**INSTRUCTIONS:**

**# Set the Working Directory**

setwd("D:/Studies/VIII Semester/R/Programs/Ex 3")

**#Get the Working Directory**

getwd()

**#Read the Dataset of CSV Format**

data\_cleaning <- read.csv("D:/Studies/VIII Semester/R/Programs/Ex 3/data\_cleaning.csv", header=TRUE)

**#View the Minimum Value in the Column**

min(data\_cleaning$Budget\_2018\_2019\_Total)

**#View the Maximum Value in the Column**

max(data\_cleaning$Budget\_2018\_2019\_Total)

**#View the Mean Value of a Column**

mean(data\_cleaning$Budget\_2018\_2019\_Total)

**#View the Standard Deviation of a Column**

sd(data\_cleaning$Budget\_2018\_2019\_Total)

**#View the Dataset**

View(data\_cleaning)

**#View the Dataset Class**

class(data\_cleaning)

**#View the Dimention of Dataset**

dim(data\_cleaning)

**#View the Number of Columns**

length(data\_cleaning)

**#View the Names of Columns**

names(data\_cleaning)

**#View the Number of Rows**

nrow(data\_cleaning)

**#View the Structure of Dataset**

str(data\_cleaning)

**#View the Summary of Dataset**

summary(data\_cleaning)

**#View the First N Rows**

head(data\_cleaning, n=10)

**#View the Last N Rows**

tail(data\_cleaning,n=10)

**#View the Dataset in Data Editor**

fix(data\_cleaning)

**#View the Glimpse of Data in a Column**

glimpse(data\_cleaning$Budget\_2018\_2019\_Total)

**#NOISY DATA**

**#binning**

bins<-5

minimumVal<-min(data\_cleaning$Budget\_2018\_2019\_Total)

maximumVal<-max(data\_cleaning$Budget\_2018\_2019\_Total)

width=(maximumVal-minimumVal)/bins;

x<-cut(data\_cleaning$Budget\_2018\_2019\_Total, breaks=seq(minimumVal, maximumVal, width))

x

barplot(data\_cleaning$Budget\_2018\_2019\_Total)

barplot(table(cut(data\_cleaning$Budget\_2018\_2019\_Total, breaks=seq(minimumVal, maximumVal, width))))

**#regression**

scatter.smooth(x=data\_cleaning$Revised\_2017\_2018\_Revenue, y=data\_cleaning$Budget\_2018\_2019\_Revenue, main="Revised\_2018\_2019\_Revenue ~ Budget\_2018\_2019\_Revenue")

boxplot(data\_cleaning$Budget\_2018\_2019\_Total, main="Budget\_2017\_2018\_Revenue", sub=paste("Outlier rows: ", boxplot.stats(data\_cleaning$Budget\_2017\_2018\_Revenue)$out)) # box plot for 'Budget\_2017\_2018\_Revenue'

**#Outlier Analysis**

data\_cleaning$Budget\_2018\_2019\_Total[data\_cleaning$Budget\_2018\_2019\_Total<0]<-NA

**#MISSING VALUES**

**#To replace the data with mean value**

for (i in which(sapply(data\_cleaning, is.numeric)))

{

data\_cleaning[is.na(data\_cleaning[, i]), i] <- mean(data\_cleaning[, i], na.rm = TRUE)

}

**#Ignoring the tuples**

data\_new<-na.omit(data\_cleaning)

data\_new

fix(data\_new)

**#Pictorial representation**

**#histogram**

hist(data\_cleaning$Budget\_2018\_2019\_Total)

**#plot**

plot(data\_cleaning$Budget\_2017\_2018\_Total,data\_cleaning$Budget\_2018\_2019\_Total)

**#boxplot**

boxplot(data\_cleaning)

boxplot(data\_cleaning, las = 2,col = c("red","sienna","red","sienna","royalblue2","red","sienna","royalblue2","red","sienna","royalblue2","red","sienna","royalblue2"),at = c(1,2, 4,5,6, 8,9,10, 12,13,14, 16,17,18),par(mar = c(14, 5, 4, 2)- 1),names = c("Index","Partriculars","Actual\_2016\_2017\_Revenue","Actual\_2016\_2017\_Capital","Actual\_2016\_2017\_Total","Budget\_2017\_2018\_Revenue","Budget\_2017\_2018\_Capital","Budget\_2017\_2018\_Total","Revised\_2017\_2018\_Revenue","Revised\_2017\_2018\_Capital","Revised\_2017\_2018\_Total","Budget\_2018\_2019\_Revenue","Budget\_2018\_2019\_Capital","Budget\_2018\_2019\_Total "))

