Breast Cancer Analysis

- Apeksha Shetty

- Yash Shah

- Roshni Suhanda

GITHUB LINK : <https://github.com/yshah2232/MVA>

wisc\_bc\_df <- read.csv("C://Users//Yshah//Downloads//Rutgers Sem 2//MVA//wisc\_bc\_data.csv")

wisc\_bc\_df

#Reading and input of data set into dataset name : wisc\_bc\_df

#Renmaing the dataset

cancer<-wisc\_bc\_df

Dataset rename to cancer for ease

library("ggplot2")

Library used to call the package ggplot2 to plot the bar plot.

library("corrplot")

Library used to call the package corrplot to plot correlation between the variables.

library("reshape")

Library used to call the package reshape for melt function(plotting a histogram of together of two categories).

#number of observations and data type

str(cancer)

Gives you the data type and examples of data displayed

569 observation with 32 variables

#gives you the summary of the dataset

summary(cancer)

Provide you values of your dataset

Minimum value

1st quarter value

Median

Mean

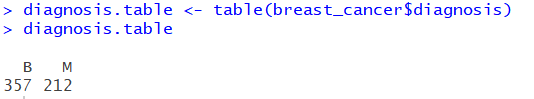
3rd quarter value

Maximum value

#Gives you frequency table

diagnosis.table <- table(cancer$diagnosis)

diagnosis.table

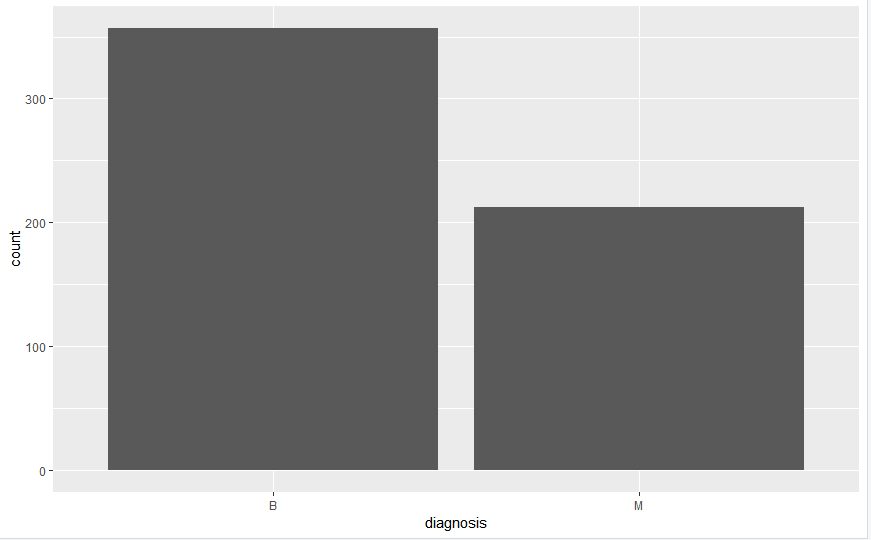


Diagnosis table provides you the frequency of two type of categories

M= Malignant (indicates prescence of infectious cancer cells)

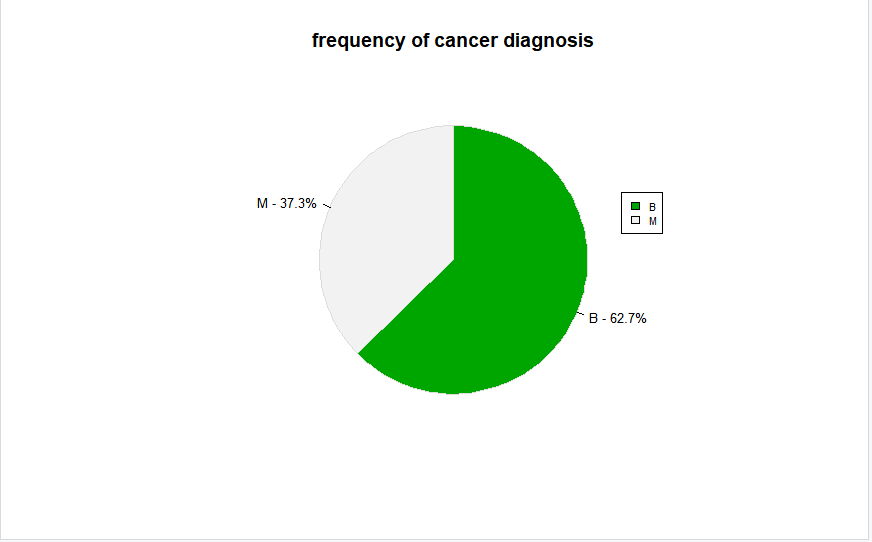
B= Benign (indicates non-infectious abscence)

