**# Set the Working Directory**

setwd("D:/Studies/VIII Semester/R/Programs/Ex 4")

**#Get the Working Directory**

getwd()

**#Read the Dataset of CSV Format**

data\_reduction <- read.csv("D:/Studies/VIII Semester/R/Programs/Ex 4/congenital\_data\_2011.csv", header=TRUE)

View(data\_reduction)

**#DIMENSIONALITY REDUCTION**

**#Principal Components Analysis**

**# load data**

data\_reduction=read.csv("congenital\_data\_2011.csv")

**# applying principal component analysis on dataset**

pca = prcomp(data\_reduction)

**# plot to show variable importance**

par(mar = rep(2, 4))

**# To know the values of pca**

pca

**# To plot the values**

plot(pca)

**# plot pca components using biplot in r**

biplot (pca , scale =0)

**# Rotating pca components**

pca$rotation=-pca$rotation

pca$x=-pca$x

**# plot pca components using biplot in r after rotating**

biplot (pca , scale =0)

**#NUMEROSITY REDUCTION**

**#Parametric Methods**

**#Regression(Linear Regression)**

**# To assign the field to a variable**

x<-data\_reduction$Paediatric

y<-data\_reduction$AlivePaed30d

**# create the relationship model using lm()**

relation<-lm(y~x)

relation

**# get the summary of the relationship**

print (summary(relation))

**# To plot a value by using x,y**

plot(x,y)

**# x is the predictor variable and y is the response variable**

a<-data.frame(x=100)

result<-predict(relation,a)

print(result)

**# visualizing the regression graphically**

abline (relation)

**#Regression(Multiple Regression)**

**#selecting three columns**

x=data\_reduction[,1:4]

#assinging attributes data to frame

y=x$Procedures

x1=x$Paediatric

x2=x$AlivePaed30d

x3= x$DeadPaed30d

**#using formula of regression**

fit <- lm( y~ x1+ x2+ x3)

print(summary(fit**)) # show results**

plot(fit)

**#prediction of future**

newdata=data.frame(x1=150,x2=80,x3=70)

m=predict(fit,newdata)

print(paste("Paediatric",newdata$x1,"AlivePaed30d",newdata$x2,"DeadPaed30d",newdata$x3,"in the Procedures be",m ))

**#Non-Parametric Methods**

**#Histogram**

x<-data\_reduction$Procedure

**#Distributing Buckets equally**

x2<-split(x,cut(x,5))

x2

x1<-tapply(x,cut(x,5),median)

x1

hist(data\_reduction$Procedure,main="Before reduction",xlab="mean width",ylab="mean frequency",label="bucket",col="blue")

hist(x1,main="Before reduction",xlab="mean width",ylab="mean frequency",label="bucket",col="blue")

**#Clustering**

plot(data\_reduction$Procedure,data\_reduction$Paediatric)

**#Normalization**

z<-data\_reduction[,-c(1,1)]

m<-apply(z,2,mean)

s<-apply(z,2,sd)

z<-scale(z,m,s)

**#Calculate Euclidean distance**

distance<-dist(z)

print(distance,digits = 3)

**#Cluster Dendrogram with complete linkage**

hc.c<-hclust(distance)

plot(hc.c,hang=-1)

**#Cluster Dendrogram with Average linkage**

hc.a<-hclust(distance,method="average")

plot(hc.a,hang=-1)