**OUTPUT:**

**#Get the Working Directory**

[1] "D:/Studies/VIII Semester/R/Programs/Ex 3"

**#View the Minimum Value in the Column**

[1] -2100.00

**#View the Maximum Value in the Column**

[1] 982.72

**#View the Mean Value of a Column**

[1] 185.11

**#View the Dataset Class**

[1] "data.frame"

**#View the Dimention of Dataset**

[1] 77 14

**#View the Number of Columns**

[1] 14

**#View the Names of Columns**

[1] "Index" "Particulars"

[3] "Actual\_2016\_2017\_Revenue" "Actual\_2016\_2017\_Capital"

[5] "Actual\_2016\_2017.\_Total" "Budget\_2017\_2018\_Revenue"

[7] "Budget\_2017\_2018\_Capital" "Budget\_2017\_2018\_Total"

[9] "Revised\_2017\_2018\_Revenue" "Revised\_2017\_2018\_Capital"

[11] "Revised\_2017\_2018\_Total" "Budget\_2018\_2019\_Revenue"

[13] "Budget\_2018\_2019\_Capital" "Budget\_2018\_2019\_Total"

**#View the Number of Rows**

[1] 77

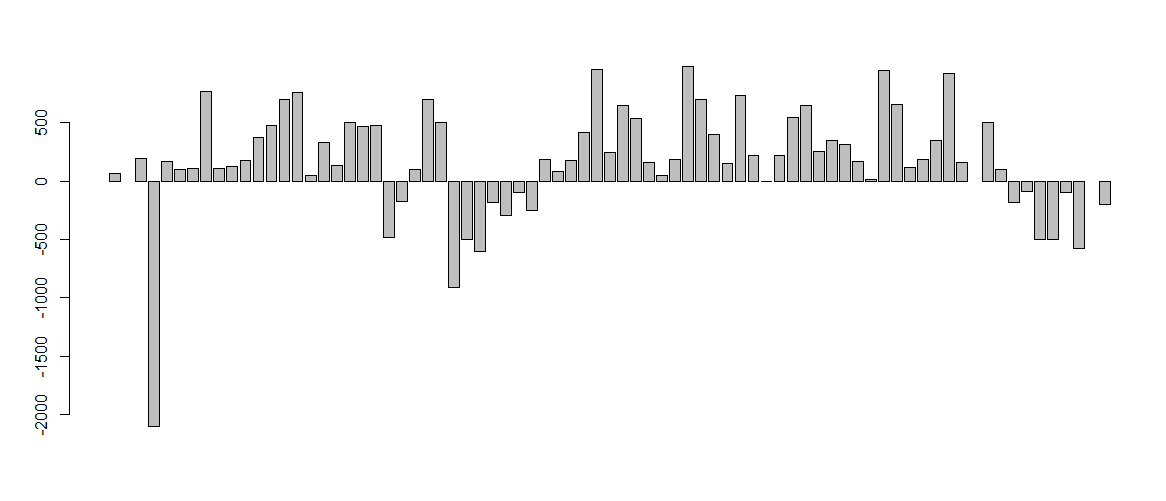
**#View the Glimpse of Data in a Column**

num [1:77] 752 193 -2100 170 ...

