NAME: KRISHNA BARFIWALA

ROLL NO: 19

Introduction:

The public sector includes government administration, departmental undertakings, government companies and statutory corporations. The private corporate sector comprises of non-government non-financial corporate enterprise. The rest is termed household sector. Thus, the household sector, being residual in character, includes a host of economic agents who engage in production/consumption activity. Savings are an important part of any economy. Savings can be invested as capital and that helps economic growth. It is known that Asian families are more inclined to saving compared to West European and North American families, which tend to borrow and spend. A low rate of saving would be below 10% of income. Asian families may save as much as a quarter of their earnings.
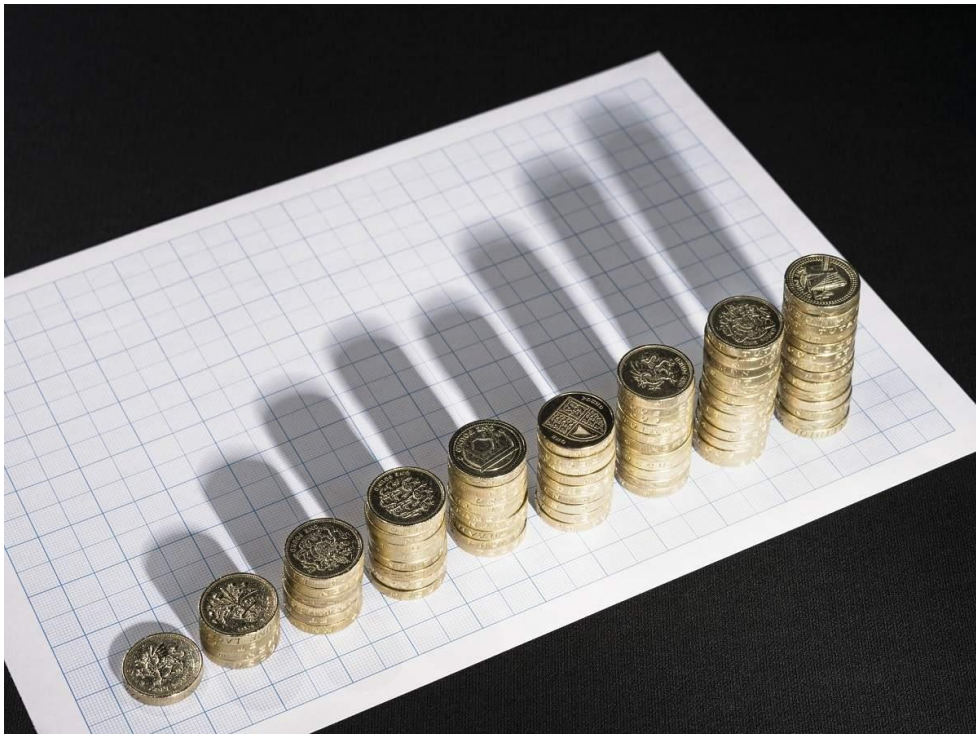
Objective:

To understand the data about savings in different sectors that is household sector, private corporate sector and public sector. To find the mean estimate, the correlation coefficient and hence finding the relationship between the two. And finally checking multivariate normality.

Data description:

Column	Description
A	Savings in Household sector (Rs. Crores)
B	Savings in Private corporate sector (Rs. Crores)
C	Savings in Public sector (Rs. Crores)
D	Total GDS (Rs. Crores)

Methodology:

1) Data cleaning: -check data info and clean through unnecessary data.

2) Visualization: - The first step is to visualize and perform multivariate analysis to explore data to find useful insights.

3).Actionable insights:- Running codes for the same.

Data analysis:

- Data used for the analysis:

	Household.sector	Private.Corporate.Sector	Public.Sector	Total
1	612	93	182	887

2	583	136	266	985
3	637	64	160	861
4	665	90	143	898
5	719	118	169	1006
6	1046	134	190	1370

- Here is the Structure of the data. This code is the best for displaying contents of a list.

```
> str(data)
'data.frame':   53 obs. of  4 variables:
 $ Household.sector      : int  612 583 637 665 719 1046 1178 997 1016 1301 ...
 $ Private.Corporate.Sector: int  93 136 64 90 118 134 155 121 140 185 ...
 $ Public.Sector         : int  182 266 160 143 169 190 251 266 251 262 ...
 $ Total                 : int  887 985 861 898 1006 1370 1584 1384 1407 1748 ...
```

