 | 13 |   
## | 92.31% | 7.69% | 7.65% |   
## -------------|-----------|-----------|-----------|  
## 2 | 3 | 3 | 6 |   
## | 50.00% | 50.00% | 3.53% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 23 | 23 |   
## | 0.00% | 100.00% | 13.53% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 56 | 56 |   
## | 0.00% | 100.00% | 32.94% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr41"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 54 | 1 | 55 |   
## | 98.18% | 1.82% | 32.35% |   
## -------------|-----------|-----------|-----------|  
## 1 | 24 | 2 | 26 |   
## | 92.31% | 7.69% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 2 | 7 | 7 | 14 |   
## | 50.00% | 50.00% | 8.24% |   
## -------------|-----------|-----------|-----------|  
## 3 | 1 | 14 | 15 |   
## | 6.67% | 93.33% | 8.82% |   
## -------------|-----------|-----------|-----------|  
## 4 | 0 | 60 | 60 |   
## | 0.00% | 100.00% | 35.29% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr42"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 43 | 1 | 44 |   
## | 97.73% | 2.27% | 25.88% |   
## -------------|-----------|-----------|-----------|  
## 1 | 13 | 4 | 17 |   
## | 76.47% | 23.53% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## 2 | 20 | 6 | 26 |   
## | 76.92% | 23.08% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 6 | 28 | 34 |   
## | 17.65% | 82.35% | 20.00% |   
## -------------|-----------|-----------|-----------|  
## 4 | 4 | 45 | 49 |   
## | 8.16% | 91.84% | 28.82% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr43"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 14 | 2 | 16 |   
## | 87.50% | 12.50% | 9.41% |   
## -------------|-----------|-----------|-----------|  
## 1 | 16 | 3 | 19 |   
## | 84.21% | 15.79% | 11.18% |   
## -------------|-----------|-----------|-----------|  
## 2 | 29 | 4 | 33 |   
## | 87.88% | 12.12% | 19.41% |   
## -------------|-----------|-----------|-----------|  
## 3 | 14 | 19 | 33 |   
## | 42.42% | 57.58% | 19.41% |   
## -------------|-----------|-----------|-----------|  
## 4 | 13 | 56 | 69 |   
## | 18.84% | 81.16% | 40.59% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr44"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 58 | 2 | 60 |   
## | 96.67% | 3.33% | 35.29% |   
## -------------|-----------|-----------|-----------|  
## 1 | 14 | 2 | 16 |   
## | 87.50% | 12.50% | 9.41% |   
## -------------|-----------|-----------|-----------|  
## 2 | 12 | 5 | 17 |   
## | 70.59% | 29.41% | 10.00% |   
## -------------|-----------|-----------|-----------|  
## 3 | 0 | 28 | 28 |   
## | 0.00% | 100.00% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 4 | 2 | 47 | 49 |   
## | 4.08% | 95.92% | 28.82% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr45"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 24 | 4 | 28 |   
## | 85.71% | 14.29% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 1 | 22 | 2 | 24 |   
## | 91.67% | 8.33% | 14.12% |   
## -------------|-----------|-----------|-----------|  
## 2 | 14 | 7 | 21 |   
## | 66.67% | 33.33% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 12 | 24 | 36 |   
## | 33.33% | 66.67% | 21.18% |   
## -------------|-----------|-----------|-----------|  
## 4 | 14 | 47 | 61 |   
## | 22.95% | 77.05% | 35.88% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr46"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 18 | 4 | 22 |   
## | 81.82% | 18.18% | 12.94% |   
## -------------|-----------|-----------|-----------|  
## 1 | 15 | 3 | 18 |   
## | 83.33% | 16.67% | 10.59% |   
## -------------|-----------|-----------|-----------|  
## 2 | 17 | 11 | 28 |   
## | 60.71% | 39.29% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 3 | 25 | 23 | 48 |   
## | 52.08% | 47.92% | 28.24% |   
## -------------|-----------|-----------|-----------|  
## 4 | 11 | 43 | 54 |   
## | 20.37% | 79.63% | 31.76% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr47"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 30 | 3 | 33 |   
## | 90.91% | 9.09% | 19.41% |   
## -------------|-----------|-----------|-----------|  
## 1 | 28 | 6 | 34 |   
## | 82.35% | 17.65% | 20.00% |   
## -------------|-----------|-----------|-----------|  
## 2 | 13 | 7 | 20 |   
## | 65.00% | 35.00% | 11.76% |   
## -------------|-----------|-----------|-----------|  
## 3 | 7 | 13 | 20 |   
## | 35.00% | 65.00% | 11.76% |   
## -------------|-----------|-----------|-----------|  
## 4 | 8 | 55 | 63 |   
## | 12.70% | 87.30% | 37.06% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr48"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 10 | 0 | 10 |   
## | 100.00% | 0.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 1 | 8 | 2 | 10 |   
## | 80.00% | 20.00% | 5.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 42 | 5 | 47 |   
## | 89.36% | 10.64% | 27.65% |   
## -------------|-----------|-----------|-----------|  
## 3 | 20 | 30 | 50 |   
## | 40.00% | 60.00% | 29.41% |   
## -------------|-----------|-----------|-----------|  
## 4 | 6 | 47 | 53 |   
## | 11.32% | 88.68% | 31.18% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr49"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 28 | 0 | 28 |   
## | 100.00% | 0.00% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 1 | 26 | 2 | 28 |   
## | 92.86% | 7.14% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 2 | 17 | 9 | 26 |   
## | 65.38% | 34.62% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 10 | 17 | 27 |   
## | 37.04% | 62.96% | 15.88% |   
## -------------|-----------|-----------|-----------|  
## 4 | 5 | 56 | 61 |   
## | 8.20% | 91.80% | 35.88% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr50"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 19 | 0 | 19 |   
## | 100.00% | 0.00% | 11.18% |   
## -------------|-----------|-----------|-----------|  
## 1 | 27 | 4 | 31 |   
## | 87.10% | 12.90% | 18.24% |   
## -------------|-----------|-----------|-----------|  
## 2 | 31 | 5 | 36 |   
## | 86.11% | 13.89% | 21.18% |   
## -------------|-----------|-----------|-----------|  
## 3 | 6 | 20 | 26 |   
## | 23.08% | 76.92% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 4 | 3 | 55 | 58 |   
## | 5.17% | 94.83% | 34.12% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr51"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 12 | 0 | 12 |   
## | 100.00% | 0.00% | 7.06% |   
## -------------|-----------|-----------|-----------|  
## 1 | 24 | 5 | 29 |   
## | 82.76% | 17.24% | 17.06% |   
## -------------|-----------|-----------|-----------|  
## 2 | 38 | 5 | 43 |   
## | 88.37% | 11.63% | 25.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 9 | 29 | 38 |   
## | 23.68% | 76.32% | 22.35% |   
## -------------|-----------|-----------|-----------|  
## 4 | 3 | 45 | 48 |   
## | 6.25% | 93.75% | 28.24% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr52"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 23 | 0 | 23 |   
## | 100.00% | 0.00% | 13.53% |   
## -------------|-----------|-----------|-----------|  
## 1 | 23 | 5 | 28 |   
## | 82.14% | 17.86% | 16.47% |   
## -------------|-----------|-----------|-----------|  
## 2 | 17 | 6 | 23 |   
## | 73.91% | 26.09% | 13.53% |   
## -------------|-----------|-----------|-----------|  
## 3 | 14 | 16 | 30 |   
## | 46.67% | 53.33% | 17.65% |   
## -------------|-----------|-----------|-----------|  
## 4 | 9 | 57 | 66 |   
## | 13.64% | 86.36% | 38.82% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr53"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 31 | 0 | 31 |   
## | 100.00% | 0.00% | 18.24% |   
## -------------|-----------|-----------|-----------|  
## 1 | 25 | 6 | 31 |   
## | 80.65% | 19.35% | 18.24% |   
## -------------|-----------|-----------|-----------|  
## 2 | 18 | 8 | 26 |   
## | 69.23% | 30.77% | 15.29% |   
## -------------|-----------|-----------|-----------|  
## 3 | 7 | 23 | 30 |   
## | 23.33% | 76.67% | 17.65% |   
## -------------|-----------|-----------|-----------|  
## 4 | 5 | 47 | 52 |   
## | 9.62% | 90.38% | 30.59% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##   
## [1] "Atr54"  
##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Row Percent |  
## |-------------------------|  
##   
## Total Observations in Table: 170   
##   
## | SLD$Class   
## SLD[, i] | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 46 | 4 | 50 |   
## | 92.00% | 8.00% | 29.41% |   
## -------------|-----------|-----------|-----------|  
## 1 | 24 | 3 | 27 |   
## | 88.89% | 11.11% | 15.88% |   
## -------------|-----------|-----------|-----------|  
## 2 | 14 | 7 | 21 |   
## | 66.67% | 33.33% | 12.35% |   
## -------------|-----------|-----------|-----------|  
## 3 | 1 | 14 | 15 |   
## | 6.67% | 93.33% | 8.82% |   
## -------------|-----------|-----------|-----------|  
## 4 | 1 | 56 | 57 |   
## | 1.75% | 98.25% | 33.53% |   
## -------------|-----------|-----------|-----------|  
## Column Total | 86 | 84 | 170 |   
## -------------|-----------|-----------|-----------|  
##   
##

## Density plots

for (i in 1:54){  
print(ggplot(data = SLD,  
 aes(x = SLD[,i],  
 fill= Class))+  
 ylab("count")+  
 xlab(names(SLD)[i])+  
 geom\_density(alpha=0.3))  
}