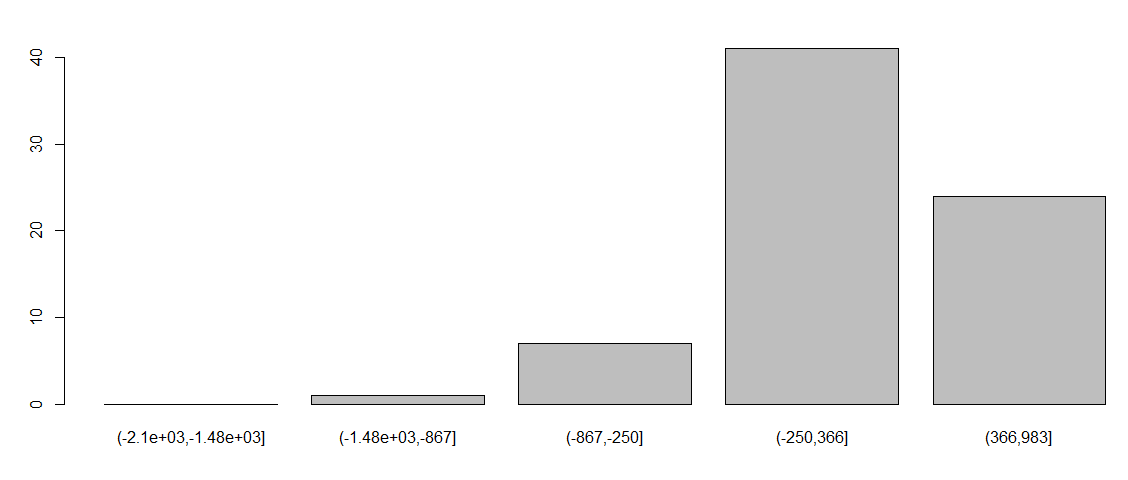
**#NOISY DATA**

**#binning**

**#before:**

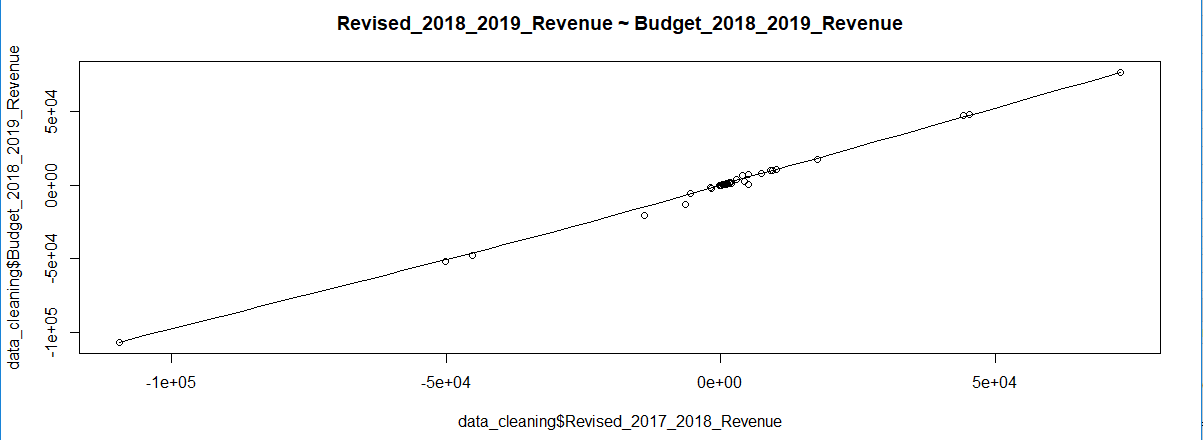
****

**After:**

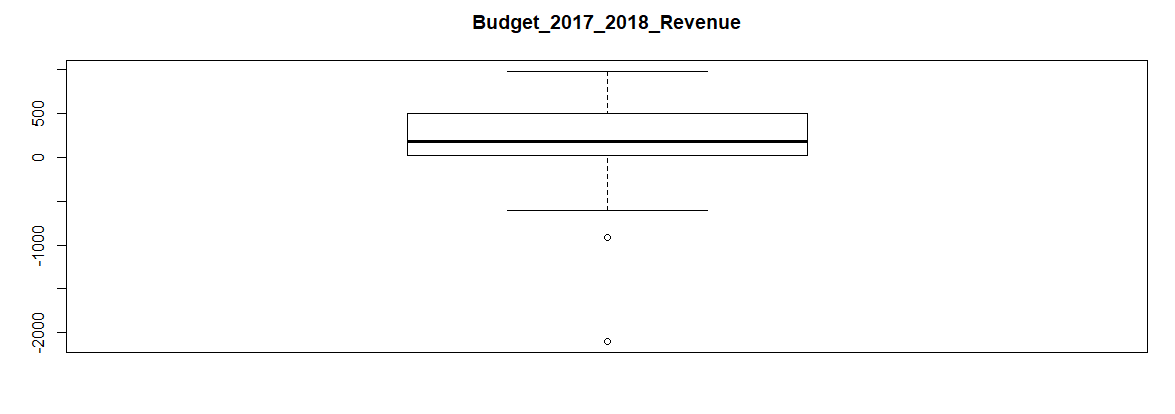
