#Analysis of dataset in R ##By Sabiha Mhatranaik

#Load Dataset packages

library(readxl)

#import the excel file into the global environment

Dataset\_Media\_ <- read\_excel("Machine Learning/Dataset\_Media -.xls")

#head function gives the first 6 rows

head(Dataset\_Media\_)

## # A tibble: 6 x 5  
## id media Preference Category Audience\_Size  
## <chr> <chr> <dbl> <chr> <dbl>  
## 1 s1 Times Of india 1 Newspaper 23  
## 2 s2 Hindustan Times 1 Newspaper 67  
## 3 s3 Mumbai Mirror 1 Newspaper 43  
## 4 s4 Lokmat 1 Newspaper 32  
## 5 s5 The Economic Times 1 Newspaper 11  
## 6 s6 Mid Day 1 Newspaper 54

#tail function gives the last 6 rows

tail(Dataset\_Media\_)

## # A tibble: 6 x 5  
## id media Preference Category Audience\_Size  
## <chr> <chr> <dbl> <chr> <dbl>  
## 1 s12 yahoo.com 3 Online 44  
## 2 s13 ndtv.com 3 Online 17  
## 3 s14 timesofIndia.com 3 Online 63  
## 4 s15 news.google.com 3 Online 54  
## 5 s16 Indiatimes.com 3 Online 49  
## 6 s17 aajtak.in 3 Online 50

#str function will structure the data

str(Dataset\_Media\_)

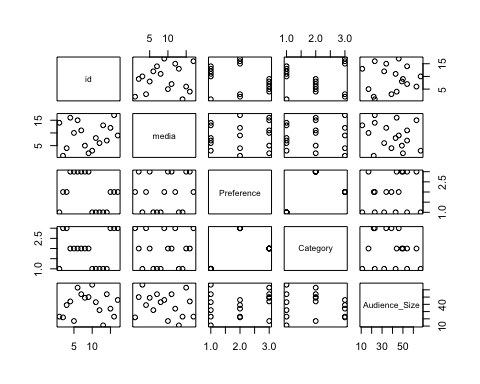
## tibble [17 × 5] (S3: tbl\_df/tbl/data.frame)  
## $ id : chr [1:17] "s1" "s2" "s3" "s4" ...  
## $ media : chr [1:17] "Times Of india" "Hindustan Times" "Mumbai Mirror" "Lokmat" ...  
## $ Preference : num [1:17] 1 1 1 1 1 1 2 2 2 2 ...  
## $ Category : chr [1:17] "Newspaper" "Newspaper" "Newspaper" "Newspaper" ...  
## $ Audience\_Size: num [1:17] 23 67 43 32 11 54 34 23 46 22 ...

#summary function will give the summary of the data

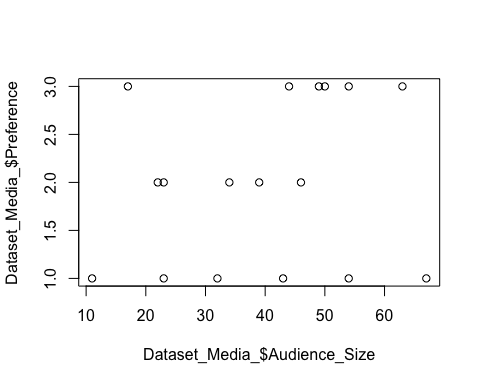
summary(Dataset\_Media\_)

## id media Preference Category   
## Length:17 Length:17 Min. :1 Length:17   
## Class :character Class :character 1st Qu.:1 Class :character   
## Mode :character Mode :character Median :2 Mode :character   
## Mean :2   
## 3rd Qu.:3   
## Max. :3   
## Audience\_Size   
## Min. :11.00   
## 1st Qu.:23.00   
## Median :43.00   
## Mean :39.47   
## 3rd Qu.:50.00   
## Max. :67.00