**OUTPUT:**

**#Get the Working Directory**

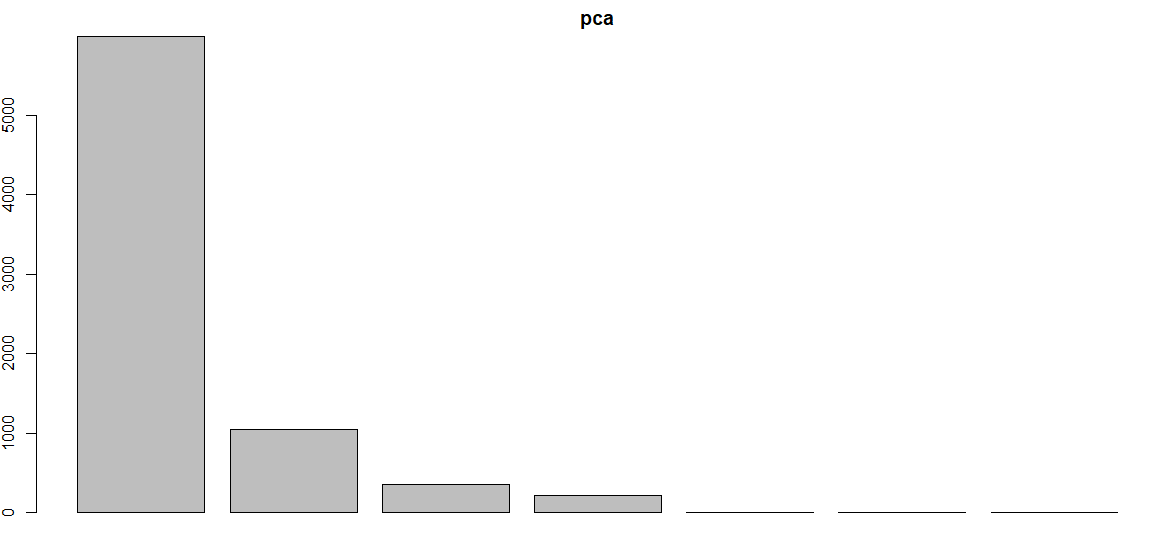
[1] "D:/Studies/VIII Semester/R/Programs/Ex 4"