> summary (data)

This is the summary of the data which includes Minimum, Quartiles, Median, Mean and Maximum.

```
 Household.sector Private.Corporate.Sector Public.Sector           Total
 Min.   :   583   Min.   :   64            Min.   :-62704.0   Min.   :   861
 1st Qu.:  1618   1st Qu.:  389            1st Qu.:   266.0   1st Qu.:  2763
 Median : 11849   Median : 1419            Median :  1379.0   Median : 17408
 Mean   : 80349   Mean   :14558            Mean   :   824.9   Mean   : 95732
 3rd Qu.: 86955   3rd Qu.:11845            3rd Qu.:  7168.0   3rd Qu.:106979
 Max.   :559258   Max.   :86142            Max.   : 24065.0   Max.   :597697
```

- Dispersion matrix :

```
                          Household.sector Private.Corporate.Sector Public.Sector
Household.sector                19325643907               3508166183    -1320836519
Private.Corporate.Sector         3508166183                670266655     -198460311
Public.Sector                   -1320836519               -198460311      224487853
Total                           21512973571               3979972527    -1294808978
                                Total
Household.sector          21512973571
Private.Corporate.Sector   3979972527
Public.Sector             -1294808978
Total                     24198137121
```

- Correlation matrix:

	Household.sector	Private.Corporate.Sector	Public.Sector
Household.sector	1.0000000	0.9747413	-0.6341407
Private.Corporate.Sector	0.9747413	1.0000000	-0.5116270
Public.Sector	-0.6341407	-0.5116270	1.0000000
Total	0.9948154	0.9882465	-0.5555438

	Total
Household.sector	0.9948154
Private.Corporate.Sector	0.9882465
Public.Sector	-0.5555438
Total	1.0000000

Interpretation:

A +1 correlation coefficient indicates a perfect positive relationship. Since each variable has to correlate perfectly with itself. Correlation coefficient between Household.sector and Private.Corporate.Sector is 0.9747413 indicating a very strong positive correlation coefficient. Correlation coefficient between Household.sector and Public.Sector is -0.6341407 indicating a strong negative correlation coefficient. Correlation coefficient between Private.Corporate.Sector and Public.Sector is -0.5116270 indicating a moderate negative correlation coefficient. Correlation coefficient between Public.Sector and Total is -0.5555438 indicating a moderate positive correlation coefficient. Correlation coefficient between Total and Household.sector is 0.9948154 indicating a very strong positive correlation coefficient. Correlation coefficient between Total and Private.Corporate.Sector is 0.9882465 indicating a very strong positive correlation coefficient.
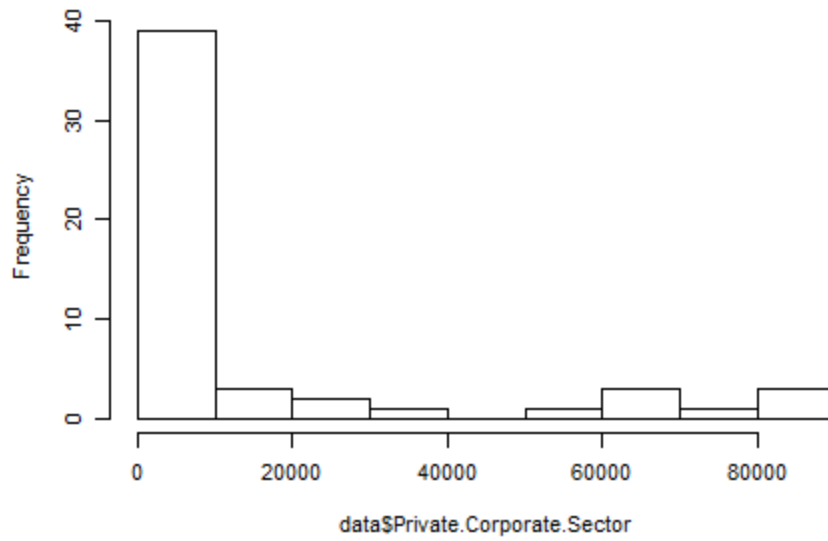
Representation of data graphically:
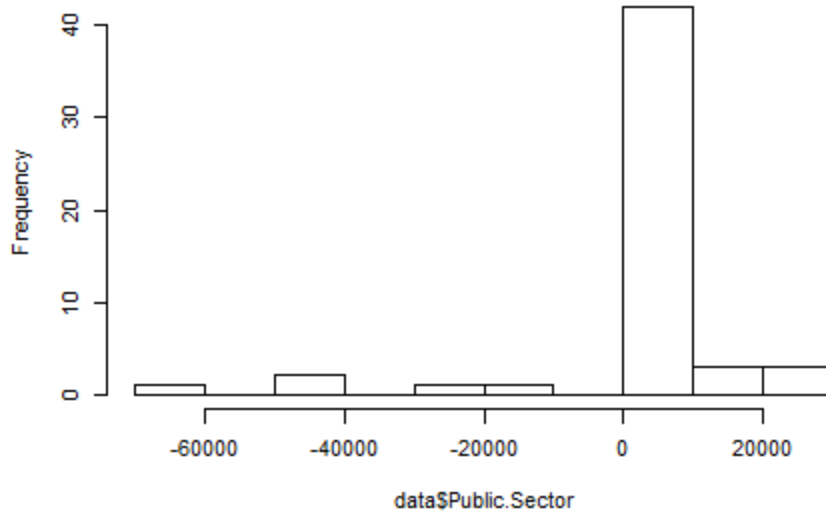
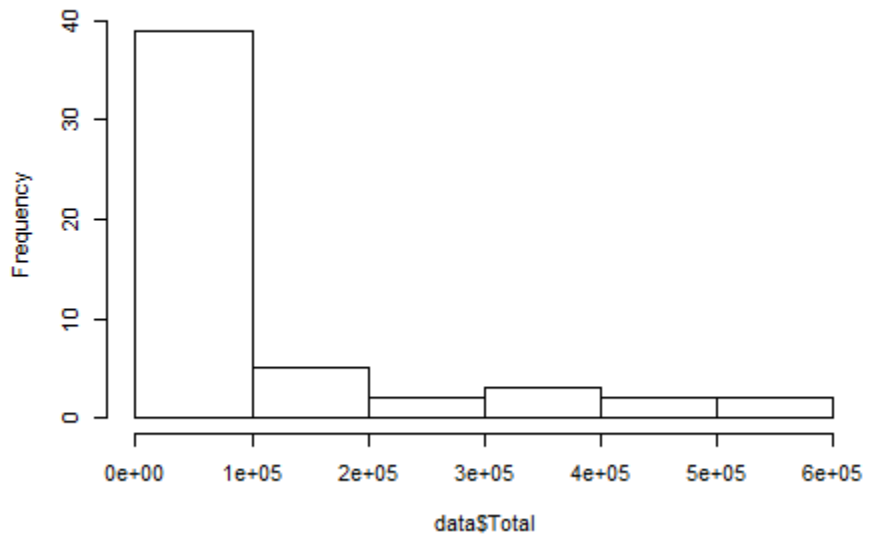# HISTOGRAM of the respective data.

