**CLASS GRADES**

#Loading the dataset

data3<-read.csv(file.choose())

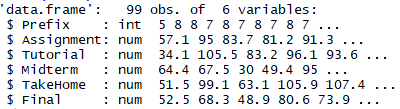
colnames(data3)

"Prefix" "Assignment" "Tutorial" "Midterm" "TakeHome" "Final"

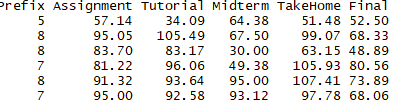
nrow(data3)

99

str(data3)



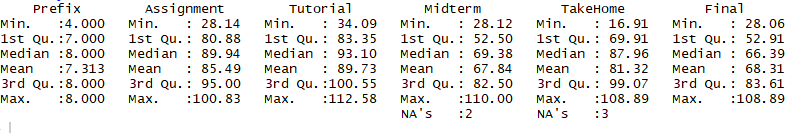
head(3)



sum(is.na(data3))

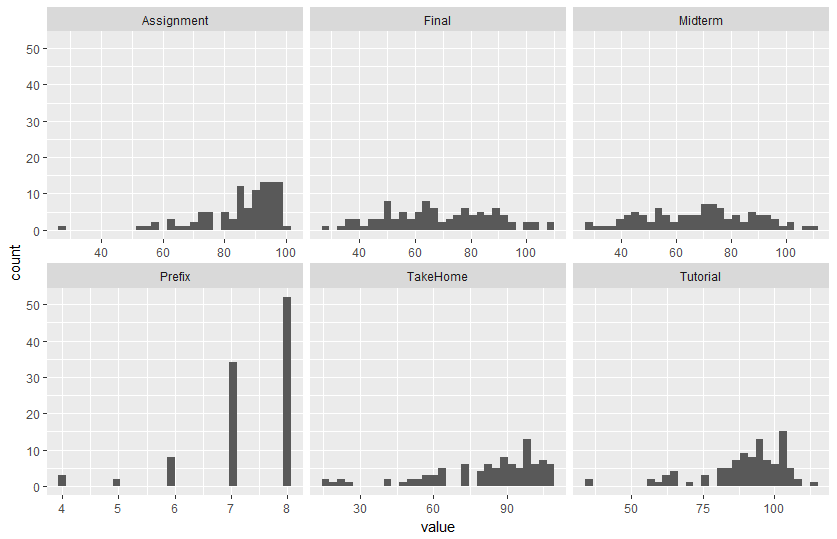
***5 missing values – 2 missing in Mid-Term and 3 missing in Take Home***

summary(data3)

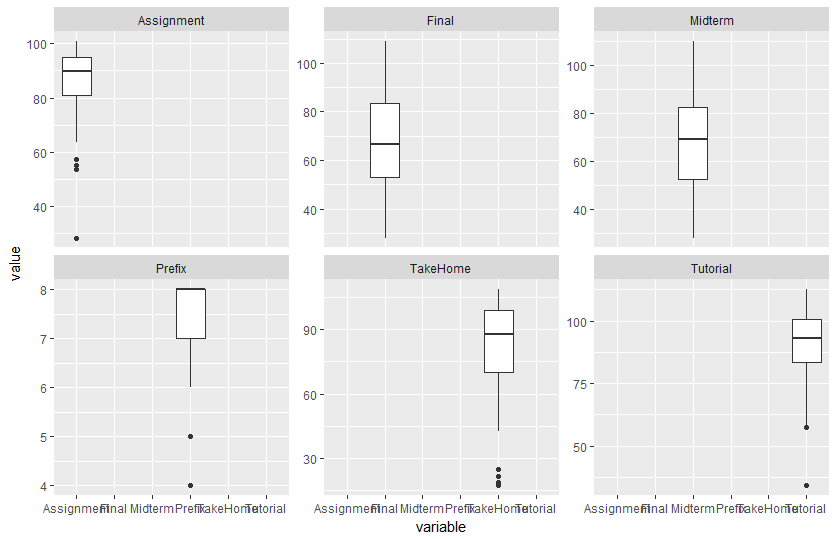


data3 %>% gather(Prefix:Final, key = "variable", value = "value") %>% ggplot(aes(x = value)) +

geom\_histogram(bins = 30) + facet\_wrap(~ variable, scales = 'free\_x')

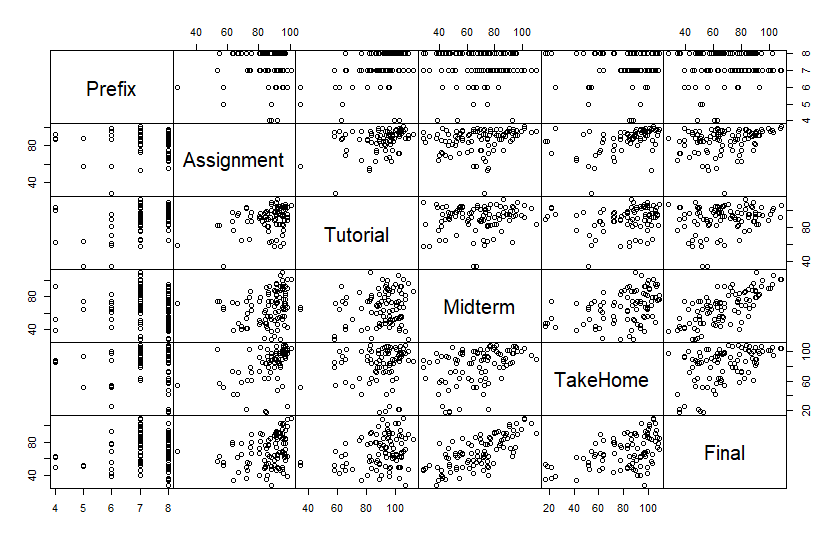


data3 %>% gather(Prefix:Final, key = "variable", value = "value") %>% ggplot(aes(x=variable,y = value)) + geom\_boxplot() + facet\_wrap(~ variable, scales = 'free\_y')