**#DIMENSIONALITY REDUCTION**

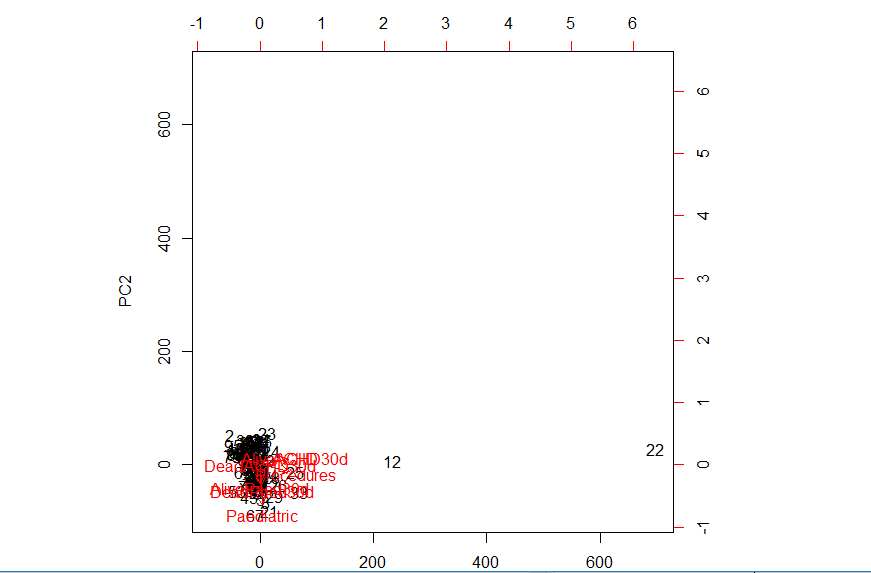
**#Principal Components Analysis**

|  |
| --- |
| Standard deviations (1, .., p=7):  [1] 7.742211e+01 3.246024e+01 1.886100e+01 1.482391e+01 1.789192e+00 3.642696e-01  [7] 9.118266e-15  Rotation (n x k) = (7 x 7):  PC1 PC2 PC3 PC4 PC5 PC6  Procedures 0.593557373 -0.140250652 0.792369933 -0.012836801 0.0021066418 -0.0002614091  Paediatric 0.041903850 -0.800510076 -0.175066955 -0.028884307 0.5709157395 0.0004020250  AlivePaed3 0d 0.004066683 -0.367435039 -0.078309081 -0.725808913 -0.5762321278 0.0007738938  DeadPaed30d 0.038132369 -0.420137329 -0.090246438 0.686933837 -0.5848133552 -0.0011707404  ACHD 0.570618144 0.118900881 -0.406746884 -0.013315281 -0.0002816899 -0.4016264141  AliveACHD30d 