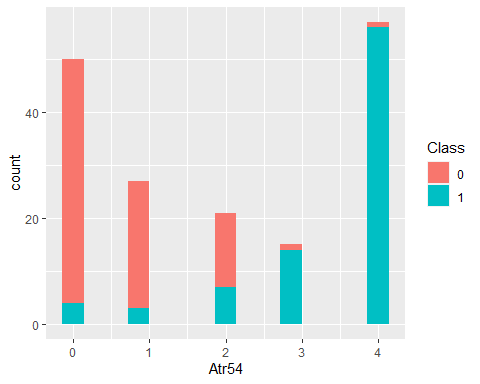
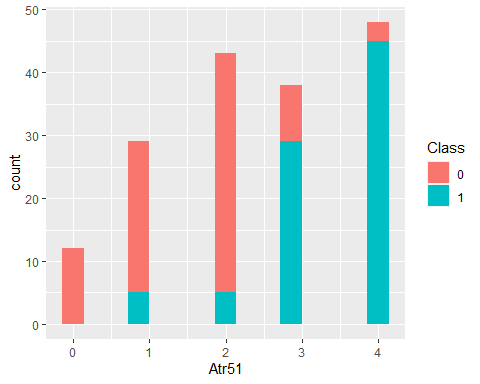
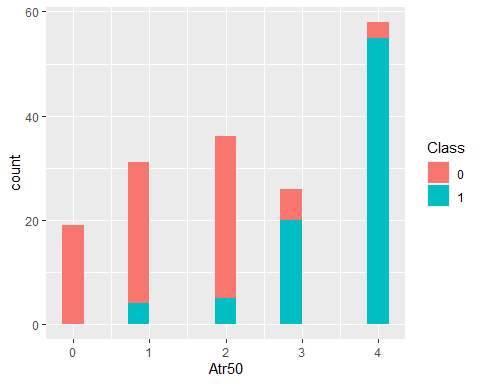
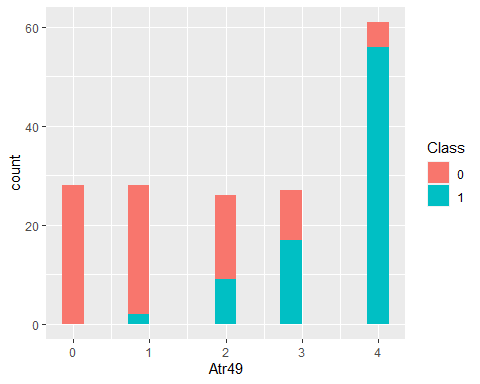
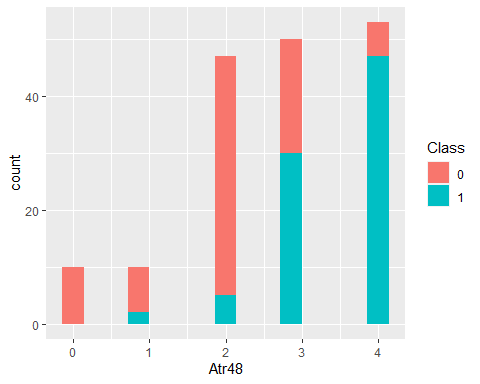
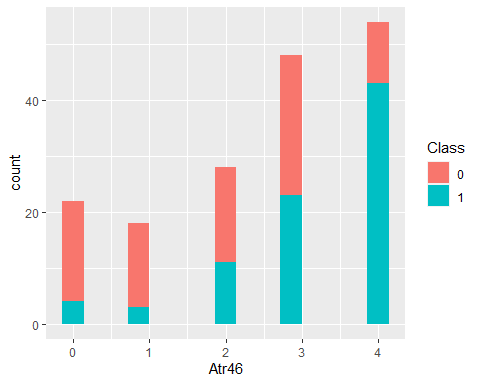
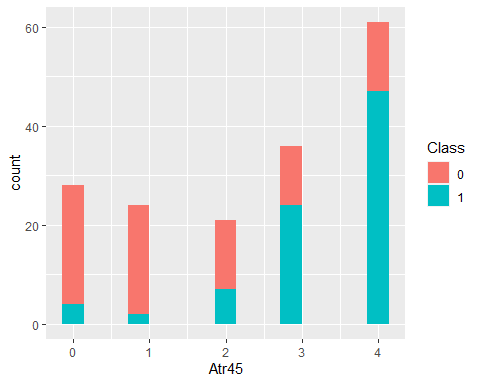
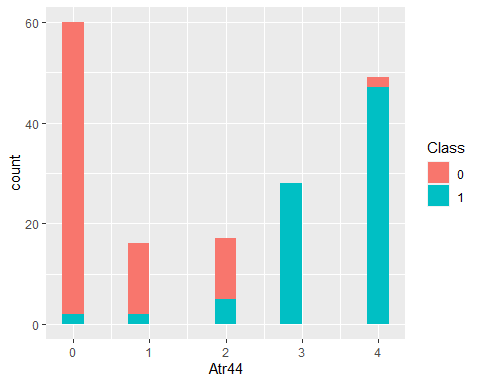
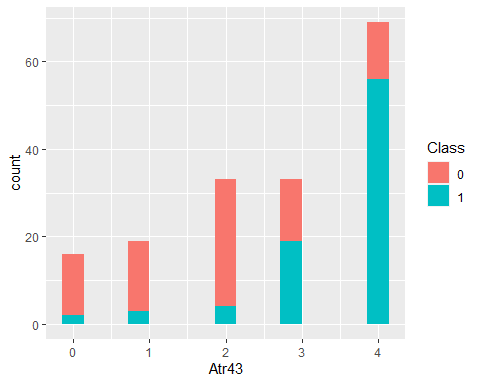
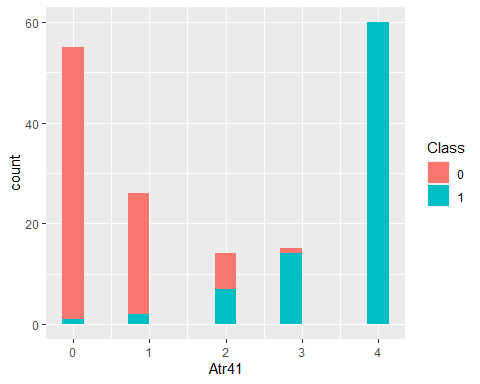
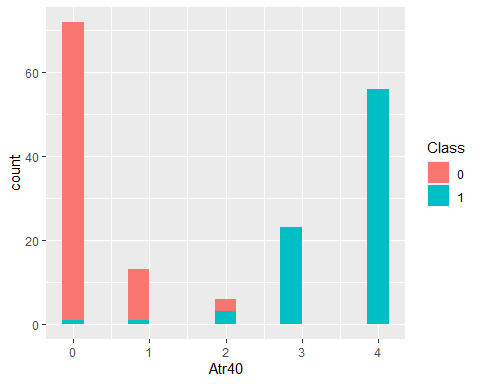
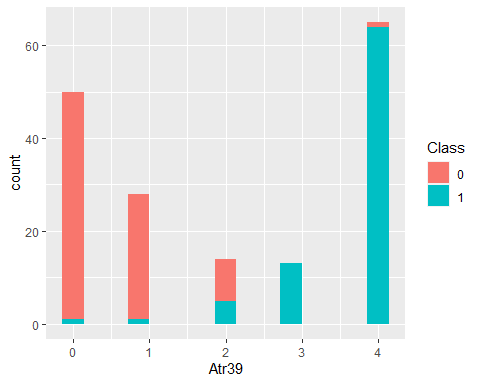
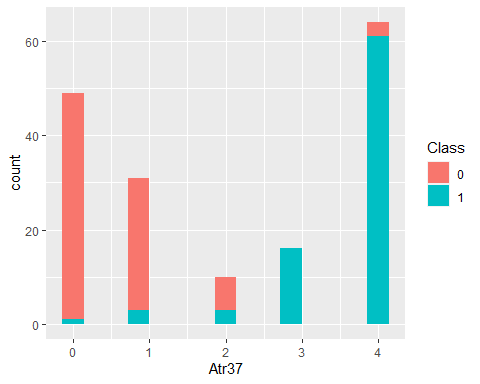
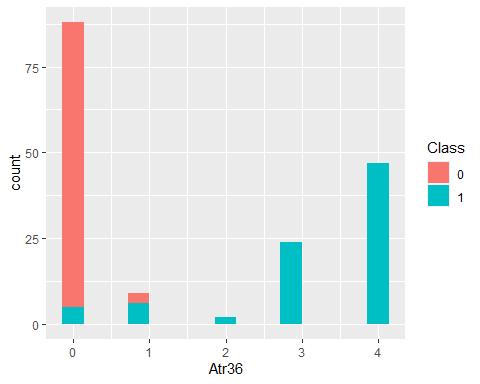
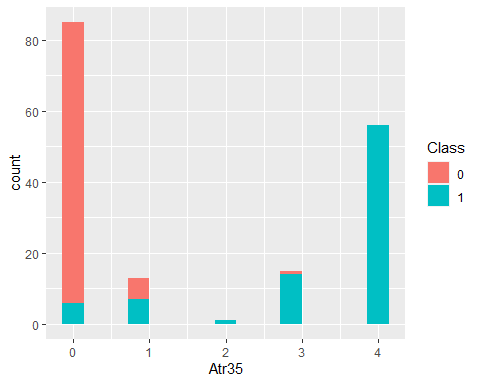
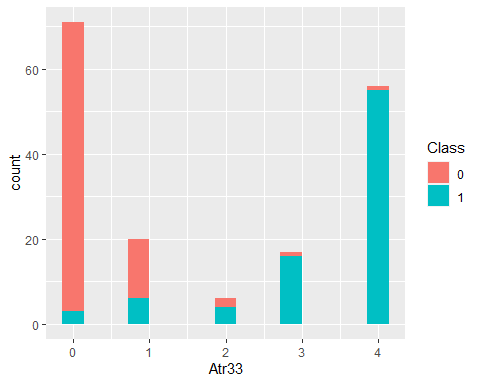
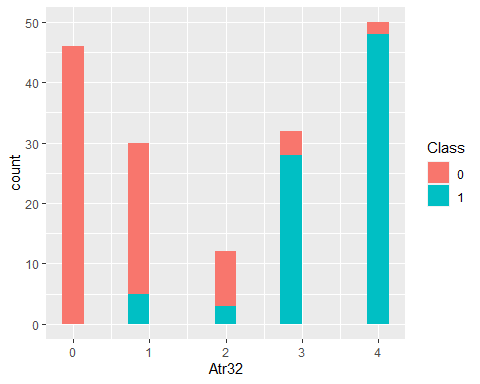
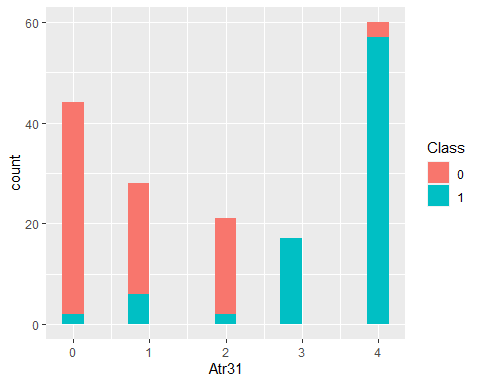
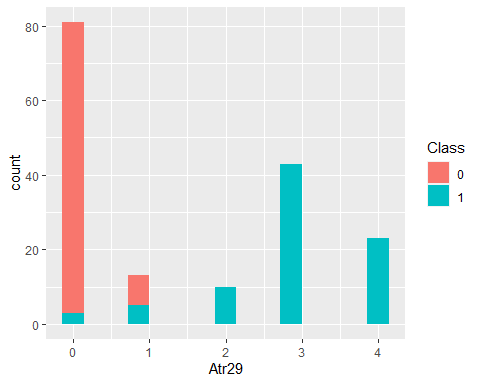
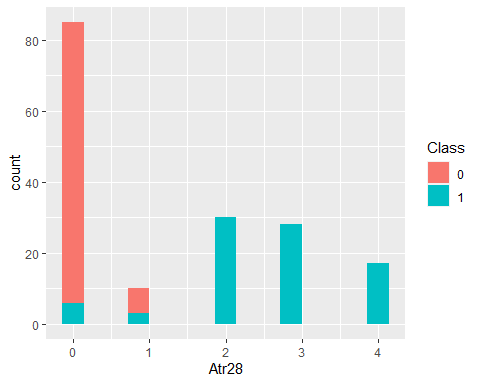
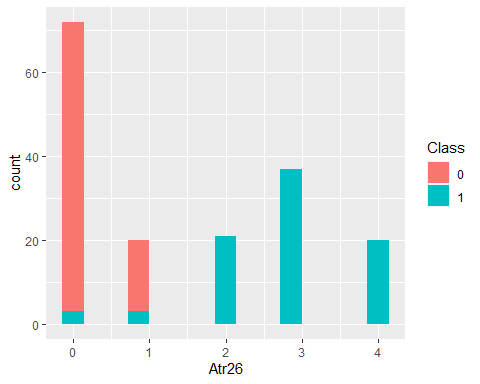