#Barplot



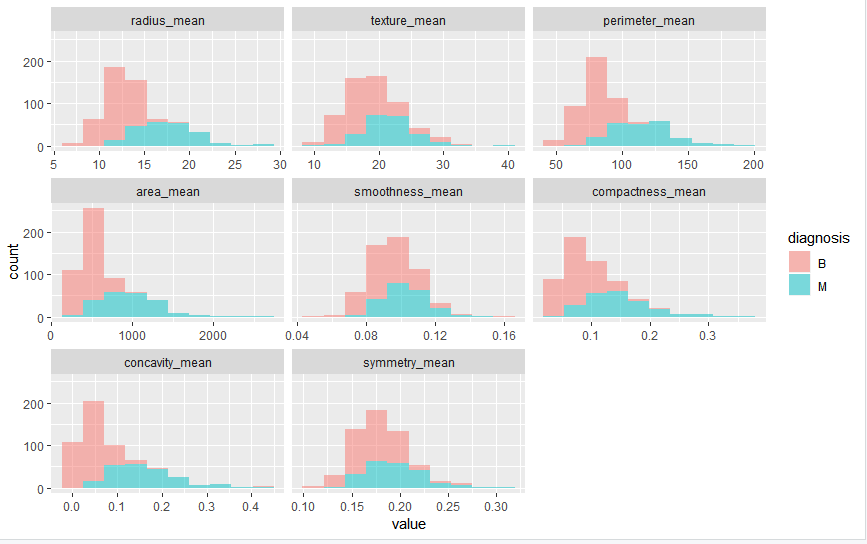
Using ggplot package we plot the frequency of the two categories as Bar plot.

## Pie Chart



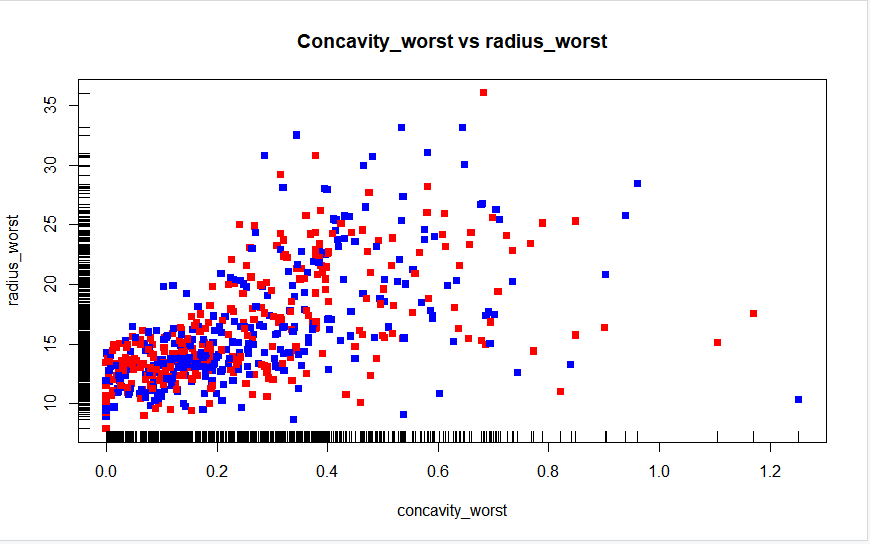
Gives you a frequency of cases of two type categories on basis of percentage.