****

**#regression**

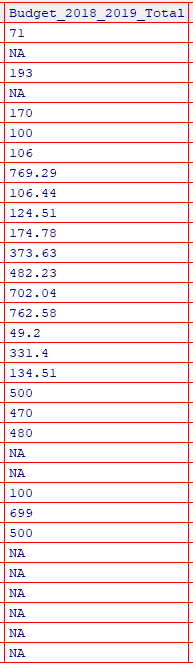
**scatter plot:**



**BOX PLOT:**

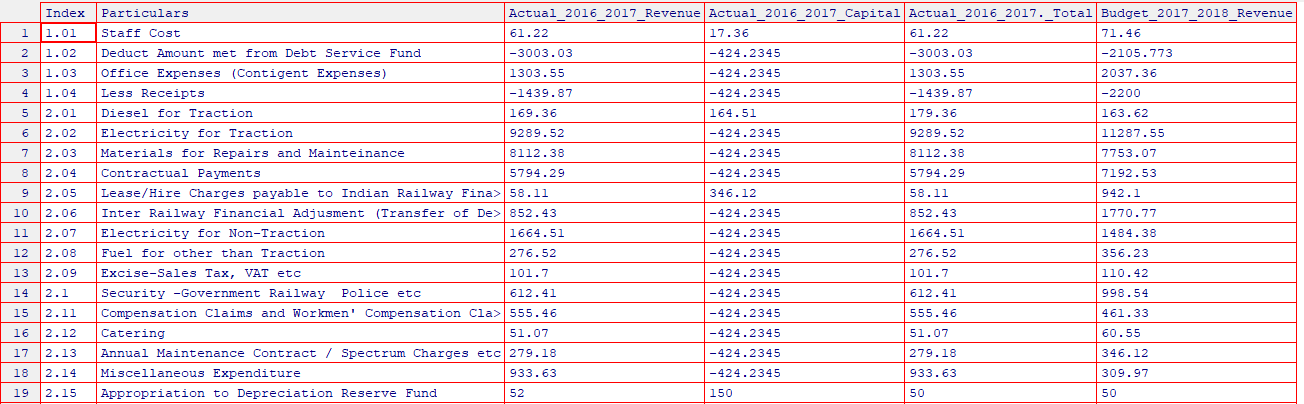
****

**#Outlier Analysis**

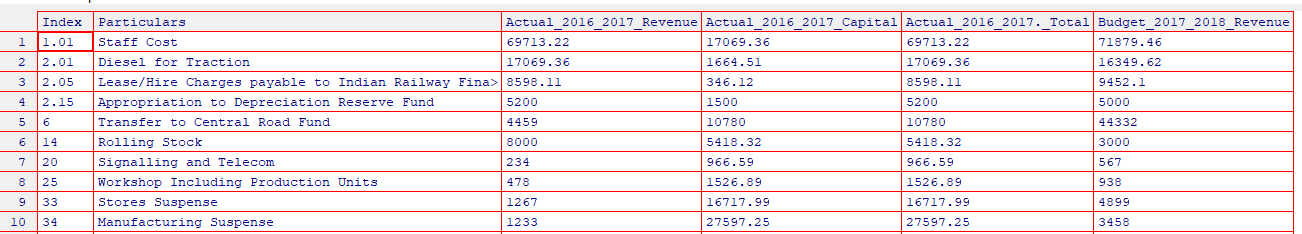