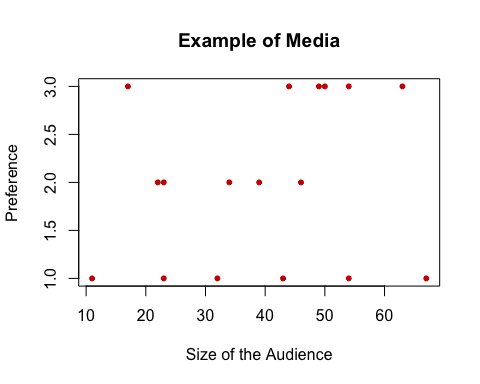
plot(Dataset\_Media\_)

 #X-Y plot for two quantitative variables(Scatterplot)

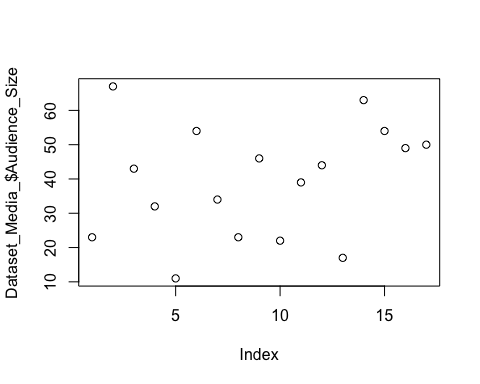
plot(Dataset\_Media\_$Audience\_Size,Dataset\_Media\_$Preference)

 #Add some Options

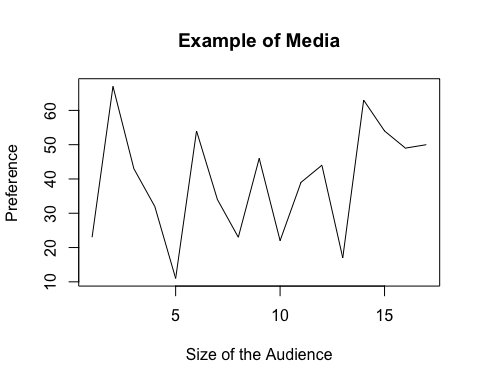
plot(Dataset\_Media\_$Audience\_Size,Dataset\_Media\_$Preference,pch=20,col="#cc0000",main="Example of Media",xlab="Size of the Audience",ylab="Preference")

 #Plotting a quantitative variable Audience\_Size

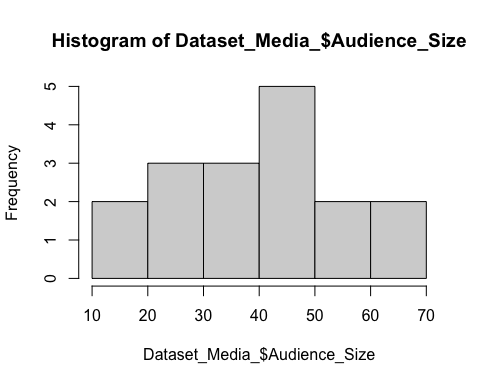
plot(Dataset\_Media\_$Audience\_Size)

 #Lineplot,Histogram,Boxplot

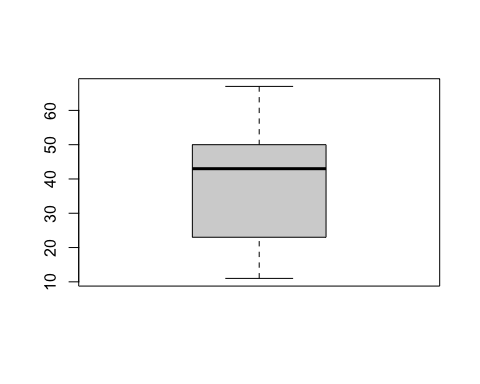
plot(Dataset\_Media\_$Audience\_Size,type="l",main="Example of Media",xlab="Size of the Audience",ylab="Preference")



hist(Dataset\_Media\_$Audience\_Size)



boxplot(Dataset\_Media\_$Audience\_Size)

 #Working with functions

min(Dataset\_Media\_$Audience\_Size)

## [1] 11

max(Dataset\_Media\_$Audience\_Size)

## [1] 67

range(Dataset\_Media\_$Audience\_Size)

## [1] 11 67

mean(Dataset\_Media\_$Audience\_Size)

## [1] 39.47059

median(Dataset\_Media\_$Audience\_Size)

## [1] 43

#The End