Histogram of data$Household.sector
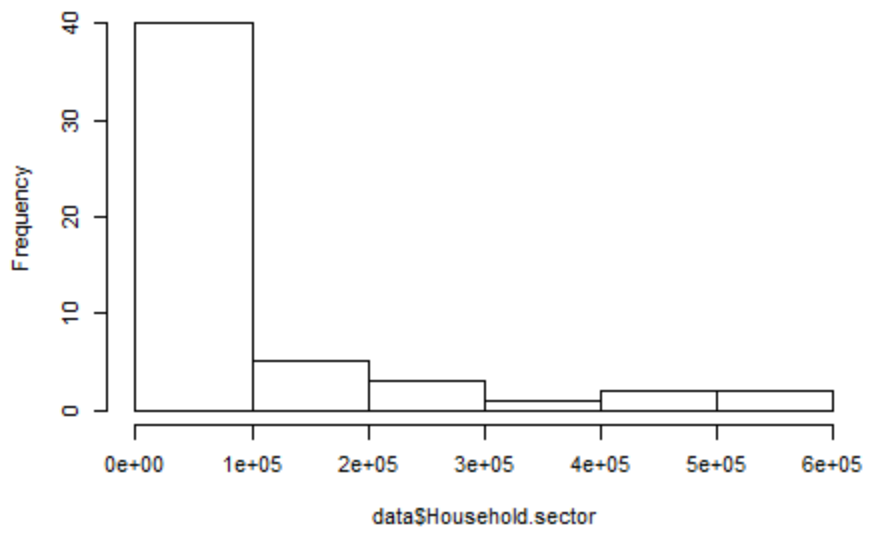
Histogram of data$Private.Corporate.Sector
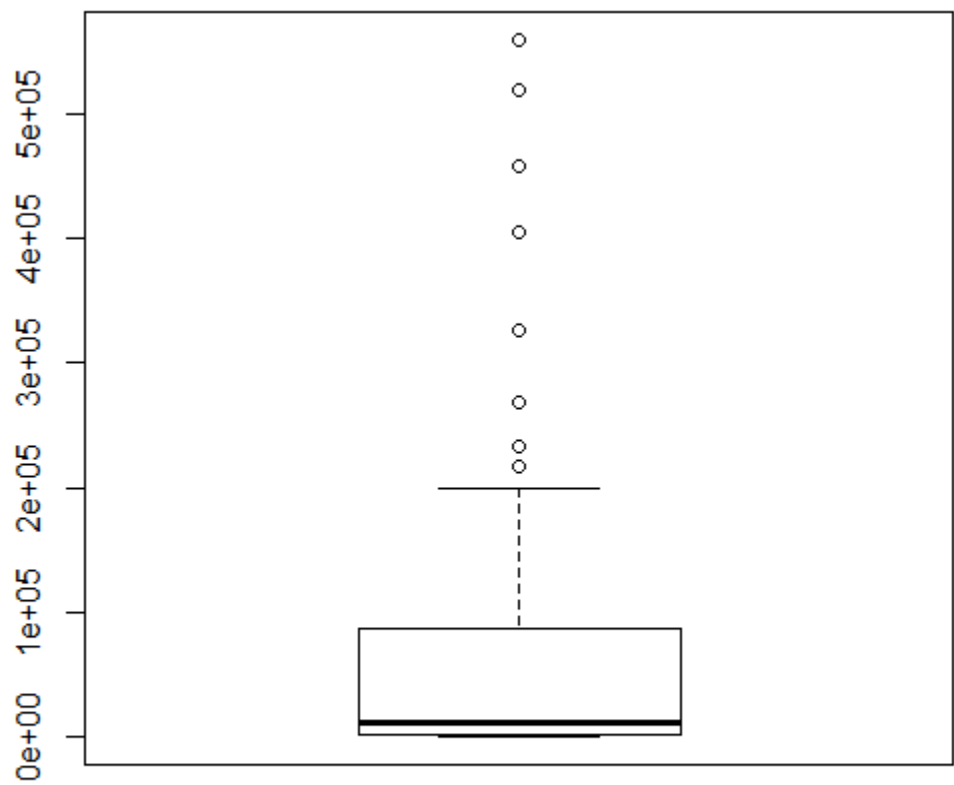
Histogram of data$Public.Sector