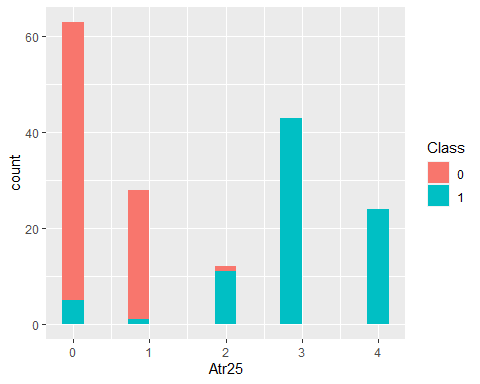
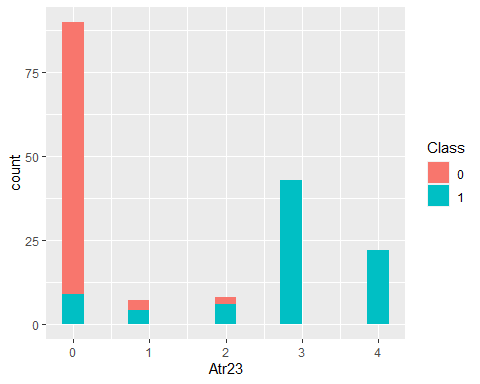
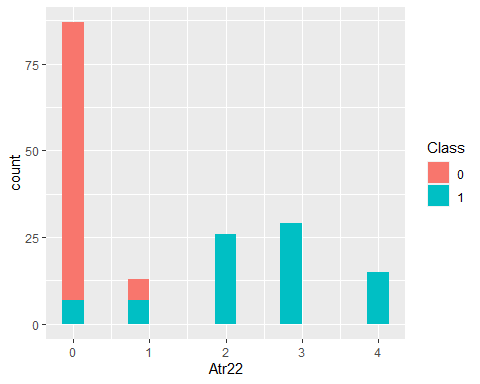
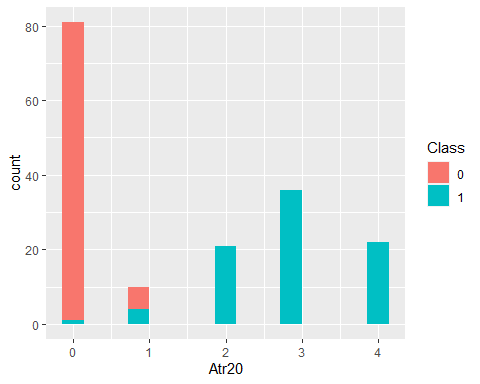
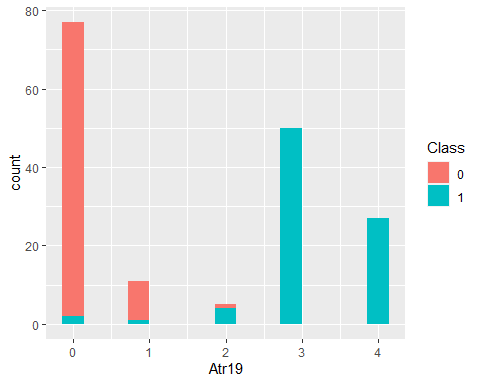
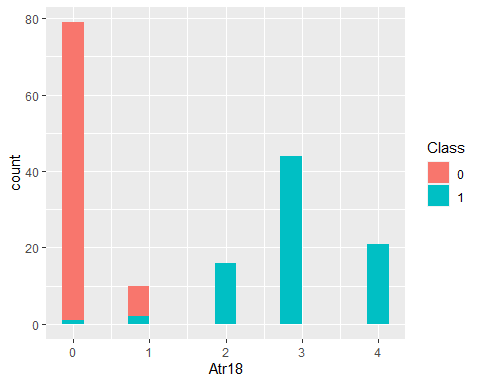
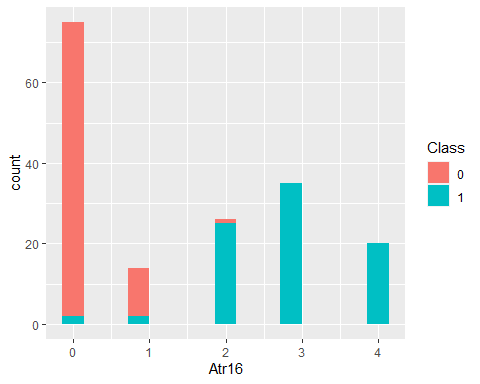
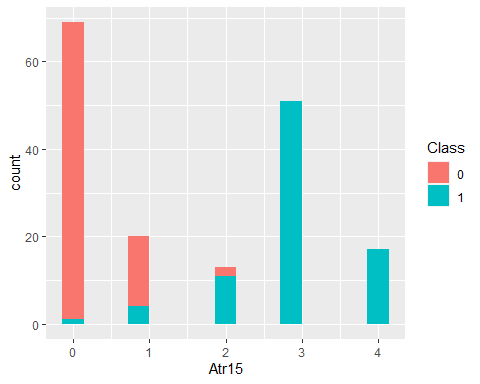
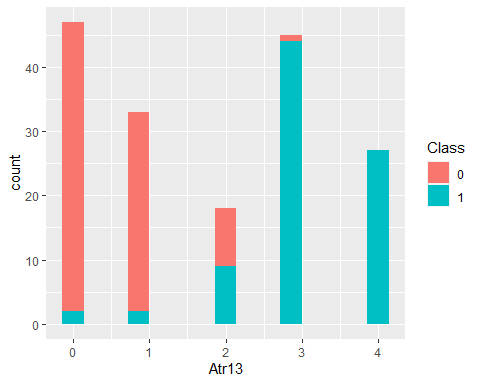
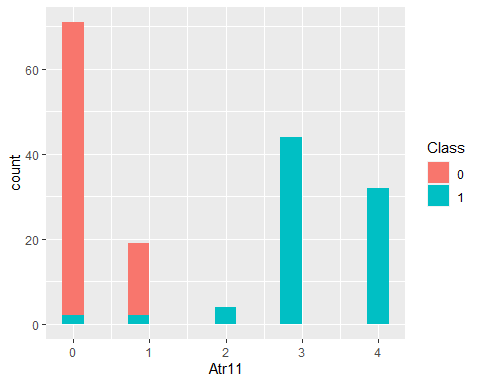
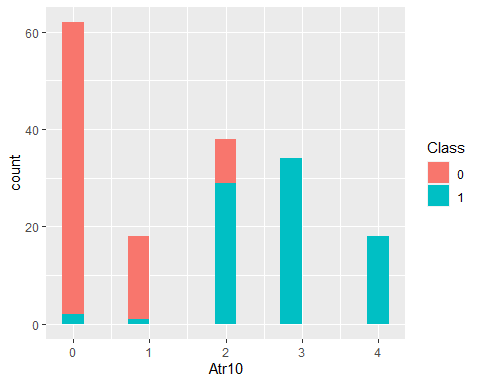
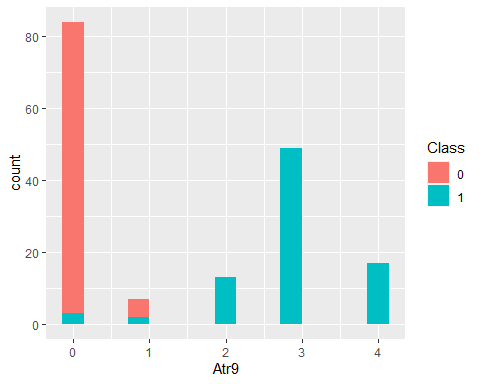
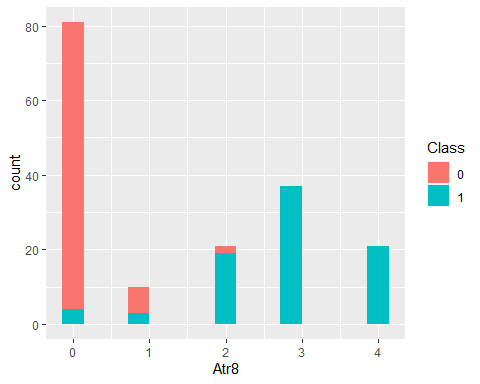
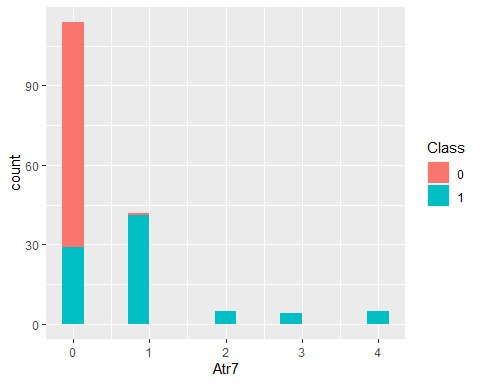
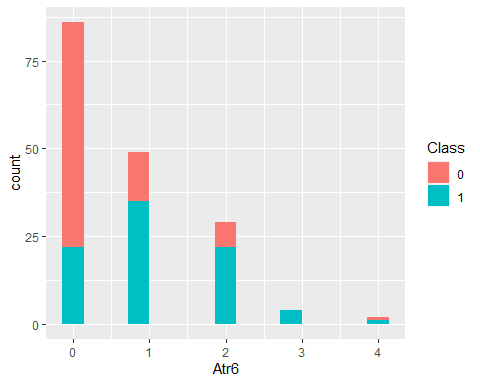
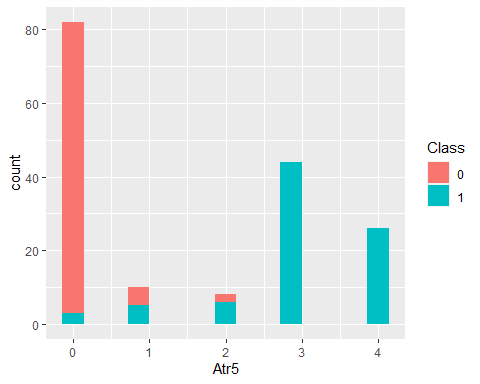
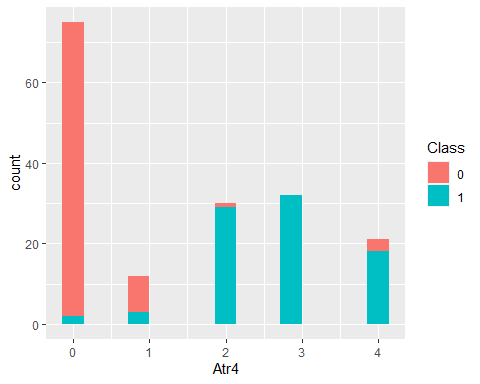
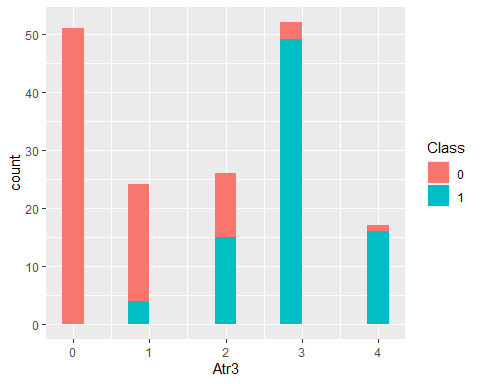
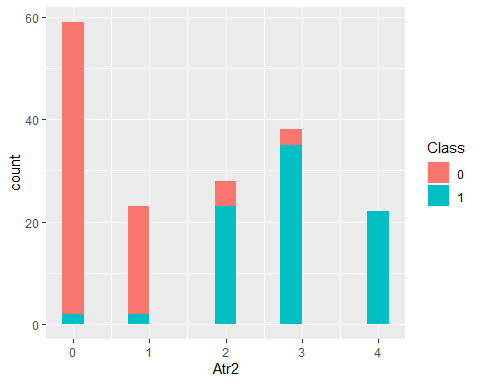
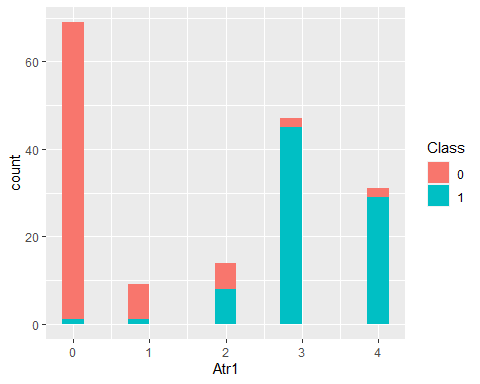