**#MISSING VALUES**

**#To replace the data with mean value**



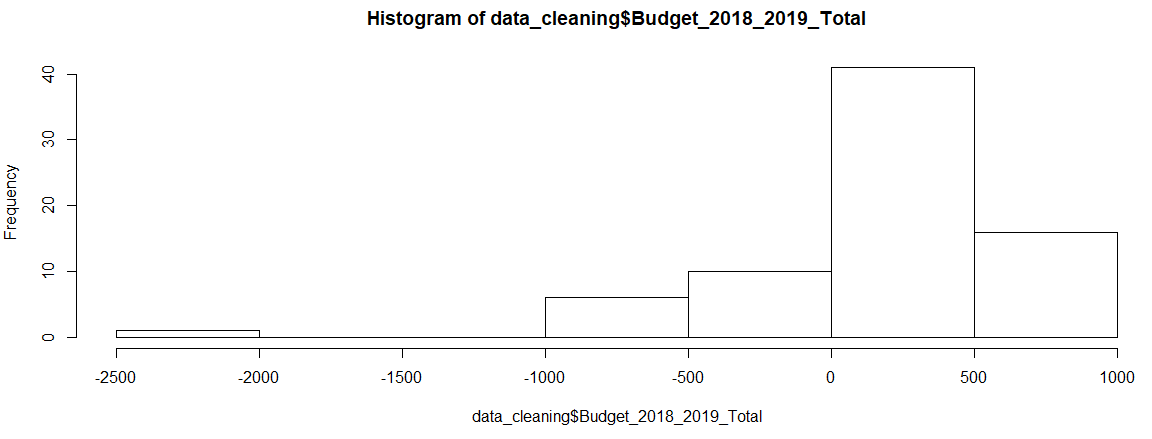
**#Ignoring the tuples**



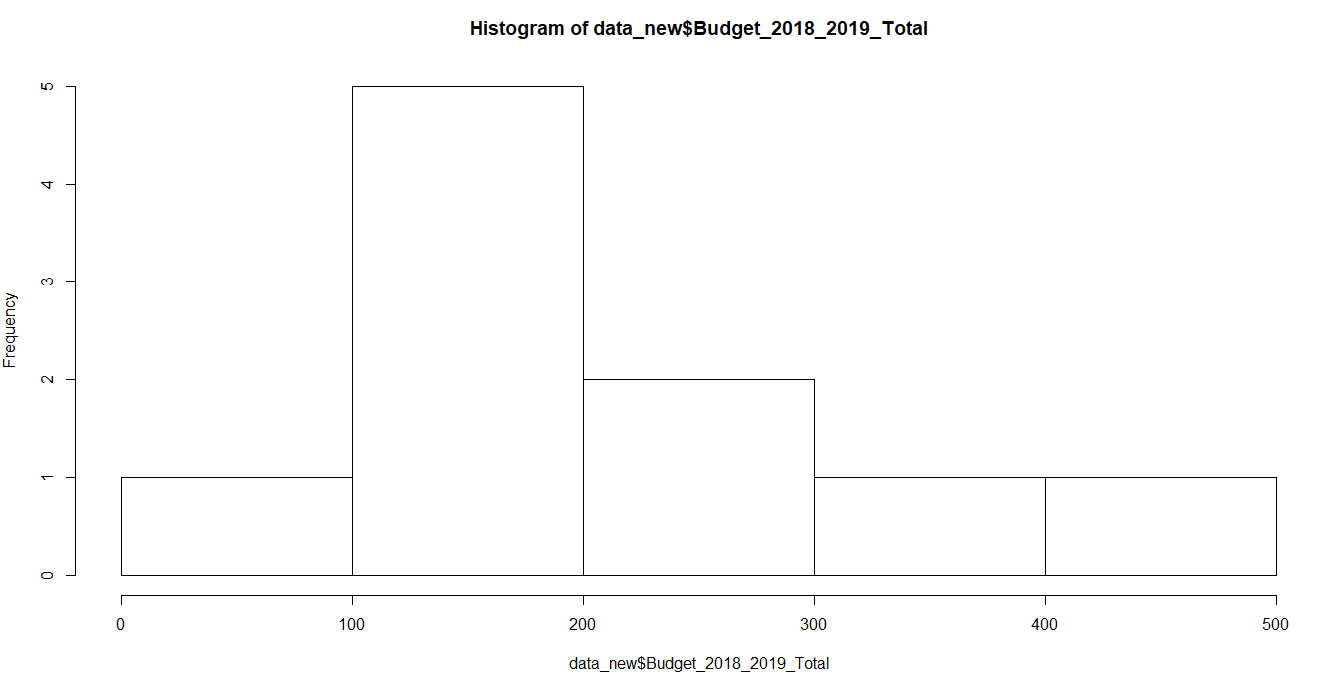
