OUTPUT

1. #mean, median and standard deviation of variables

mean(age)

mean(gmat\_tot)

mean(salary)

mean(work\_yrs)

median(salary)

median(age)

median(gmat\_tot)

mean(s\_avg)

mean(f\_avg)

o/p :

> mean(age)

[1] 27.35766

> mean(gmat\_tot)

[1] 619.4526

> mean(salary)

[1] 39025.69

> mean(work\_yrs)

[1] 3.872263

> median(salary)

[1] 999

> median(age)

[1] 27

> median(gmat\_tot)

[1] 620

> mean(s\_avg)

[1] 3.025401

> mean(f\_avg)

[1] 3.061533

2)#Draw scatter plots to find relationships between variables

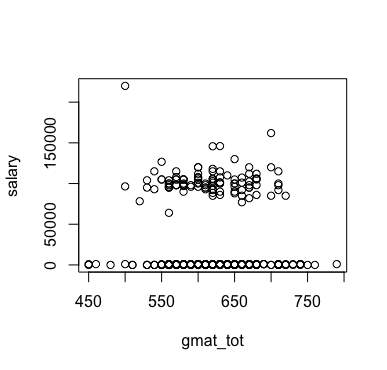
no\_data.df <- subset(MBA.df, satis!=998 & satis!=999)

mean(no\_data.df$satis)

o/p: [1] 5.565789

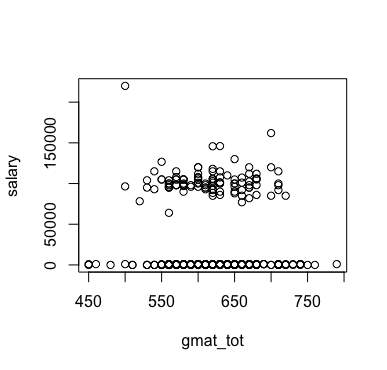
plot(gmat\_tot, salary)

o/p:



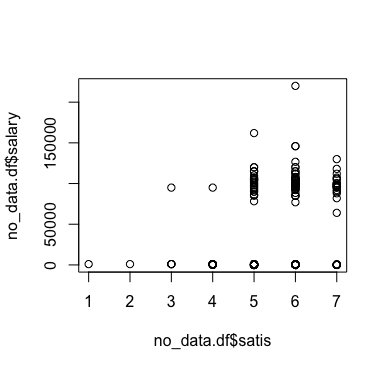
plot(work\_yrs, salary)

o/p:



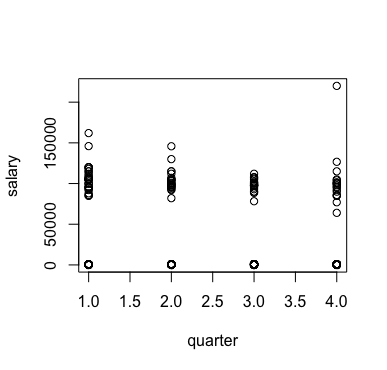
plot(no\_data.df$satis, no\_data.df$salary)

o/p:



plot(quarter, salary)

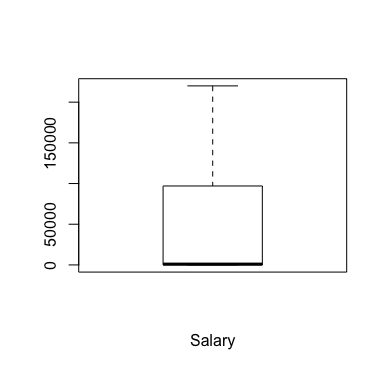
o/p:



3)#draw boxplots

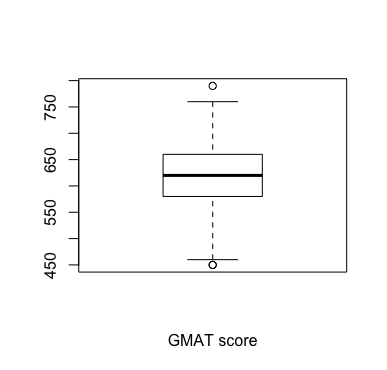
boxplot(salary, xlab='Salary')

o/p:



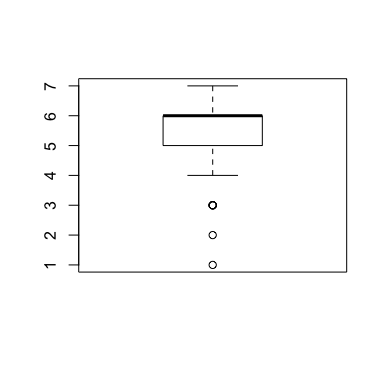
boxplot(gmat\_tot, xlab='GMAT score')

o/p:



boxplot(no\_data.df$satis)