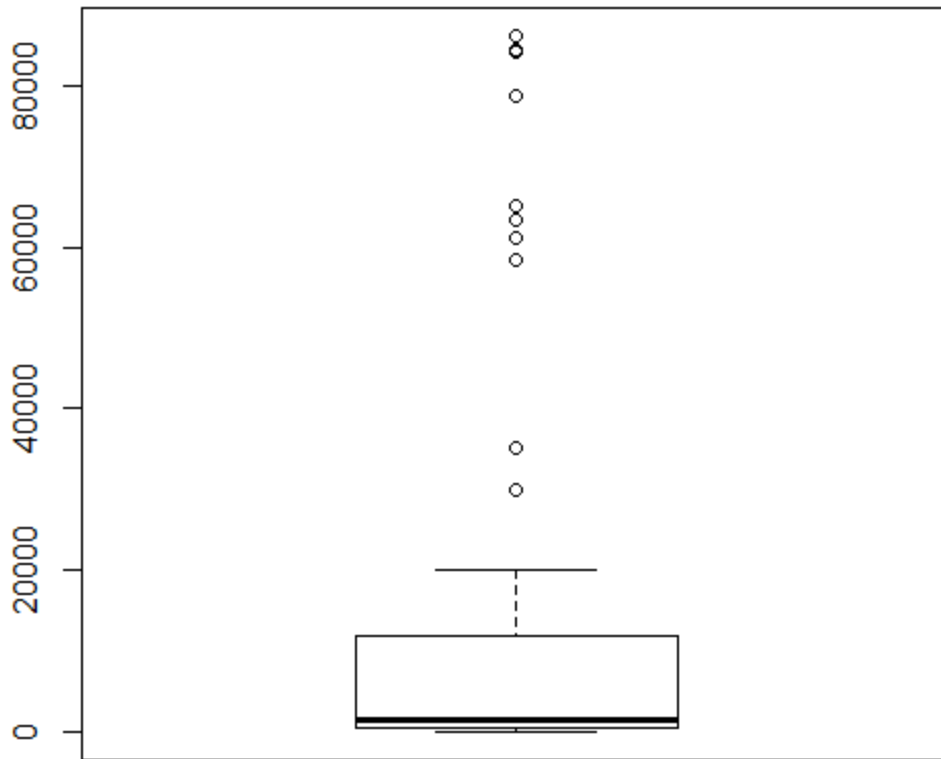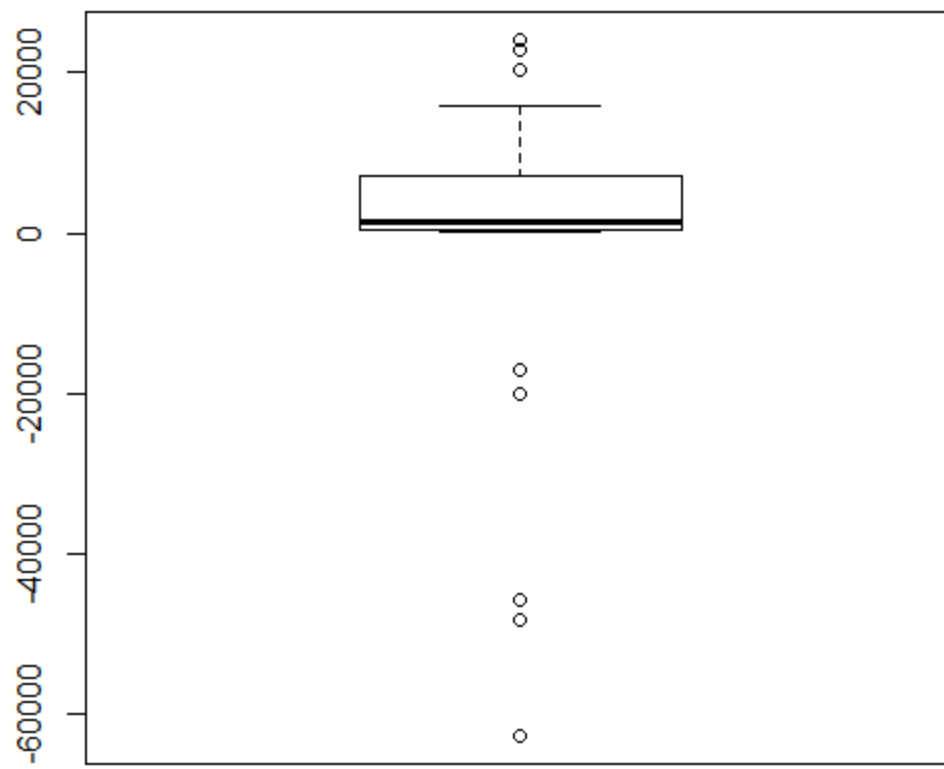
Histogram of data$Total

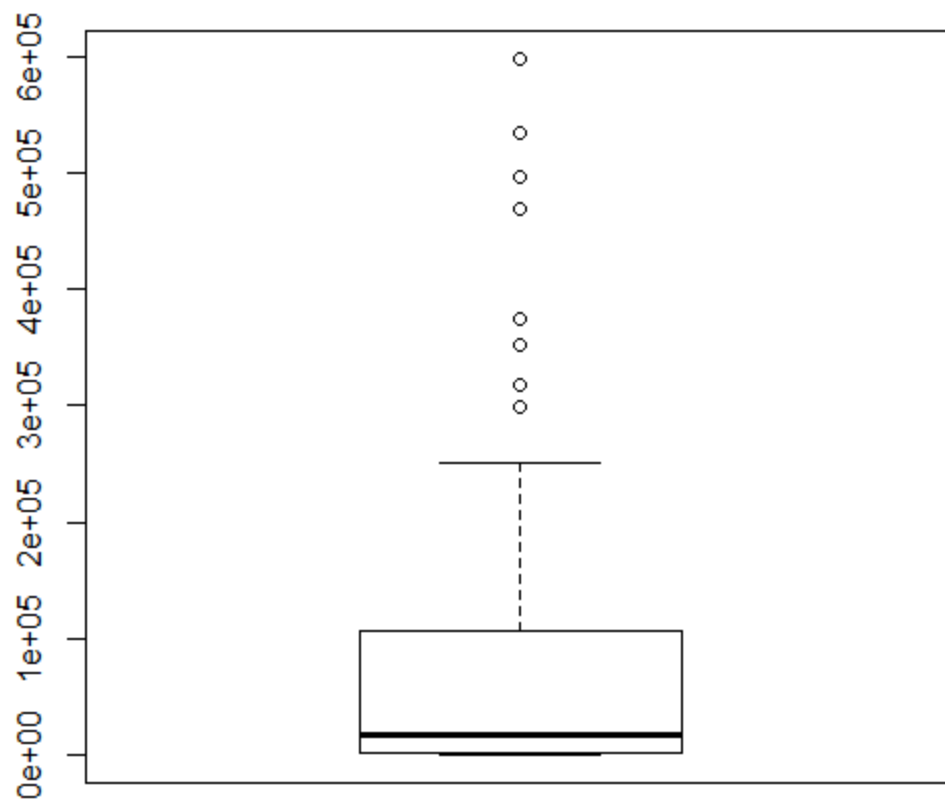#BOXPLOT of the data.