ggplot(data = SLD,  
 aes(x = SLD[,15]))+  
 ylab("count")+  
 xlab(names(SLD)[15])+  
 geom\_density(alpha=0.3)



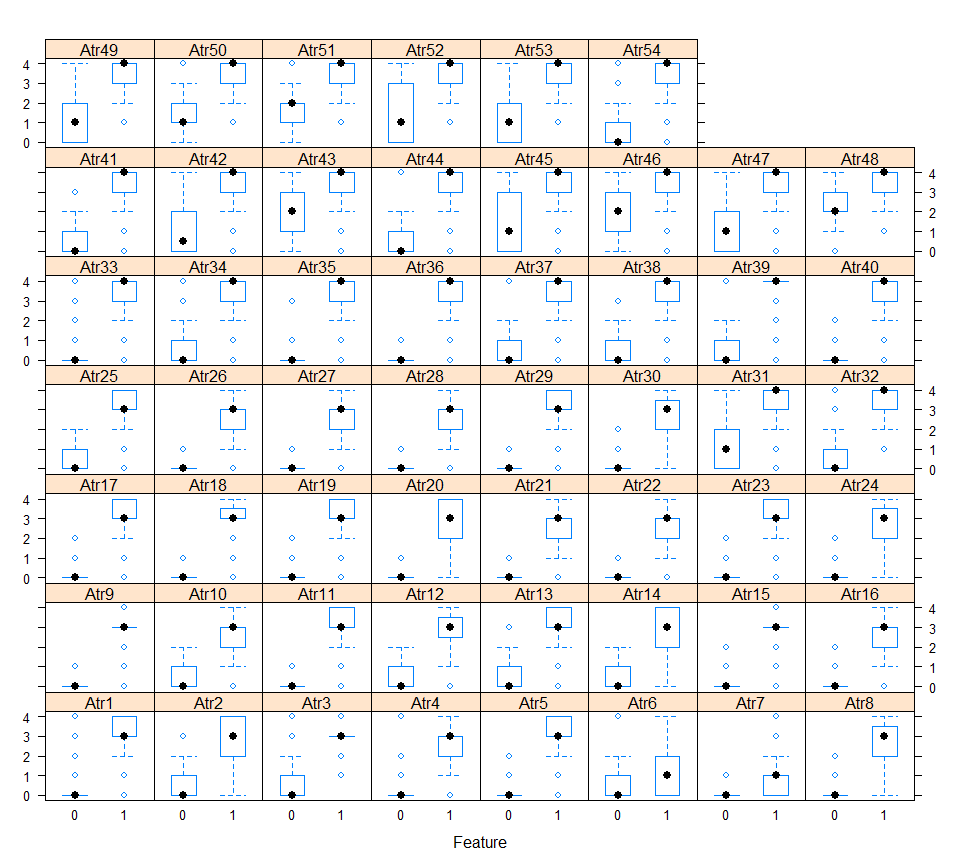
## Bar plots

for (i in 1:54){  
print(ggplot(data = SLD,  
 aes(x = SLD[,i],  
 fill= Class))+  
 ylab("count")+  
 xlab(names(SLD)[i])+  
 geom\_histogram(bins = 15))  
}



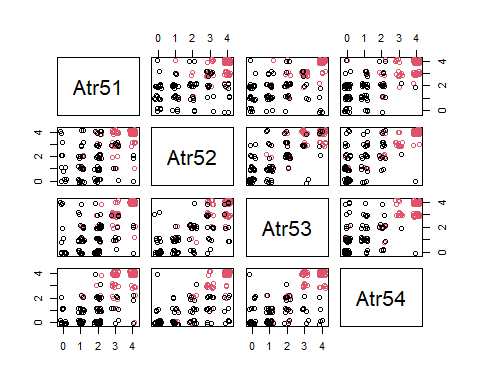
## Boxplots

featurePlot(x=SLD[,1:54],y=SLD[,55],plot = "box")



## Scaterplots with jitter

jittered <- sapply(SLD[,1:54],jitter)  
pairs(jittered[,51:54],col=SLD$Class)

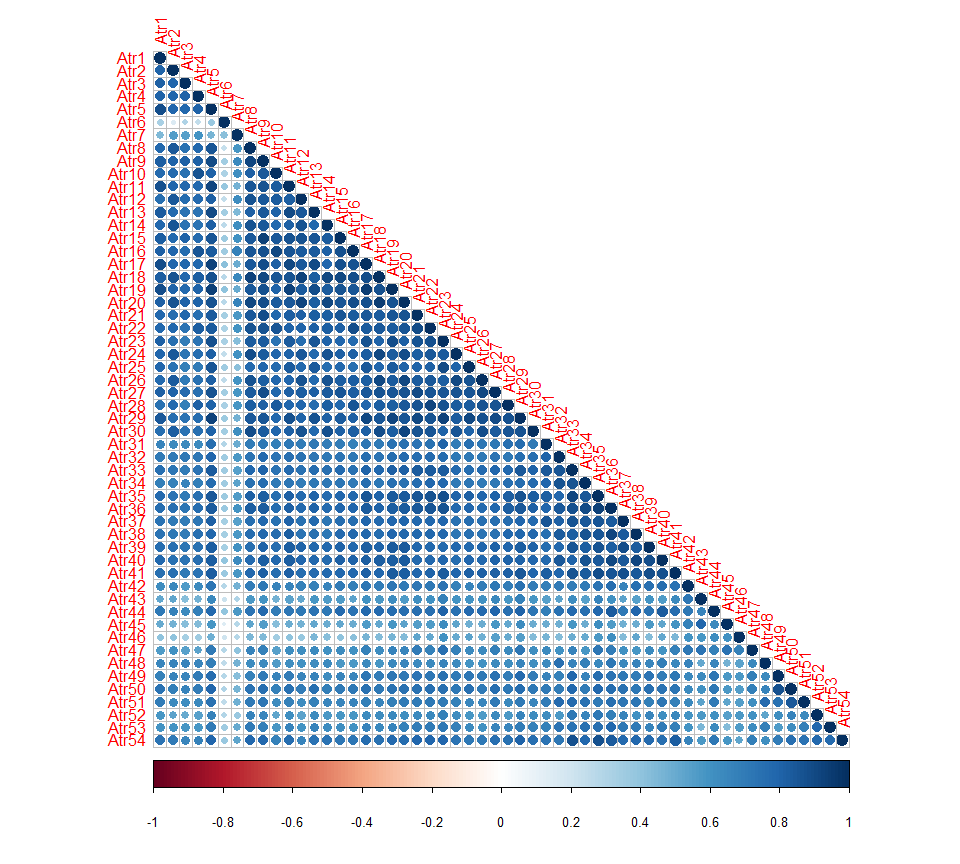


# Data spliting

seed <- 1821  
set.seed(seed)  
   
  
idx <-sample(nrow(SLD),   
 nrow(SLD)\*0.75,  
 replace = F)  
  
daTrain <- SLD[idx,]  
daTest <- SLD[-idx,]  
  
#   
# idx <- sample(seq(1, 2),   
# size = nrow(SLD),   
# replace = TRUE,   
# prob = c(.75, .25))  
  
# daTrain <- SLD[idx == 1,]  
# daTest <- SLD[idx == 2,]

# Correlated Variables

## create correlation matrix  
predictor\_variables = setdiff(colnames(daTrain), "Class")  
 correlation\_matrix = cor(SLD[ , predictor\_variables],  
 method ="spearman")  
  