##Histogram



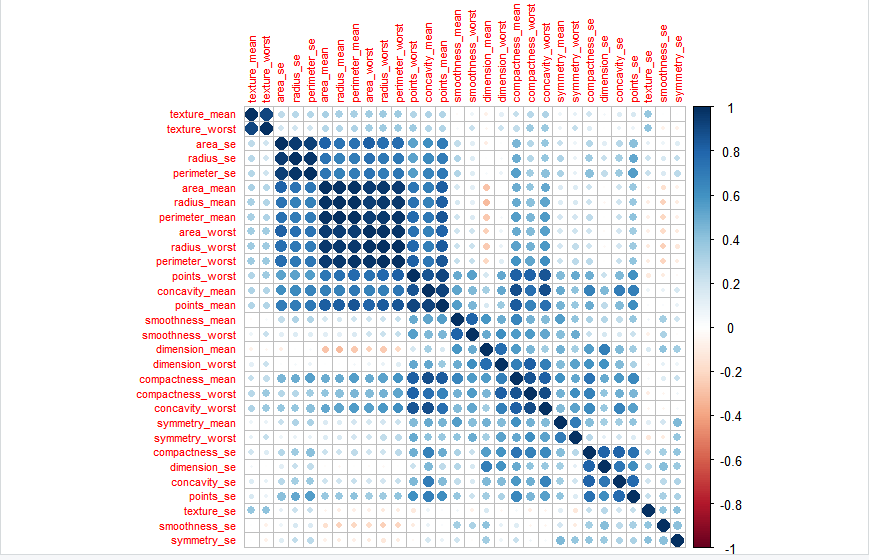
Plot histogram of two categories in one graph for better comparison and visualization.

##Scatterplot



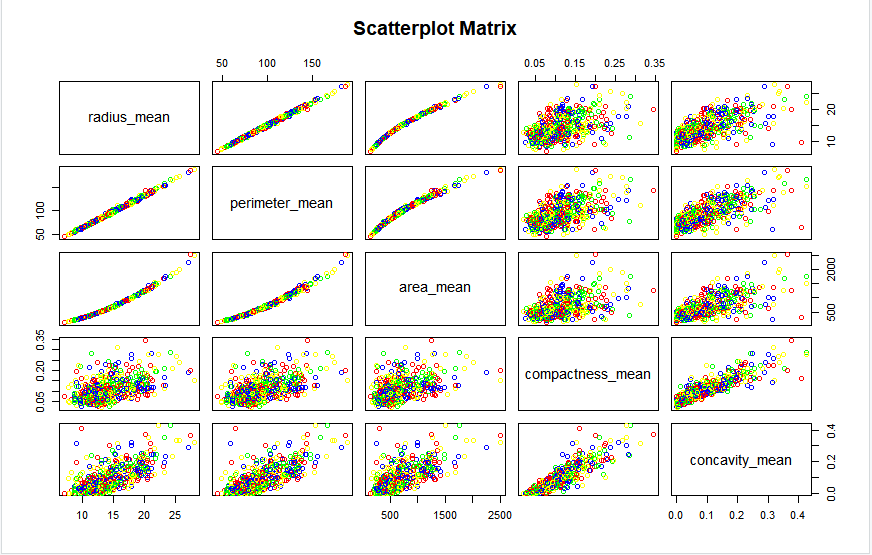
Plots scatter plot with values of two variable that is radius\_worst and concavity\_worst representing you the concentration of most occurring values and can be used to plot outliers as well as develop a relationship between the for the making an analysis assuming the more the radius and more of concavity may interpret a worst case scenario of breast cancer analysis.

##Correlation Matrix



Plots scatter plot with values of two variable that is radius\_worst and concavity\_worst representing you the concentration of most occurring values and can be used to plot outliers as well as develop a relationship between the for the making an analysis assuming the more the radius and more of concavity may interpret a worst case scenario of breast cancer analysis.

##Scatterplot Matrix



A scatter plot matrix is table of scatter plots. Each plot is small so that many plots can be fit on a page. When you need to look at several plots.