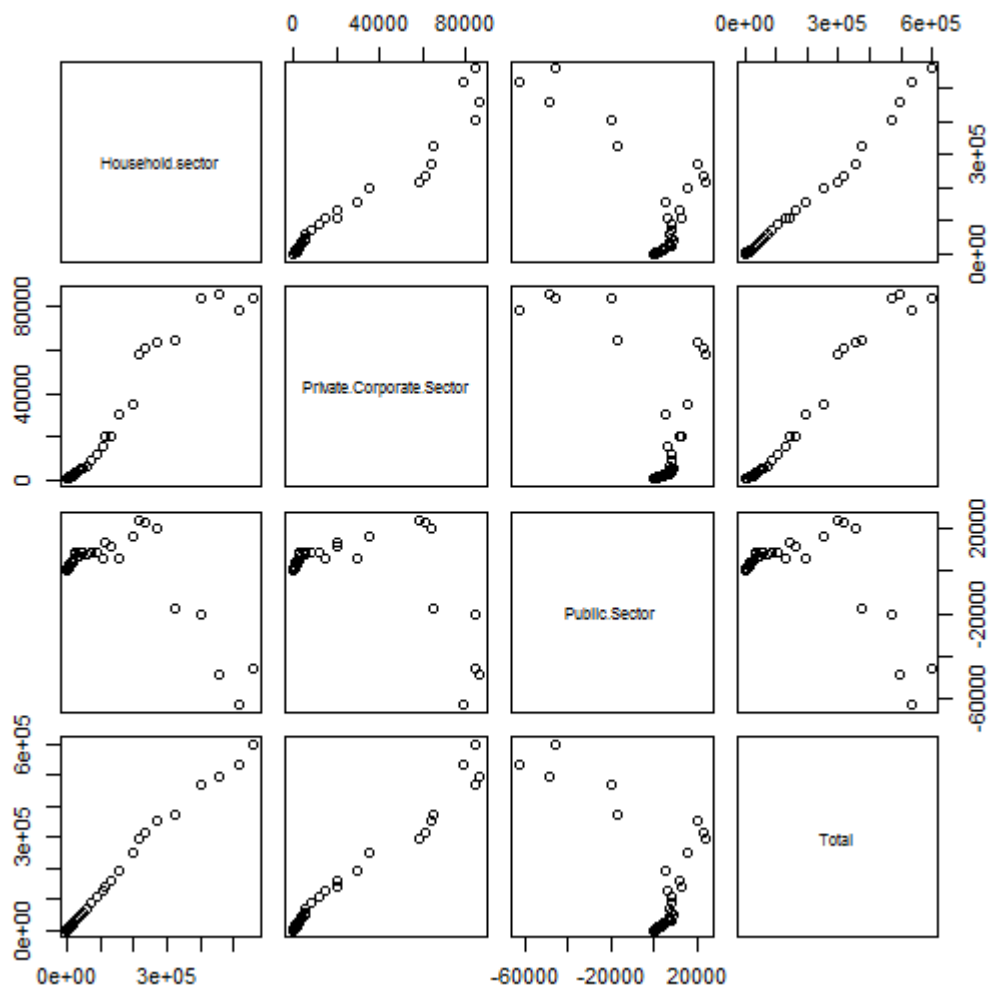#Below is the boxplot for household sector:

Below is the Boxplot for private corporate sector:
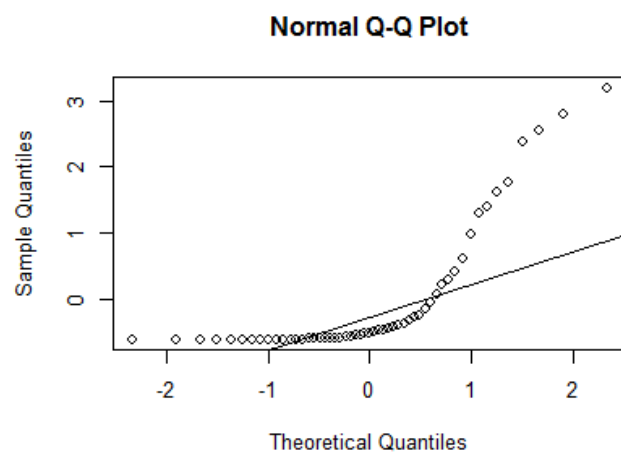
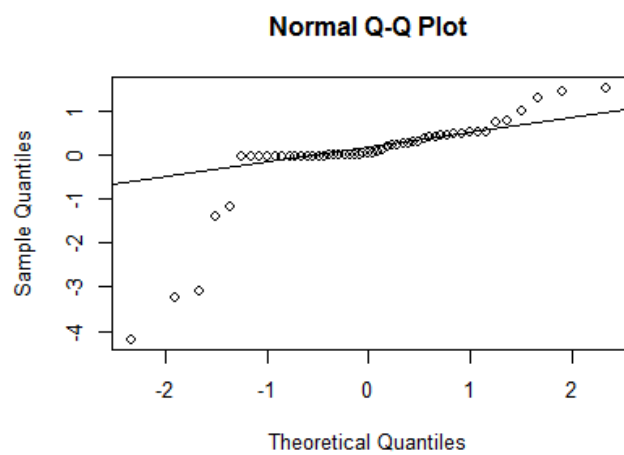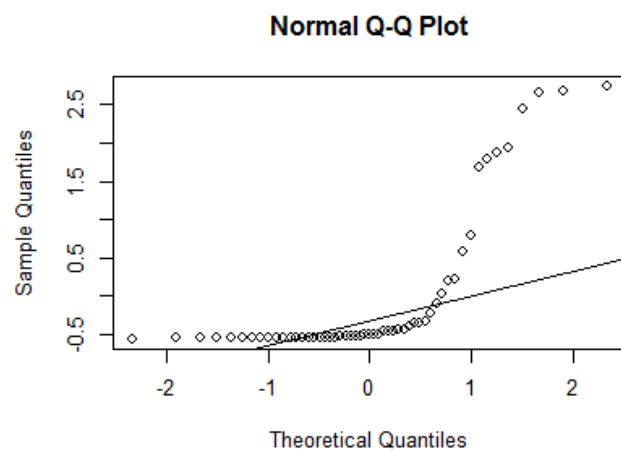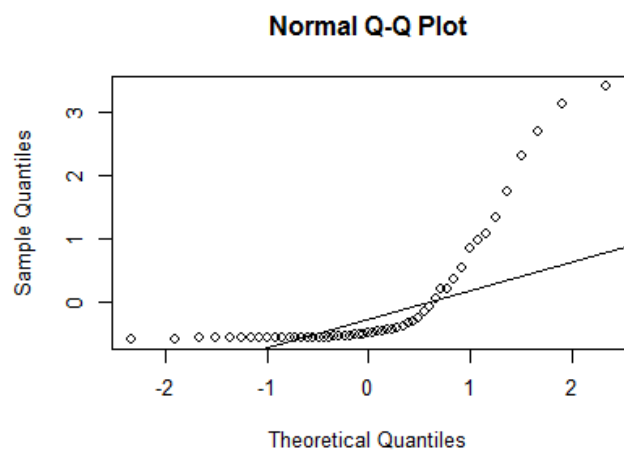# Below is the Boxplot for public sector:

Below is the Boxplot for the total:
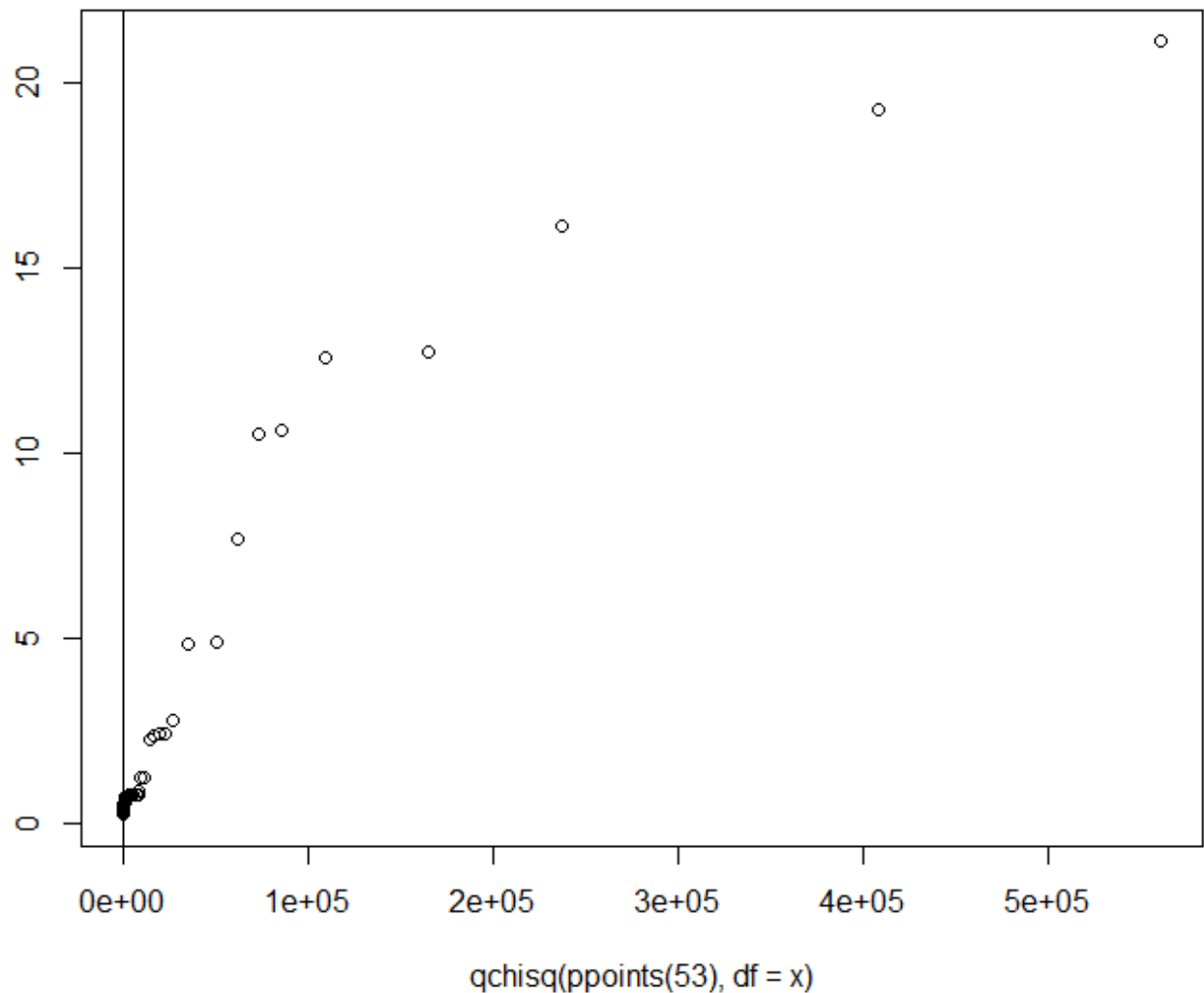
representing the entire data graphically:

##USING THE NORMAL Q-Q PLOT TO CHECK NORMALITY:

Normal Q-Q Plot (top-left)

Normal Q-Q Plot (top-right)

Normal Q-Q Plot (bottom-left)

Normal Q-Q Plot (bottom-right)

Since almost all the points are lying on the line, the data is normally distributed.

- TO CHECK FOR MULTIVARIATE NORMALITY:

qchisq(ppoints(53), df = x)

Since all the points lie on the straight line, it follows multivariate normal distribution.

- To check whether mean vector is significant or not we take the hypothesis :

Ho: There is no significant difference between the means.

H1: There is a significant difference between the means.

Doing the ANALYSIS;

Taking p=3 and n=53 , we get fcal= 0.2575362 and pval= 0.8556051.

The decision criteria is to reject Ho(null hypothesis) if pval < fcal then we do no reject the null hypothesis.

Hence here pval>fcal therefore we reject the null hypothesis and conclude that there is a significant difference between the means.