0.564643547 0.117688573 -0.402195193 -0.012056414 -0.0006585077 0.4148336624  DeadACHD30d 0.005974598 0.001212308 -0.004551691 -0.001258867 0.0003768179 -0.8164600765  PC7  Procedures 9.379496e-17  Paediatric -3.106502e-16  AlivePaed30d 4.250357e-16  DeadPaed30d 2.692870e-16  ACHD 5.773503e-01  AliveACHD30d -5.773503e-01  DeadACHD30d -5.773503e-01 |
|  |
| |  | | --- | |  | |

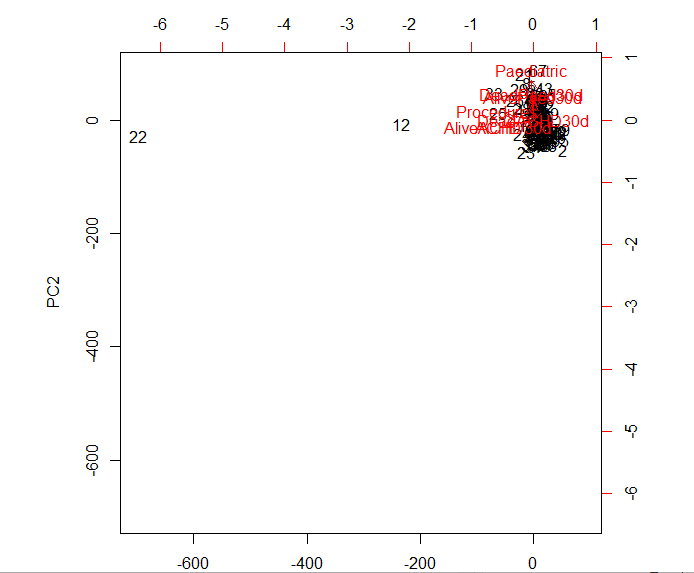
**# To plot the values**



**# plot pca components using biplot in r**



**# plot pca components using biplot in r after rotating**



**#NUMEROSITY REDUCTION**

**#Parametric Methods**

**#Regression(Linear Regression)**

**# create the relationship model using lm()**

Call:

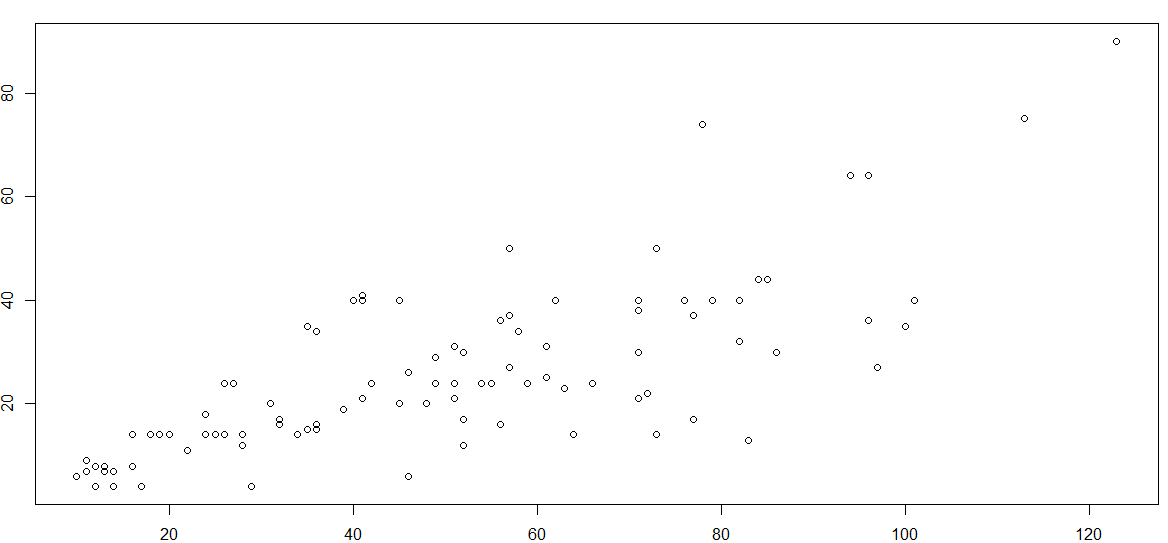
lm(formula = y ~ x)

