SESSION 8 – ASSIGNMENT 8.2

Date: 29th Jan 2019

library(RcmdrPlugin.IPSUR)
data(RcmdrTestDrive)
Perform the below operations: -

a. **Compute the measures of central tendency for salary and reduction which variable has highest center?**
b. **Which measure of center is more appropriate for before and after?**

```
#PART A
library(RcmdrPlugin.IPSUR)
x<- c(mean(RcmdrTestDrive$salary),median(RcmdrTestDrive$salary))
x
```

```
> x
  [1]  618.65  544.56  550.24  616.16  543.39  692.09  800.48  703.79  540.06  522.28  377.17  525.96  548.88  537.70  500.20  597.73  578.95  690.06  748.98
 [20]  811.71  660.58  586.29  387.59  524.54  536.87  503.64  496.09  701.91  595.70  759.30  717.91  808.63  682.60  623.09  550.28  646.25  635.43  437.19
 [39]  619.29  593.68  546.26  704.83  764.15  859.67  724.25  631.62  478.39  652.79  545.66  515.95  612.27  633.12  671.35  643.83  794.66  888.00  602.94
 [58]  716.78  606.12  704.90  620.32  515.92  655.72  619.44  640.48  844.32  918.03  933.49  699.63  593.27  634.24  686.98  618.68  631.20  608.88  686.28
 [77]  715.44  754.66  865.89  890.88  777.91  680.56  594.61  651.73  601.11  626.71  643.80  724.52  745.57  842.05  880.47 1016.21  726.13  780.21  704.08
 [96]  785.89  662.98  621.30  521.17  714.58  728.94  812.26  924.78 1001.31  724.99  822.35  653.58  642.28  730.12  708.30  629.17  790.33  788.05  849.25
[115] 1036.06 1149.92  854.31  768.94  666.74  639.72  744.38  584.08  712.00  789.76  719.06  903.34 1044.98 1027.36  855.36  796.51  771.74  780.27  808.65
[134]  632.05  681.58  823.38  754.55  938.47 1072.65 1021.69  785.75  882.78  762.43  863.78  745.97  809.26  668.26  780.61  749.43  889.55 1025.09 1156.16
[153]  777.93  835.96  668.69  870.52  827.18  689.23  662.17  820.52  780.51  980.09 1084.21 1073.50  908.11  793.42  804.78  790.82
```

```
#for reduction
y<- c(median(RcmdrTestDrive$reduction),mean(RcmdrTestDrive$reduction))
y
```

```
> y<- c(median(RcmdrTestDrive$reduction),mean(RcmdrTestDrive$reduction))
> y
[1] 139.500 223.631
```

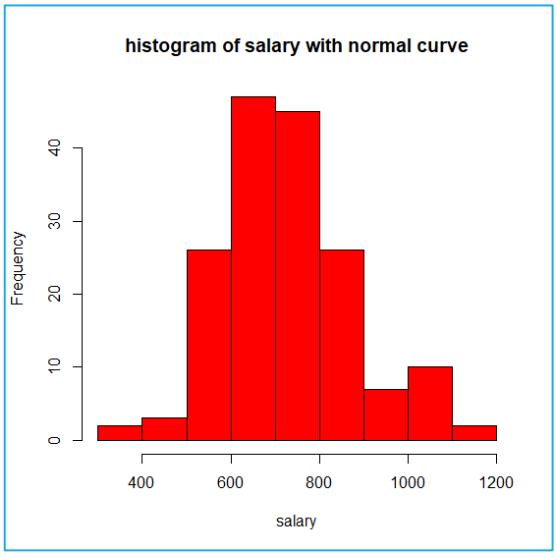
#now since we are looking for variable which has highest center we can check for this by plotting histogram or by checking kurtosis which describes the amount of peak of a distribution.

```
library(psych)
kurtosi(RcmdrTestDrive$salary)
kurtosi(RcmdrTestDrive$reduction)
```