## plot correlation  
corrplot(correlation\_matrix,   
 # method = "ellipse",   
 type = "lower")



# PCA analysis

## create pca object using train dataset  
pca = prcomp(daTrain[ ,predictor\_variables], scale=TRUE, center=TRUE)  
  
## view pca summary  
summary(pca)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6 PC7  
## Standard deviation 6.3752 1.48216 1.22581 1.09705 0.95419 0.85309 0.84251  
## Proportion of Variance 0.7527 0.04068 0.02783 0.02229 0.01686 0.01348 0.01314  
## Cumulative Proportion 0.7527 0.79334 0.82116 0.84345 0.86031 0.87379 0.88693  
## PC8 PC9 PC10 PC11 PC12 PC13 PC14  
## Standard deviation 0.78261 0.75954 0.69365 0.64983 0.59879 0.5594 0.54656  
## Proportion of Variance 0.01134 0.01068 0.00891 0.00782 0.00664 0.0058 0.00553  
## Cumulative Proportion 0.89827 0.90896 0.91787 0.92569 0.93233 0.9381 0.94365  
## PC15 PC16 PC17 PC18 PC19 PC20 PC21  
## Standard deviation 0.52226 0.49616 0.47536 0.44010 0.42667 0.40793 0.40453  
## Proportion of Variance 0.00505 0.00456 0.00418 0.00359 0.00337 0.00308 0.00303  
## Cumulative Proportion 0.94871 0.95326 0.95745 0.96104 0.96441 0.96749 0.97052  
## PC22 PC23 PC24 PC25 PC26 PC27 PC28  
## Standard deviation 0.38382 0.35817 0.34199 0.33114 0.32216 0.30428 0.29018  
## Proportion of Variance 0.00273 0.00238 0.00217 0.00203 0.00192 0.00171 0.00156  
## Cumulative Proportion 0.97325 0.97562 0.97779 0.97982 0.98174 0.98346 0.98502  
## PC29 PC30 PC31 PC32 PC33 PC34 PC35  
## Standard deviation 0.2850 0.26627 0.26372 0.25566 0.24683 0.22700 0.21814  
## Proportion of Variance 0.0015 0.00131 0.00129 0.00121 0.00113 0.00095 0.00088  
## Cumulative Proportion 0.9865 0.98783 0.98912 0.99033 0.99146 0.99241 0.99329  
## PC36 PC37 PC38 PC39 PC40 PC41 PC42  
## Standard deviation 0.20450 0.19781 0.18774 0.18282 0.17415 0.15791 0.15044  
## Proportion of Variance 0.00077 0.00072 0.00065 0.00062 0.00056 0.00046 0.00042  
## Cumulative Proportion 0.99407 0.99479 0.99545 0.99607 0.99663 0.99709 0.99751  
## PC43 PC44 PC45 PC46 PC47 PC48 PC49  
## Standard deviation 0.14508 0.13411 0.13099 0.12508 0.11135 0.10733 0.10145  
## Proportion of Variance 0.00039 0.00033 0.00032 0.00029 0.00023 0.00021 0.00019  
## Cumulative Proportion 0.99790 0.99823 0.99855 0.99884 0.99907 0.99928 0.99947  
## PC50 PC51 PC52 PC53 PC54  
## Standard deviation 0.08837 0.08559 0.07244 0.06717 0.06034  
## Proportion of Variance 0.00014 0.00014 0.00010 0.00008 0.00007  
## Cumulative Proportion 0.99962 0.99975 0.99985 0.99993 1.00000

## Save PCA components in a new train set

## save PC components for train dataset  
train.pc = as.data.frame(pca$x)  
train.pc$Class = daTrain$Class  
head(train.pc)

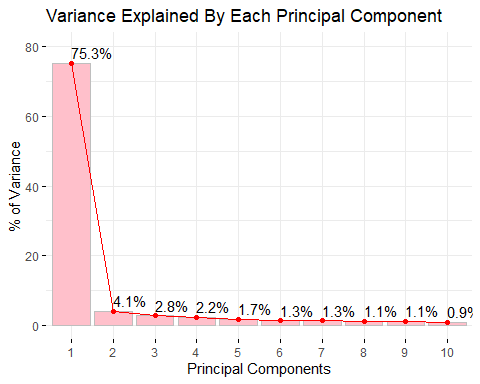
## PC1 PC2 PC3 PC4 PC5 PC6  
## 46 -7.027907 -0.9315128 0.6830216 0.1824142 -0.8168369 0.45327064  
## 161 5.237180 0.4032814 0.5450589 -0.3580583 0.1999248 -0.73095098  
## 62 -6.219161 -1.1503136 1.3680131 1.7973590 -0.2497481 0.54901732  
## 167 5.362912 -2.0384566 0.7718106 0.2333332 -0.7724820 0.45045310  
## 14 -8.806375 -0.2336933 -0.4877372 -1.5832140 -1.1889899 0.22871909  
## 49 -6.982932 -1.6624705 1.3940544 1.7456398 1.2736935 -0.06895638  
## PC7 PC8 PC9 PC10 PC11 PC12  
## 46 -0.2866837 0.2239027 -0.4926415 0.35082619 -0.4284974 -0.0006151886  
## 161 0.4705434 -1.3799090 -0.7289693 0.06706370 0.8956756 -0.2359041330  
## 62 0.2496765 -0.1246046 -0.9799316 0.18449857 -0.2760797 0.0440601738  
## 167 -0.4794212 0.6663166 -0.6509587 -0.26624885 0.3118439 -0.9008465058  
## 14 -0.4880462 0.0217516 -1.1661549 0.27371977 -0.4137592 0.3827950615  
## 49 0.7097358 -0.4162127 0.8273576 -0.07259166 0.3337429 -0.2013262222  
## PC13 PC14 PC15 PC16 PC17 PC18  
## 46 0.26616260 0.3253395 -0.7875684 0.40898507 0.32960378 -0.093466594  
## 161 0.32381403 -0.6659314 -0.3468525 -0.74383353 0.48210078 -0.005602887  
## 62 0.42562252 0.4193250 -0.4416028 0.32719692 0.39050681 0.045348266  
## 167 -0.11681482 -0.4456221 -0.1017010 0.27693957 0.24747221 0.432545805  
## 14 -0.00521788 0.1701332 -0.4565992 0.05670445 -0.02563222 0.130555524  
## 49 0.13865233 -0.3874551 0.3923291 0.22225039 0.14245444 0.059267373  
## PC19 PC20 PC21 PC22 PC23 PC24  
## 46 -0.15443467 0.04387679 0.1427649 -0.24013422 -0.252120268 -0.11880072  
## 161 0.59005095 0.92220229 0.3038433 -0.30385636 -0.215601206 0.24970543  
## 62 -0.03825611 0.10407967 0.1716721 -0.27331671 -0.138544461 -0.12756898  
## 167 -0.42485894 0.04641175 -0.1540401 -0.33920390 0.000343586 -0.14027515  
## 14 0.04916385 -0.08468241 -0.1154625 -0.18352815 -0.260307839 -0.01021067  
## 49 -0.06464054 0.05173198 0.1798850 0.07660692 0.046949073 -0.03073604  
## PC25 PC26 PC27 PC28 PC29 PC30  
## 46 0.096012918 0.203849146 -0.086570370 -0.039266955 -0.06076505 -0.08956715  
## 161 0.321752404 0.572788108 -0.074136510 -0.004550195 -1.07198011 -0.26843741  
## 62 0.111394747 0.147419937 -0.024505676 -0.070217933 0.04172334 -0.08572846  
## 167 -0.030263687 -0.335266236 0.237515183 0.042569571 -0.03295366 0.33210226  
## 14 0.093224355 0.067164640 -0.021856141 0.059428794 -0.07815301 0.03352975  
## 49 0.003999998 0.006163418 0.009351068 -0.075194691 0.08175904 -0.02717519  
## PC31 PC32 PC33 PC34 PC35 PC36  
## 46 -0.10534284 0.044944671 -0.005627192 -0.02569782 -0.03605582 0.008634286  
## 161 0.06902846 0.011807975 0.423854000 -0.06891526 0.13632605 0.192056854  
## 62 -0.07716802 -0.001612189 0.025434254 -0.03855740 -0.05471501 0.026359737  
## 167 -0.00248621 0.038839977 -0.053137776 -0.09024609 -0.10238756 0.224399250  
## 14 -0.06933430 0.031584664 -0.062635331 -0.02406706 0.02067180 0.006911335  
## 49 0.01746433 -0.061243743 -0.052822226 -0.01721126 0.01205941 -0.003920824  
## PC37 PC38 PC39 PC40 PC41 PC42  
## 46 -0.0092935955 -0.05136161 -0.034662927 0.065744161 -0.008389571 -0.01506495  
## 161 -0.5346505006 0.40318016 -0.499205032 0.190136381 0.196766704 -0.33576637  
## 62 -0.0264723175 -0.04752941 -0.038637854 0.075200181 -0.016559948 0.01475172  
## 167 0.1502509601 0.33843902 -0.150433181 0.170463506 0.045856306 0.08052771  
## 14 0.0005474739 0.02086392 0.042909250 0.009746427 0.009428965 -0.03194269  
## 49 -0.0174066720 0.01272855 -0.004791539 0.002487481 0.010223788 0.01845978  
## PC43 PC44 PC45 PC46 PC47  
## 46 0.058160710 -0.036069394 -0.01100156 -0.094138259 -0.025752972  
## 161 0.104265112 0.128250579 0.11551512 -0.029643382 -0.067525774  
## 62 0.034723274 -0.041192378 -0.03470800 -0.133723297 -0.000788924  
## 167 -0.020710441 0.061820247 0.04894593 0.061327181 0.139326299  
## 14 0.004397782 0.006859060 0.04558180 0.008286186 -0.030328671  
## 49 -0.010195516 -0.003405889 -0.01012149 -0.034920525 0.027721401  
## PC48 PC49 PC50 PC51 PC52  
## 46 8.374575e-05 -0.020194756 -0.001412414 0.006323236 -0.001875460  
## 161 -1.831508e-01 0.201935646 -0.178042470 0.107837865 0.106974812  
## 62 1.949348e-02 -0.029513406 0.022624092 -0.003887316 0.007061537  
## 167 -6.163793e-02 0.081176127 -0.018726644 0.042726935 0.026411871  
## 14 -9.881355e-03 0.008693074 -0.007969232 -0.004675058 0.005621510  
## 49 -4.111061e-03 0.001551605 0.004883071 -0.009754961 0.009115978  
## PC53 PC54 Class  
## 46 -0.008014864 -0.004578361 1  
## 161 0.105882723 0.062263511 0  
## 62 0.003468133 -0.005446668 1  
## 167 0.003415148 -0.050431818 0  
## 14 -0.006310924 0.008802237 1  
## 49 0.018969258 0.013092299 1

## Save PCA components for the test set too

## save PC components for test dataset  
test.pc = predict(pca, newdata = daTest)  
test.pc = as.data.frame(test.pc)  
test.pc$Class = daTest$Class

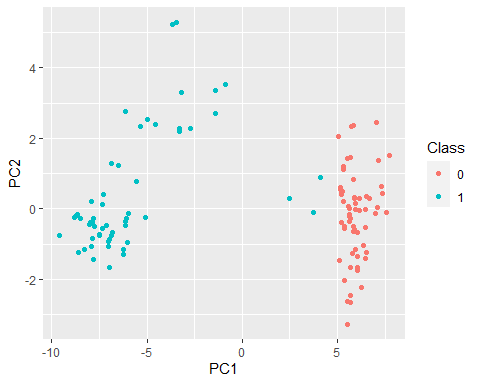
## Visualize Variance Explained by Principal Components Analysis

fviz\_eig(pca,   
 addlabels = TRUE,   
 ylim = c(0,80),  
 geom = c("bar", "line"),  
 barfill = "pink",   
 barcolor = "grey",  
 linecolor = "red",  
 ncp = 10) +  
labs(title = "Variance Explained By Each Principal Component",  
 x = "Principal Components",   
 y = "% of Variance")