Coefficients:

(Intercept) x

2.7918 0.4576

**# To plot a value by using x,y**

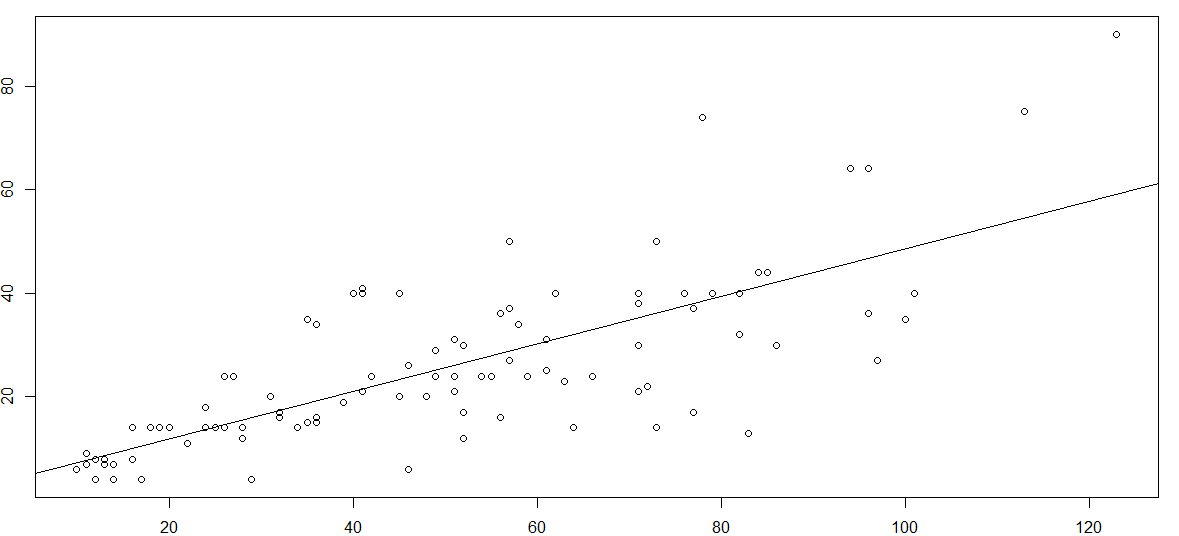


**# Result**

1

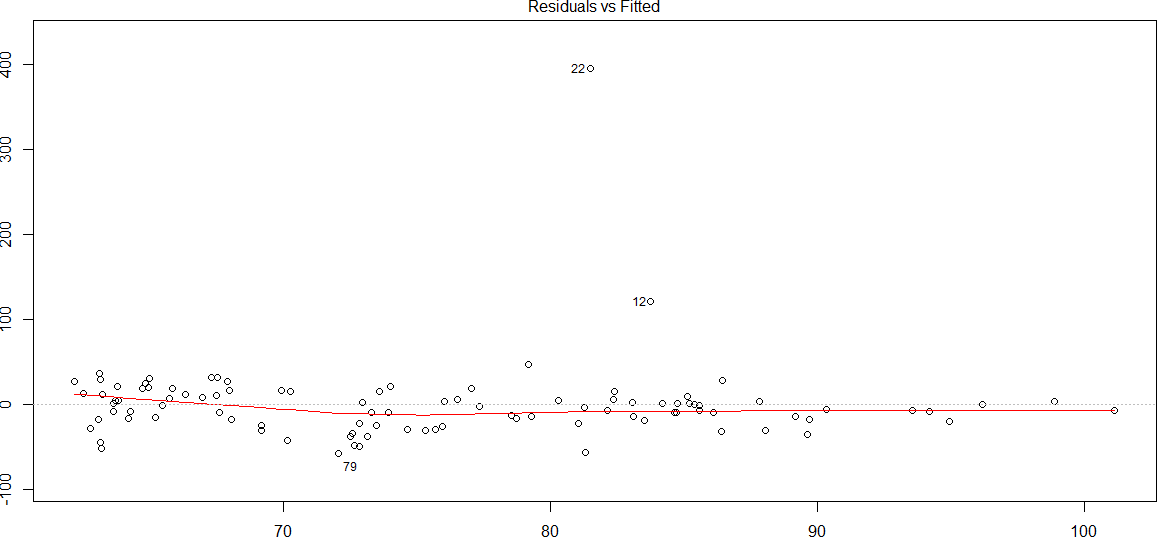
48.55627

**# visualizing the regression graphically**

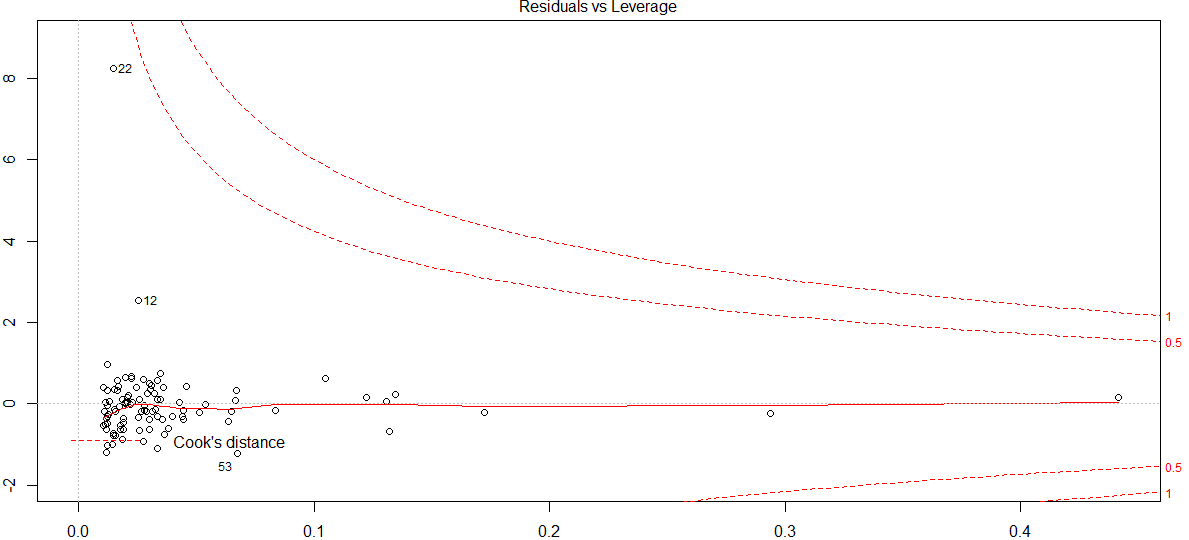