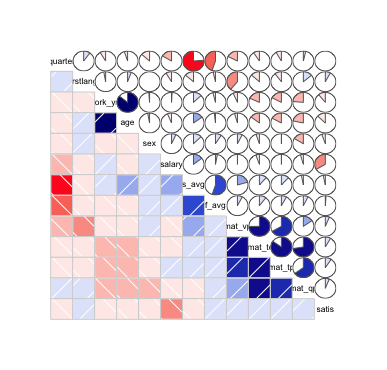
o/p:



4)#corrogram

library(corrgram)

corrgram(MBA.df, order=TRUE, upper.panel=panel.pie, text.panel=panel.txt)



5)# subset of the dataset consisting of only those people who actually got a job

job <- subset(MBA.df, salary>0 & salary!=998 & salary!=999 & work\_yrs>0)

View(job)

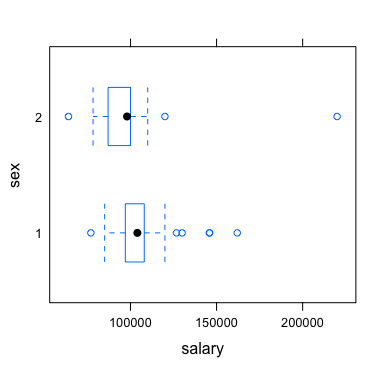
mean(job$age)

ans) 26.80392

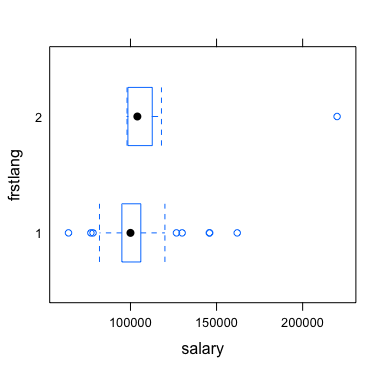
6)#Bwplot

library(lattice)

bwplot(sex~salary, data = job, horizontal = TRUE)



bwplot(frstlang~salary, data=job, horizontal = TRUE)



7)#Run chi square test

a) mytable<-xtabs(~sex+salary, data = job)

addmargins(mytable)

chisq.test(mytable)

o/p:

Pearson's Chi-squared test

data: mytable

X-squared = 52.971, df = 41, p-value = 0.09963

b) mytable<-xtabs(~gmat\_tot+salary, data = job)

addmargins(mytable)

chisq.test(mytable)

o/p:

Pearson's Chi-squared test

data: mytable

X-squared = 923.63, df = 820, p-value = 0.006665

c) mytable<-xtabs(~frstlang+salary, data = job)

addmargins(mytable)

chisq.test(mytable)

o/p:

Pearson's Chi-squared test

data: mytable

X-squared = 69.145, df = 41, p-value = 0.003885

d) mytable<-xtabs(~work\_yrs+salary, data = job)

addmargins(mytable)

chisq.test(mytable)

o/p:

Pearson's Chi-squared test

data: mytable

X-squared = 516.83, df = 410, p-value = 0.0002591

8)#Fitting regression models

a)model1 = lm(salary~., data = job)

summary(model1)

o/p:

Residual standard error: 15510 on 89 degrees of freedom

Multiple R-squared: 0.3414, Adjusted R-squared: 0.2526

F-statistic: 3.845 on 12 and 89 DF, p-value: 9.89e-05

b)model2 = lm(salary~work\_yrs+gmat\_tpc+age+sex+s\_avg+frstlang, data = job)

summary(model2)

o/p:

Residual standard error: 15640 on 95 degrees of freedom

Multiple R-squared: 0.2846, Adjusted R-squared: 0.2394

F-statistic: 6.299 on 6 and 95 DF, p-value: 1.33e-05

c)model3 = lm(salary~frstlang+gmat\_tot+gmat\_tpc+age+sex, data = job)

summary(model3)

o/p:

Residual standard error: 15500 on 96 degrees of freedom

Multiple R-squared: 0.2907, Adjusted R-squared: 0.2538

F-statistic: 7.87 on 5 and 96 DF, p-value: 3.029e-06

d)model = lm(salary~gmat\_tot+s\_avg+gmat\_tpc+quarter+work\_yrs+satis, data = job)

summary(model)

o/p:

Residual standard error: 16120 on 95 degrees of freedom

Multiple R-squared: 0.2406, Adjusted R-squared: 0.1926

F-statistic: 5.016 on 6 and 95 DF, p-value: 0.0001663

9) #comparison between those who got placed and those who didn't

nojob <- subset(MBA.df, MBA.df$salary==0 )

summary(nojob)

mean(nojob$age)

o/p: 28.51111

mean(nojob$work\_yrs)

o/p:4.588889