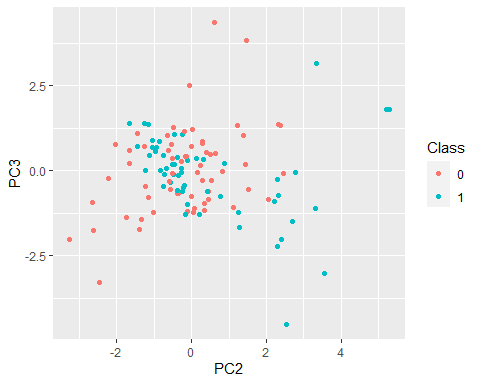


## Effectiveness of First Principal Components in Separating Out Data Points

## first and second principal components  
ggplot(train.pc) +   
geom\_point(aes(x = PC1,   
 y = PC2,   
 color = Class))



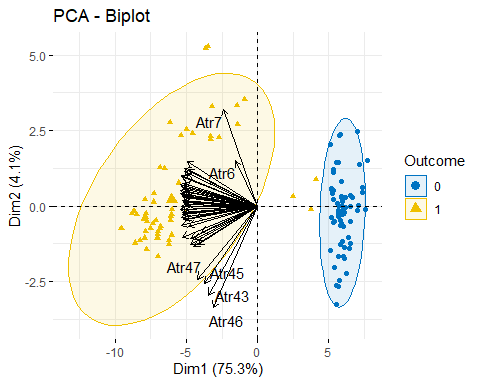
## second and third principal components  
ggplot(train.pc) +   
geom\_point(aes(x = PC2,  
 y = PC3,  
 color = Class))



## Biplot of Principal Components

fviz\_pca\_biplot(pca,   
 col.ind = daTrain$Class,   
 col = "black",  
 palette = "jco",   
 geom = "point",  
 repel = TRUE,  
 legend.title = "Outcome",  
 addEllipses = TRUE)

## Warning: ggrepel: 48 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps



# Logistic regression with PCA1 component as predictor

## create model  
fit.glm.PCA1 = glm(formula = Class ~ PC1,   
 data = train.pc,   
 family = "binomial")

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Predictions on train set

## create prediction probabilities (on train dataset)  
train\_probs = predict(fit.glm.PCA1, type="response")  
  
## create predictions (on train dataset)  
train\_preds = as.factor(ifelse(train\_probs > 0.5, "1", "0"))  
  
## evaluate performance (on train dataset)  
confusionMatrix(train\_preds, train.pc$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 64 0  
## 1 0 63  
##   
## Accuracy : 1   
## 95% CI : (0.9714, 1)  
## No Information Rate : 0.5039   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5039   
## Detection Rate : 0.5039   
## Detection Prevalence : 0.5039   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : 0   
##

## Predictions on test set

## create prediction probabilities (on test dataset)  
test\_probs = predict(fit.glm.PCA1,   
 type = "response",   
 newdata = test.pc)  
  
## create predictions (on test dataset)  
test\_preds = as.factor(ifelse(test\_probs > 0.5, "1", "0"))  
  
## evaluate performance (on test dataset)  
confusionMatrix(test\_preds, test.pc$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 22 0  
## 1 0 21  
##   
## Accuracy : 1   
## 95% CI : (0.9178, 1)  
## No Information Rate : 0.5116   
## P-Value [Acc > NIR] : 3.055e-13   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5116   
## Detection Rate : 0.5116   
## Detection Prevalence : 0.5116   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : 0   
##

## Summary of glm model with PCA1 as predictor

summary(fit.glm.PCA1)

##   
## Call:  
## glm(formula = Class ~ PC1, family = "binomial", data = train.pc)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.476e-04 -2.100e-08 -2.100e-08 2.100e-08 1.345e-04   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 177.75 40497.05 0.004 0.996  
## PC1 -38.64 8588.35 -0.004 0.996  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.7605e+02 on 126 degrees of freedom  
## Residual deviance: 4.4171e-08 on 125 degrees of freedom  
## AIC: 4  
##   
## Number of Fisher Scoring iterations: 25

# Logistic Regression with Original Variables

fit.glm = glm(formula = Class ~ .,   
 data = daTrain,   
 family = "binomial")

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Predictions on train set

## create prediction probabilities (on train dataset)  
train\_probs = predict(fit.glm, type="response")  
  
## create predictions (on train dataset)  
train\_preds = as.factor(ifelse(train\_probs > 0.5, "1", "0"))  
  
## evaluate performance (on train dataset)  
confusionMatrix(train\_preds, daTrain$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 64 0  
## 1 0 63  
##   
## Accuracy : 1   
## 95% CI : (0.9714, 1)  
## No Information Rate : 0.5039   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5039   
## Detection Rate : 0.5039   
## Detection Prevalence : 0.5039   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : 0   
##

## Predictions on test set

## create prediction probabilities (on test dataset)  
test\_probs = predict(fit.glm,   
 type = "response",   
 newdata = daTest)  
  
## create predictions (on test dataset)  
test\_preds = as.factor(ifelse(test\_probs > 0.5, "1", "0"))  
  
## evaluate performance (on test dataset)  
confusionMatrix(test\_preds, daTest$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 22 2  
## 1 0 19  
##   
## Accuracy : 0.9535   
## 95% CI : (0.8419, 0.9943)  
## No Information Rate : 0.5116   
## P-Value [Acc > NIR] : 2.642e-10   
##   
## Kappa : 0.9067   
##   
## Mcnemar's Test P-Value : 0.4795   
##   
## Sensitivity : 1.0000   
## Specificity : 0.9048   
## Pos Pred Value : 0.9167   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5116   
## Detection Rate : 0.5116   
## Detection Prevalence : 0.5581   
## Balanced Accuracy : 0.9524   
##   
## 'Positive' Class : 0   
##

## Summary of glm model with all original variables as predictors

summary(fit.glm)

##   
## Call:  
## glm(formula = Class ~ ., family = "binomial", data = daTrain)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.989e-06 -2.372e-06 -2.110e-08 1.315e-06 9.210e-06   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -2.329e+01 3.265e+05 0 1  
## Atr1 5.719e-01 2.840e+05 0 1  
## Atr2 1.240e+00 1.994e+05 0 1  
## Atr3 -7.999e-01 1.996e+05 0 1  
## Atr4 2.048e+00 2.120e+05 0 1  
## Atr5 4.587e+00 2.195e+05 0 1  
## Atr6 4.488e+00 1.708e+05 0 1  
## Atr7 2.133e+00 2.254e+05 0 1  
## Atr8 7.239e-01 1.852e+05 0 1  
## Atr9 1.935e-01 2.824e+05 0 1  
## Atr10 -5.938e-01 2.508e+05 0 1  
## Atr11 -1.208e+00 3.228e+05 0 1  
## Atr12 -2.594e+00 4.537e+05 0 1  
## Atr13 -6.100e+00 2.597e+05 0 1  
## Atr14 4.126e+00 2.615e+05 0 1  
## Atr15 3.155e+00 2.048e+05 0 1  
## Atr16 -3.273e+00 2.181e+05 0 1  
## Atr17 5.094e+00 3.284e+05 0 1  
## Atr18 4.272e+00 3.232e+05 0 1  
## Atr19 -3.486e-01 2.531e+05 0 1  
## Atr20 -6.856e+00 3.151e+05 0 1  
## Atr21 2.836e+00 2.680e+05 0 1  
## Atr22 -8.892e+00 2.909e+05 0 1  
## Atr23 -2.645e-01 3.349e+05 0 1  
## Atr24 -2.133e+00 2.498e+05 0 1  
## Atr25 -8.874e-01 2.653e+05 0 1  
## Atr26 7.473e+00 2.613e+05 0 1  
## Atr27 -4.186e+00 2.687e+05 0 1  
## Atr28 1.086e+01 4.313e+05 0 1  
## Atr29 8.156e-02 4.494e+05 0 1  
## Atr30 -8.671e-01 2.673e+05 0 1  
## Atr31 1.125e+00 1.299e+05 0 1  
## Atr32 -2.277e+00 2.000e+05 0 1  
## Atr33 1.055e+00 2.818e+05 0 1  
## Atr34 1.007e+00 2.498e+05 0 1  
## Atr35 1.056e+00 2.634e+05 0 1  
## Atr36 -1.531e+00 3.632e+05 0 1  
## Atr37 6.231e-01 2.142e+05 0 1  
## Atr38 4.956e+00 3.435e+05 0 1  
## Atr39 -3.421e+00 3.436e+05 0 1  
## Atr40 9.921e+00 2.852e+05 0 1  
## Atr41 -4.510e-01 1.246e+05 0 1  
## Atr42 -1.535e+00 1.390e+05 0 1  
## Atr43 -2.950e+00 1.648e+05 0 1  
## Atr44 -2.994e-01 1.299e+05 0 1  
## Atr45 2.312e+00 1.207e+05 0 1  
## Atr46 -2.676e+00 9.491e+04 0 1  
## Atr47 2.728e+00 1.526e+05 0 1  
## Atr48 -3.598e+00 1.392e+05 0 1  
## Atr49 -1.126e-01 1.507e+05 0 1  
## Atr50 2.449e+00 2.074e+05 0 1  
## Atr51 -1.649e+00 1.732e+05 0 1  
## Atr52 5.013e+00 1.180e+05 0 1  
## Atr53 -1.096e+00 1.574e+05 0 1  
## Atr54 -5.392e+00 2.372e+05 0 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.7605e+02 on 126 degrees of freedom  
## Residual deviance: 9.5353e-10 on 72 degrees of freedom  
## AIC: 110  
##   
## Number of Fisher Scoring iterations: 25

# 10 fold-Cross Validation repeated 3 times

ctrl <- trainControl(method = "repeatedcv",  
 number = 10,  
 repeats = 3)

# Random Forest model with PCA1 component as predictor

set.seed(seed)  
metric <- "Accuracy"  
fit.rf.PCA1 <- train(Class~PC1,  
 data = train.pc,  
 method = "rf",  
 metric = metric,  
 trControl = ctrl,  
 ntree = 500)

## Random forest results with PCA1 component as predictor on train set

print(fit.rf.PCA1)

## Random Forest   
##   
## 127 samples  
## 1 predictor  
## 2 classes: '0', '1'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 114, 113, 114, 115, 114, 113, ...   
## Resampling results:  
##   
## Accuracy Kappa  
## 1 1   
##   
## Tuning parameter 'mtry' was held constant at a value of 2

## Random Forest predictions with PCA1 component as predictor on test set

pred <- predict(fit.rf.PCA1,newdata = test.pc,type="raw")  
confusionMatrix(pred,test.pc$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 22 0  
## 1 0 21  
##   
## Accuracy : 1   
## 95% CI : (0.9178, 1)  
## No Information Rate : 0.5116   
## P-Value [Acc > NIR] : 3.055e-13   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5116   
## Detection Rate : 0.5116   
## Detection Prevalence : 0.5116   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : 0   
##