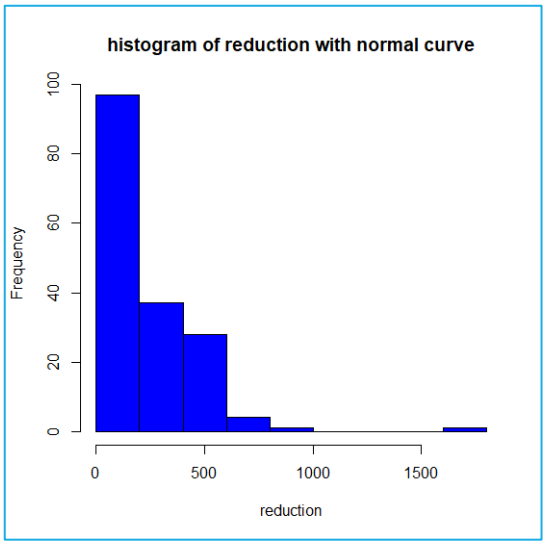
```
[1] 139.500 223.631
> library(psych)
> kurtosi(RcmdrTestDrive$salary)
[1] 0.2006576
> kurtosi(RcmdrTestDrive$reduction)
[1] 10.01655
>
```

#thus we can see variable reduction has more kurtosis thus more peaked hence more highest center or by plotting histogram we can also check that

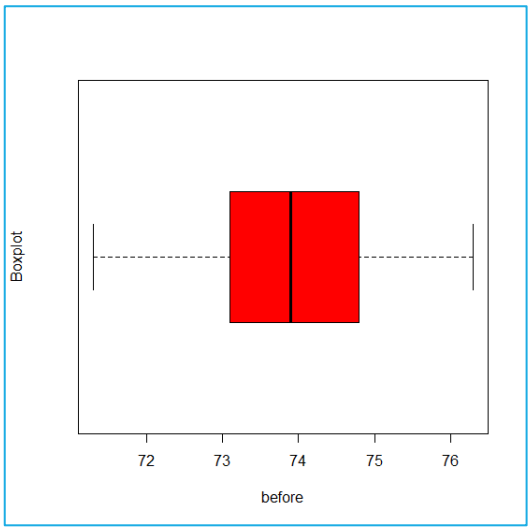
x<-RcmdrTestDrive$salary
h<- hist(x,breaks = 10,col = "red",xlab = "salary",main= "histogram of salary with normal curve")



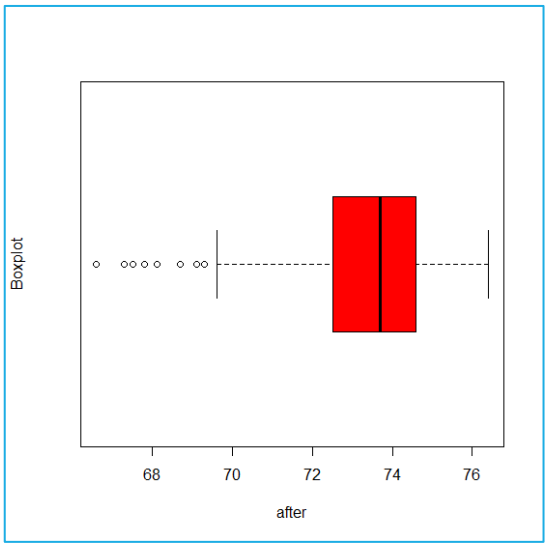
y<-RcmdrTestDrive$reduction
h<- hist(y,breaks = 10,col = "blue",xlab = "reduction",main= "histogram of reduction with normal curve")



#PART b
#by boxplot we can check for median where it lies
boxplot(RcmdrTestDrive$before,horizontal = T,col = "red",xlab="before",ylab="Boxplot")



#normal distributed
boxplot(RcmdrTestDrive$after,horizontal = T,col = "red",xlab="after",ylab="Boxplot")



#left skewed as the data is assymetrical distributed

#if we check the skewness of variables
skew (RcmdrTestDrive$before)
skew (RcmdrTestDrive$after)

```
> skew (RcmdrTestDrive$before)
[1] -0.03510369
> skew (RcmdrTestDrive$after)
[1] -1.164056
>
```

#after more negative so data more on right side as compare to before variable
#thus, the median would likely be a good choice and it is more appropriate