**Assignment 8.2**

**Question No.1:**

Compute the measures of central tendency for salary and reduction which variable has highest center?

> mean(RcmdrTestDrive$salary)

[1] 724.5164

> mean(RcmdrTestDrive$reduction)

[1] 223.631

> median(RcmdrTestDrive$salary)

[1] 710.15

> median(RcmdrTestDrive$reduction)

[1] 139.5

For mode we can initially check whether the length of unique column and the original column is the same. If the length of the unique column is lesser than the original column, then there is a need to proceed with mode, since there are some repetitions of values. In the case of salary the mode is indeterminate

> length(RcmdrTestDrive$salary)

[1] 168

> length(unique(RcmdrTestDrive$salary))

[1] 168

In the case of reduction the mode can be estimated since the length of unique column is lesser than the original column.

> length(RcmdrTestDrive$reduction)

[1] 168

> length(unique(RcmdrTestDrive$reduction, incomparables = FALSE))

[1] 132

> names(table(RcmdrTestDrive$reduction)[table(RcmdrTestDrive$reduction)==max(table(RcmdrTestDrive$reduction))]

+ )

[1] "116"

Rcmdr> numSummary(RcmdrTestDrive[,"salary", drop=FALSE], statistics=c("mean",

Rcmdr+ "quantiles"), quantiles=c(0.5))

**mean Median n**

**724.5164 710.15 168**

Rcmdr> binnedCounts(RcmdrTestDrive[,"salary", drop=FALSE])

distribution of salary

(300, 400] (400, 500] (500, 600] (600, 700] (700, 800] (800, 900]

2 3 26 47 45 26

(900, 1000] (1000, 1100] (1100, 1200]

7 10 2

Rcmdr> numSummary(RcmdrTestDrive[,"reduction", drop=FALSE], statistics=c("mean",

Rcmdr+ "quantiles"), quantiles=c(0.5))

**mean Median n**

**223.631 139.5 168**

Rcmdr> binnedCounts(RcmdrTestDrive[,"reduction", drop=FALSE])

distribution of reduction

(0, 200] (200, 400] (400, 600] (600, 800] (800, 1000] (1000, 1200]

97 37 28 4 1 0

(1200, 1400] (1400, 1600] (1600, 1800]

0 0 1

**Question No.2:**

Which of the measure of central tendency are appropriate for before and after data

Rcmdr> numSummary(RcmdrTestDrive[,c("after", "before"), drop=FALSE],

Rcmdr+ statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))

mean sd IQR 0% 25% 50% 75% 100% n

after 73.26726 1.884303 2.1 66.6 72.5 73.7 74.6 76.4 168

before 73.96607 1.076412 1.7 71.3 73.1 73.9 74.8 76.3 168

> names(table(RcmdrTestDrive$before)[table(RcmdrTestDrive$before)==max(table(RcmdrTestDrive$before))])

[1] "73.6"

> names(table(RcmdrTestDrive$after)[table(RcmdrTestDrive$after)==max(table(RcmdrTestDrive$after))])

[1] "72.5" "73.1" "73.8" "74.2" "74.3"

Mean Median Mode

After 73.27 73.7 Multi Modal

Before 73.97 73.9 73.6

Therefore the Mean and Median could be better measures of central tendencies for before and after variable. Since both of them appear to be numeric mean and median would definitely be meaningful.