# Random Forest model with all original variables as predictors

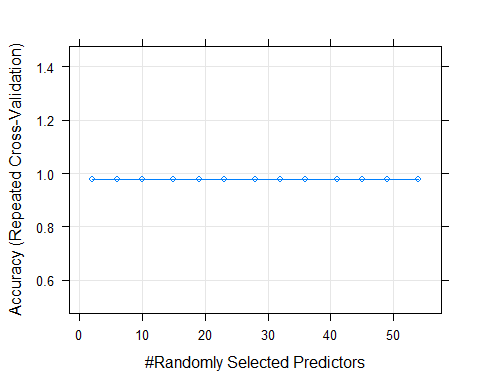
set.seed(seed)  
metric <- "Accuracy"  
fit.rf <- train(Class~.,  
 data = daTrain,  
 method = "rf",  
 metric = metric,  
 trControl = ctrl,  
 ntree = 500,  
 tuneLength = 13)

## Random forest results with all original variables as predictors on train set

print(fit.rf)

## Random Forest   
##   
## 127 samples  
## 54 predictor  
## 2 classes: '0', '1'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 114, 113, 114, 115, 114, 113, ...   
## Resampling results across tuning parameters:  
##   
## mtry Accuracy Kappa   
## 2 0.9760684 0.9521605  
## 6 0.9760684 0.9521605  
## 10 0.9760684 0.9521605  
## 15 0.9760684 0.9521605  
## 19 0.9760684 0.9521605  
## 23 0.9760684 0.9521605  
## 28 0.9760684 0.9521605  
## 32 0.9760684 0.9521605  
## 36 0.9760684 0.9521605  
## 41 0.9760684 0.9521605  
## 45 0.9760684 0.9521605  
## 49 0.9760684 0.9521605  
## 54 0.9760684 0.9521605  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was mtry = 2.

plot(fit.rf)



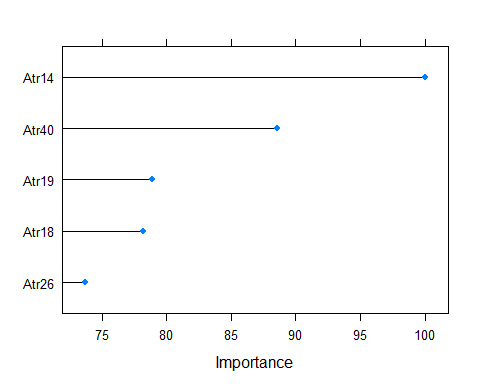
## Random Forest predictions with all original variables as predictors on test set

pred <- predict(fit.rf,newdata = daTest,type="raw")  
confusionMatrix(pred,daTest$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 22 1  
## 1 0 20  
##   
## Accuracy : 0.9767   
## 95% CI : (0.8771, 0.9994)  
## No Information Rate : 0.5116   
## P-Value [Acc > NIR] : 1.285e-11   
##   
## Kappa : 0.9534   
##   
## Mcnemar's Test P-Value : 1   
##   
## Sensitivity : 1.0000   
## Specificity : 0.9524   
## Pos Pred Value : 0.9565   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5116   
## Detection Rate : 0.5116   
## Detection Prevalence : 0.5349   
## Balanced Accuracy : 0.9762   
##   
## 'Positive' Class : 0   
##

# Most important variables for the predictions according to Random Forest Model

imp <- varImp(fit.rf)  
plot(imp,top = 5)



# Testing some more Linear and Non-Linear Algorithms

Linear: Linear Discriminate Analysis(LDA)

Non-Linear: k-Nearest Neighbors(KNN), Classification and Regression Trees(CART),Naive Bayes(Naive\_Bayes) and Support Vector Machines with Radial Basis Functions(SVM).

All the models will be tested with PCA1 as predictor

seed <- 1821  
set.seed(seed)  
metric <- "Accuracy"  
dat <- train.pc  
  
  
  
# LDA  
fit.lda <- train(Class~PC1,  
 data = dat,  
 method ="lda",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson"),  
 na.action = na.omit)  
  
  
  
# KNN  
fit.knn <- train(Class~PC1,  
 data = dat,  
 method = "knn",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson"),  
 na.action = na.omit)  
  
  
# CART  
fit.cart <- train(Class~PC1,  
 data = dat,  
 method = "rpart",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson"),  
 na.action = na.omit)  
  
  
# Naive Bayes  
fit.nb <- train(Class~PC1,  
 data = dat,  
 method = "nb",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson")  
)  
  
  
# Support Vector Machines with Radial Basis Function Kernel  
fit.svmRadial<- train(Class~PC1,  
 data = dat,  
 method = "svmRadial",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson"),  
 na.action = na.omit)  
  
  
# Compare algorithms  
results <- resamples(list(LDA = fit.lda,  
 KNN = fit.knn,  
 CART = fit.cart,  
 NB = fit.nb,  
 SVM = fit.svmRadial))  
summary(results)

##   
## Call:  
## summary.resamples(object = results)  
##   
## Models: LDA, KNN, CART, NB, SVM   
## Number of resamples: 30   
##   
## Accuracy   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## LDA 0.8461538 0.9423077 1 0.9760684 1 1 0  
## KNN 0.8461538 0.9423077 1 0.9762821 1 1 0  
## CART 1.0000000 1.0000000 1 1.0000000 1 1 0  
## NB 1.0000000 1.0000000 1 1.0000000 1 1 0  
## SVM 1.0000000 1.0000000 1 1.0000000 1 1 0  
##   
## Kappa   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## LDA 0.6976744 0.8852941 1 0.9521605 1 1 0  
## KNN 0.6976744 0.8852941 1 0.9526180 1 1 0  
## CART 1.0000000 1.0000000 1 1.0000000 1 1 0  
## NB 1.0000000 1.0000000 1 1.0000000 1 1 0  
## SVM 1.0000000 1.0000000 1 1.0000000 1 1 0

#dotplot(results)

# Ensemble methods

Let’s look at some boosting and and bagging algorithms.

Bagging: Bagged CART(BAG) and Random Forest(RF)

Boosting: Stochastic Gradient Boosting(GBM) and C5.0(C50)

set.seed(seed)  
  
# Bagged CART  
fit.treebag<-train(Class~PC1,  
 data = dat,  
 method = "treebag",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson")  
)  
  
# Random Forest  
fit.rf<-train(Class~PC1,  
 data = dat,  
 method = "rf",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson")  
 )  
  
# Stochastic gradient boosting  
fit.gbm<-train(Class~PC1,  
 data = dat,  
 method = "gbm",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson"),  
 verbose = F)  
  
  
# C5.0  
fit.c50<-train(Class~PC1,  
 data = dat,  
 method = "C5.0",  
 metric = metric,  
 trControl = ctrl,  
 # preProcess=c("center",  
 # "scale",  
 # "YeoJohnson")  
 )  
  
  
# Compare results  
ensembleResults <- resamples(list(BAG = fit.treebag,  
 RF = fit.rf,  
 GBM = fit.gbm,  
 C50 = fit.c50))  
summary(ensembleResults)

##   
## Call:  
## summary.resamples(object = ensembleResults)  
##   
## Models: BAG, RF, GBM, C50   
## Number of resamples: 30   
##   
## Accuracy   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## BAG 1.0000000 1 1 1.0000000 1 1 0  
## RF 0.9230769 1 1 0.9948718 1 1 0  
## GBM 1.0000000 1 1 1.0000000 1 1 0  
## C50 0.8461538 1 1 0.9897131 1 1 0  
##   
## Kappa   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## BAG 1.0000000 1 1 1.0000000 1 1 0  
## RF 0.8433735 1 1 0.9895582 1 1 0  
## GBM 1.0000000 1 1 1.0000000 1 1 0  
## C50 0.6976744 1 1 0.9796050 1 1 0

# dotplot(ensembleResults)  
  
  
  
ensembleResults2 <- resamples(list(BAG = fit.treebag,  
 GBM = fit.gbm,  
 C50 = fit.c50))  
summary(ensembleResults2)

##   
## Call:  
## summary.resamples(object = ensembleResults2)  
##   
## Models: BAG, GBM, C50   
## Number of resamples: 30   
##   
## Accuracy   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## BAG 1.0000000 1 1 1.0000000 1 1 0  
## GBM 1.0000000 1 1 1.0000000 1 1 0  
## C50 0.8461538 1 1 0.9897131 1 1 0  
##   
## Kappa   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## BAG 1.0000000 1 1 1.000000 1 1 0  
## GBM 1.0000000 1 1 1.000000 1 1 0  
## C50 0.6976744 1 1 0.979605 1 1 0

#dotplot(ensembleResults2)

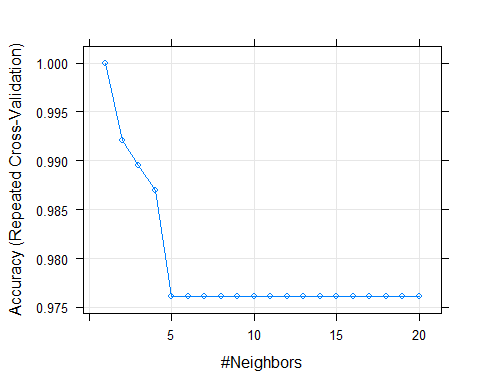
# Tuning KNN

The KNN implementation has one parameter that we can tune with caret: k the number of closest instances to collect in order to make a prediction.Let’s try all k values between 1 and 20.

set.seed(seed)  
 grid <- expand.grid(.k=seq(1,20,by =1))  
  
fit.knn <- train(Class~PC1,  
 data = dat,  
 method = "knn",  
 metric = metric,  
 trControl = ctrl,  
 tuneGrid = grid,  
 # tuneLength = 20,  
 preProcess = c("center",  
 "scale",  
 "YeoJohnson"))  
print(fit.knn)

## k-Nearest Neighbors   
##   
## 127 samples  
## 1 predictor  
## 2 classes: '0', '1'   
##   
## Pre-processing: centered (1), scaled (1), Yeo-Johnson transformation (1)   
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 114, 113, 114, 115, 114, 113, ...   
## Resampling results across tuning parameters:  
##   
## k Accuracy Kappa   
## 1 1.0000000 1.0000000  
## 2 0.9920940 0.9841255  
## 3 0.9895299 0.9791460  
## 4 0.9869658 0.9740480  
## 5 0.9760684 0.9521605  
## 6 0.9760684 0.9521605  
## 7 0.9760684 0.9521605  
## 8 0.9760684 0.9521605  
## 9 0.9760684 0.9521605  
## 10 0.9760684 0.9521605  
## 11 0.9760684 0.9521605  
## 12 0.9760684 0.9521605  
## 13 0.9760684 0.9521605  
## 14 0.9760684 0.9521605  
## 15 0.9760684 0.9521605  
## 16 0.9760684 0.9521605  
## 17 0.9760684 0.9521605  
## 18 0.9760684 0.9521605  
## 19 0.9760684 0.9521605  
## 20 0.9760684 0.9521605  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was k = 1.