***Assignment, Take Home and Tutorial are left skewed with outliers.***

pairs(data3, gap = 0, pch = 21)

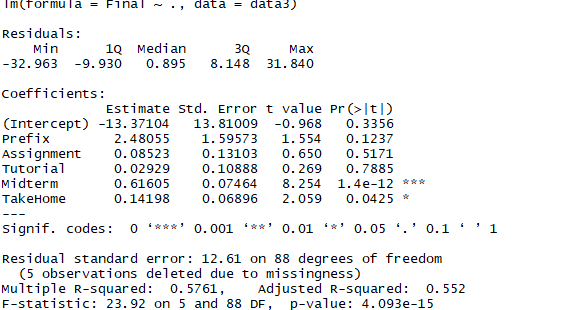


**Regression Imputation technique**

First we shall create a linear regression model.

linear\_model = lm(Final~.,data=data3)

summary(linear\_model)



The summary shows that Midterm and TakeHome are significant variables.

Now, we shall regress Take Home on all the remaining attributes

TakeHome\_all <- lm(TakeHome~Prefix+Assignment+Tutorial+Midterm,data=data3)

first\_df<- data.frame(Prefix=c(7,6,8),Assignment=c(72.85,90.74,71.79),Tutorial=c(86.85,89.64,102.87),Midterm=c(60,61.25,41.88))

#Predicting the missing values

TakeHome\_all.pred = predict(TakeHome\_all,first\_df)

TakeHome\_all.pred



Hence, we have the missing values in Take Home predicted as below:

data3$TakeHome[8]=68.64

data3$TakeHome[39]=84.93

data3$TakeHome[61]=59.40

Now, we shall regress Mid Term on all the remaining attributes

MidTerm\_all<-lm(Midterm~Prefix+Assignment+Tutorial+TakeHome,data=data3)

second\_df<-data.frame(Prefix=c(8,8),Assignment=c(63.4,82.45),Tutorial=c(86.21,86.65))

MidTerm\_all.pred = predict(MidTerm\_all,second\_df)

MidTerm\_all.pred



Hence, we have the missing values in MidTerm predicted as below:

data3$Midterm[21]=59.43

data3$Midterm[27]=64.58

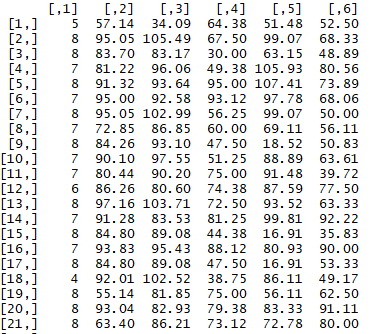
**HotDeck Imputation technique**

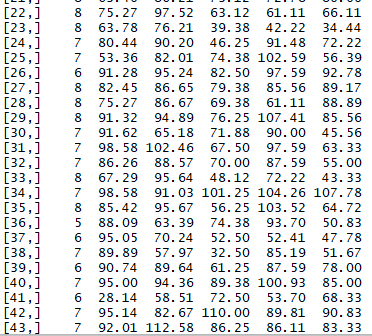
library(HotDeckImputation)

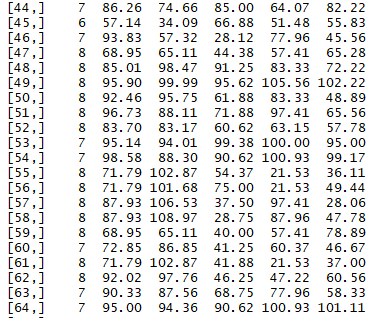
data3\_matrix = as.matrix(data3)

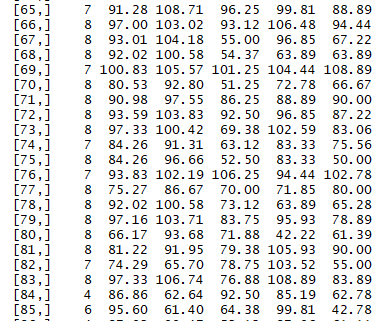
fitHD= impute.NN\_HD(DATA=data3\_matrix,distance="man")

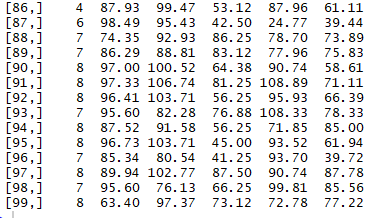
fitHD











**Table of the imputed values from both the techniques**

We can write the Imputed values from HotDeck Imputation into a csv file

*write.csv(fitHD, file = "Imputed.csv")*

*For Saving it in Tabular Form, we can write as:*

*as.table(fitHD)*

*and*

*as.table(MidTerm\_all.pred,TakeHome\_all.pred) # From Regression Imputation*