**#PICTORIAL REPRESENTATION**

**#HISTOGRAM**

**Before:**

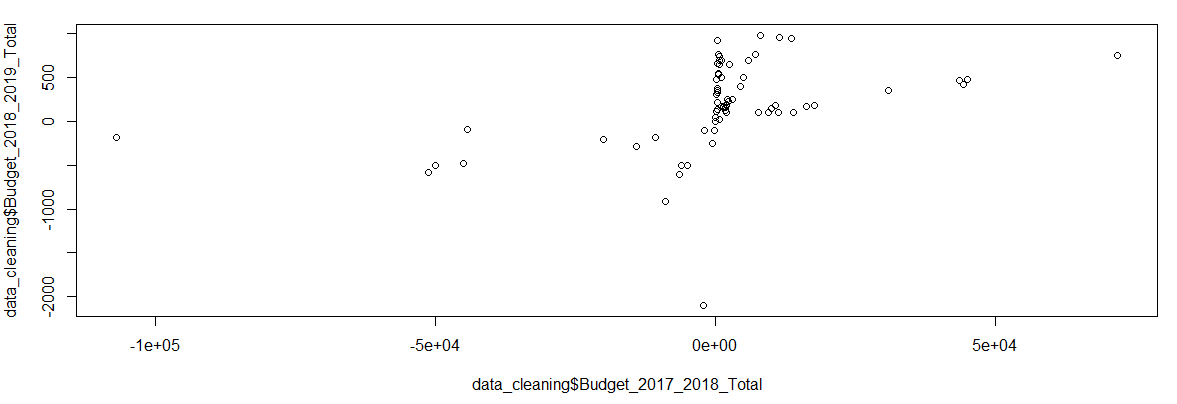
****

**After:**

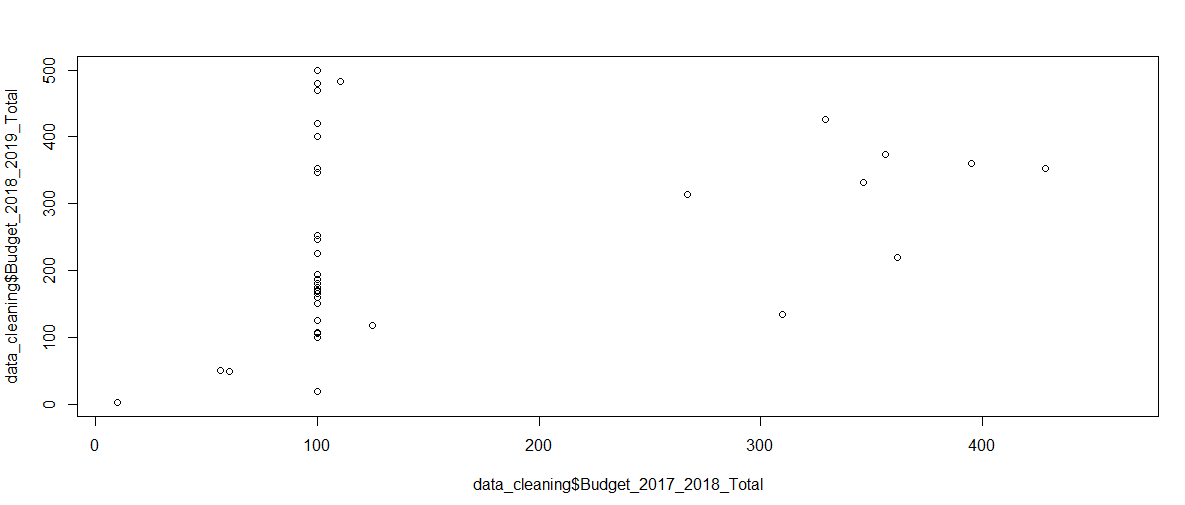
****

**PLOT:**

**Before:**



**After:**

****