plot(fit.knn)

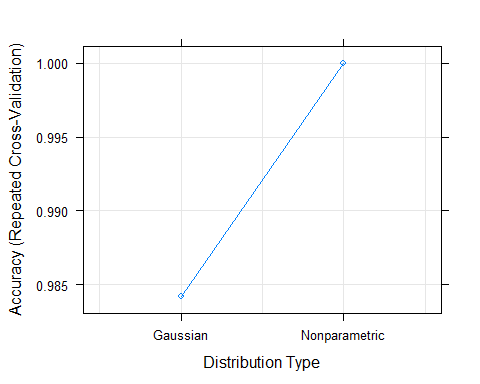


# Tuning NaiveBayes

set.seed(seed)  
  
fit.nb<- train(Class~PC1,  
 data = dat,  
 method = "nb",  
 metric = metric,  
 trControl = ctrl,  
 tuneLength = 20,  
 preProcess = c("center",  
 "scale",  
 "YeoJohnson"))  
print(fit.nb)

## Naive Bayes   
##   
## 127 samples  
## 1 predictor  
## 2 classes: '0', '1'   
##   
## Pre-processing: centered (1), scaled (1), Yeo-Johnson transformation (1)   
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 114, 113, 114, 115, 114, 113, ...   
## Resampling results across tuning parameters:  
##   
## usekernel Accuracy Kappa   
## FALSE 0.984188 0.968251  
## TRUE 1.000000 1.000000  
##   
## Tuning parameter 'fL' was held constant at a value of 0  
## Tuning  
## parameter 'adjust' was held constant at a value of 1  
## Accuracy was used to select the optimal model using the largest value.  
## The final values used for the model were fL = 0, usekernel = TRUE and adjust  
## = 1.

plot(fit.nb)



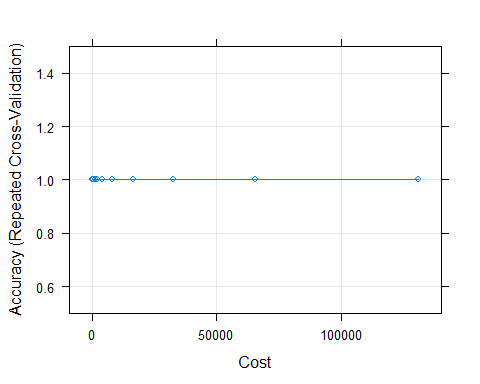
# Tuning SVM

The SVM implementation has two parameters that we can tune with caret package. The Sigma which is a smoothing term, and C which is a cost constraint.

set.seed(seed)  
# Support Vector Machines with Radial Basis Function Kernel.  
# The Radial basis function kernel, also called the RBF kernel,  
# or Gaussian kernel.  
  
# grid<- expand.grid(.sigma = c(0.025,0.05,0.1,0.15),  
# .C = seq(1,10,by=1))  
  
fit.svmRadial.tuned<- train(Class~PC1,  
 data = dat,   
 method = "svmRadial",  
 metric = metric,  
 trControl = ctrl,  
 preProcess = c("center",  
 "scale",  
 "YeoJohnson"),  
 # tuneGrid = grid)  
 tuneLength = 20)  
  
print(fit.svmRadial.tuned)

## Support Vector Machines with Radial Basis Function Kernel   
##   
## 127 samples  
## 1 predictor  
## 2 classes: '0', '1'   
##   
## Pre-processing: centered (1), scaled (1), Yeo-Johnson transformation (1)   
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 114, 113, 114, 115, 114, 113, ...   
## Resampling results across tuning parameters:  
##   
## C Accuracy Kappa  
## 0.25 1 1   
## 0.50 1 1   
## 1.00 1 1   
## 2.00 1 1   
## 4.00 1 1   
## 8.00 1 1   
## 16.00 1 1   
## 32.00 1 1   
## 64.00 1 1   
## 128.00 1 1   
## 256.00 1 1   
## 512.00 1 1   
## 1024.00 1 1   
## 2048.00 1 1   
## 4096.00 1 1   
## 8192.00 1 1   
## 16384.00 1 1   
## 32768.00 1 1   
## 65536.00 1 1   
## 131072.00 1 1   
##   
## Tuning parameter 'sigma' was held constant at a value of 139.8535  
## Accuracy was used to select the optimal model using the largest value.  
## The final values used for the model were sigma = 139.8535 and C = 0.25.

plot(fit.svmRadial.tuned)



# Final model for SVM

fit.svmRadial.tuned$finalModel

## Support Vector Machine object of class "ksvm"   
##   
## SV type: C-svc (classification)   
## parameter : cost C = 0.25   
##   
## Gaussian Radial Basis kernel function.   
## Hyperparameter : sigma = 139.853464616555   
##   
## Number of Support Vectors : 64   
##   
## Objective Function Value : -5.2354   
## Training error : 0

# Predictions of SVM

pred<-predict(fit.svmRadial.tuned, newdata = test.pc,type = "raw")  
 confusionMatrix(pred, test.pc$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 22 0  
## 1 0 21  
##   
## Accuracy : 1   
## 95% CI : (0.9178, 1)  
## No Information Rate : 0.5116   
## P-Value [Acc > NIR] : 3.055e-13   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5116   
## Detection Rate : 0.5116   
## Detection Prevalence : 0.5116   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : 0   
##

# XGBoost

set.seed(seed)  
tune\_grid <- expand.grid(nrounds = 300,  
 max\_depth = c(14),  
 eta = c(0.01),  
 gamma = 0.01,  
 colsample\_bytree = 0.85,  
 min\_child\_weight = 1,  
 subsample = 0.5)  
  
fit.xgb <- train(Class~PC1,   
 data = dat,   
 method = "xgbTree",  
 trControl = ctrl,  
 tuneGrid = tune\_grid,  
 # tuneLength = 15,  
 preProcess = c("center",  
 "scale",  
 "YeoJohnson"))  
print(fit.xgb)

## eXtreme Gradient Boosting   
##   
## 127 samples  
## 1 predictor  
## 2 classes: '0', '1'   
##   
## Pre-processing: centered (1), scaled (1), Yeo-Johnson transformation (1)   
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 114, 113, 114, 115, 114, 113, ...   
## Resampling results:  
##   
## Accuracy Kappa   
## 0.9948718 0.9899225  
##   
## Tuning parameter 'nrounds' was held constant at a value of 300  
## Tuning  
##   
## Tuning parameter 'min\_child\_weight' was held constant at a value of 1  
##   
## Tuning parameter 'subsample' was held constant at a value of 0.5

# XGBoost predictions on test set

pred<-predict(fit.xgb, newdata = test.pc,type = "raw")  
 confusionMatrix(pred, test.pc$Class)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 22 0  
## 1 0 21  
##   
## Accuracy : 1   
## 95% CI : (0.9178, 1)  
## No Information Rate : 0.5116   
## P-Value [Acc > NIR] : 3.055e-13   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5116   
## Detection Rate : 0.5116   
## Detection Prevalence : 0.5116   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : 0   
##

# Neural network model

library(h2o)

##   
## ----------------------------------------------------------------------  
##   
## Your next step is to start H2O:  
## > h2o.init()  
##   
## For H2O package documentation, ask for help:  
## > ??h2o  
##   
## After starting H2O, you can use the Web UI at http://localhost:54321  
## For more information visit https://docs.h2o.ai  
##   
## ----------------------------------------------------------------------

##   
## Attaching package: 'h2o'

## The following objects are masked from 'package:stats':  
##   
## cor, sd, var

## The following objects are masked from 'package:base':  
##   
## %\*%, %in%, &&, ||, apply, as.factor, as.numeric, colnames,  
## colnames<-, ifelse, is.character, is.factor, is.numeric, log,  
## log10, log1p, log2, round, signif, trunc

h2oInstance <- h2o.init(ip="localhost")

##   
## H2O is not running yet, starting it now...  
##   
## Note: In case of errors look at the following log files:  
## C:\Users\izzyd\AppData\Local\Temp\Rtmp4qm66l\file9db849a02eda/h2o\_izzyd\_started\_from\_r.out  
## C:\Users\izzyd\AppData\Local\Temp\Rtmp4qm66l\file9db84c9735e0/h2o\_izzyd\_started\_from\_r.err  
##   
##   
## Starting H2O JVM and connecting: Connection successful!  
##   
## R is connected to the H2O cluster:   
## H2O cluster uptime: 3 seconds 940 milliseconds   
## H2O cluster timezone: Europe/Athens   
## H2O data parsing timezone: UTC   
## H2O cluster version: 3.32.1.3   
## H2O cluster version age: 3 months and 3 days   
## H2O cluster name: H2O\_started\_from\_R\_izzyd\_vyn879   
## H2O cluster total nodes: 1   
## H2O cluster total memory: 1.76 GB   
## H2O cluster total cores: 8   
## H2O cluster allowed cores: 8   
## H2O cluster healthy: TRUE   
## H2O Connection ip: localhost   
## H2O Connection port: 54321   
## H2O Connection proxy: NA   
## H2O Internal Security: FALSE   
## H2O API Extensions: Amazon S3, Algos, AutoML, Core V3, TargetEncoder, Core V4   
## R Version: R version 4.0.3 (2020-10-10)

set.seed(seed)  
h2oInstance

## IP Address: localhost   
## Port : 54321   
## Name : NA   
## Session ID: \_sid\_865f   
## Key Count : 0

trH <- as.h2o(train.pc,"trH")

## | | | 0% | |======================================================================| 100%

tsH <- as.h2o(test.pc,"tsH")

## | | | 0% | |======================================================================| 100%

deep.fit <- h2o.deeplearning(  
 x = 1:length(train.pc)-1,  
 y = length(train.pc),  
 training\_frame = trH,  
 hidden = c(10,10),  
 epochs = 50,  
 nfolds = 2,  
 standardize = F,  
 autoencoder = F,  
 mini\_batch\_size = 4,  
 # categorical\_encoding = c("LabelEncoder"),  
 # balance\_classes = T,  
 # class\_sampling\_factors = c(2.49,8),  
 # max\_after\_balance\_size = 5,  
 use\_all\_factor\_levels = T,  
 seed = seed)

## | | | 0% | |===================================================================== | 98% | |======================================================================| 100%

## Predictions

h2o.confusionMatrix(deep.fit,tsH)

## Confusion Matrix (vertical: actual; across: predicted) for max f1 @ threshold = 0.952944838382645:  
## 0 1 Error Rate  
## 0 22 0 0.000000 =0/22  
## 1 0 21 0.000000 =0/21  
## Totals 22 21 0.000000 =0/43

h2o.shutdown(prompt = F)