**#Regression(Multiple Regression)**

**plot**



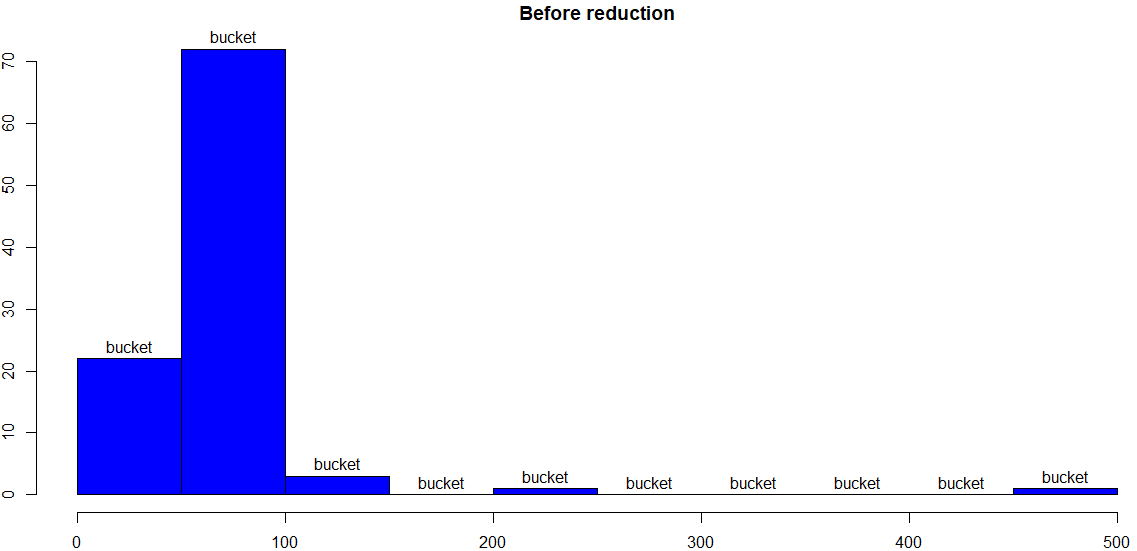
**#prediction of future**



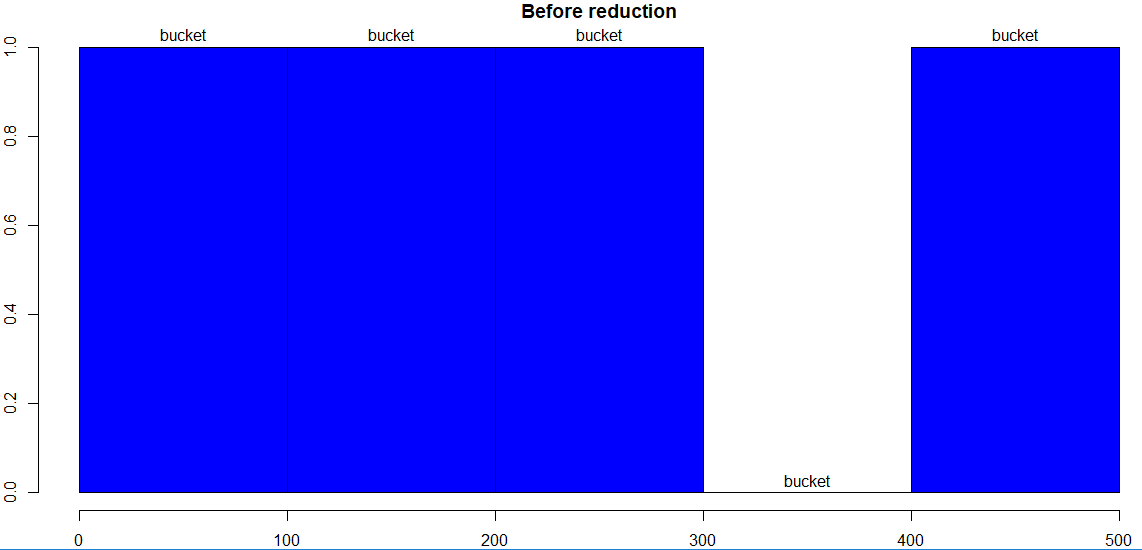
**#Non-Parametric Methods**

**#Histogram**

**Before**

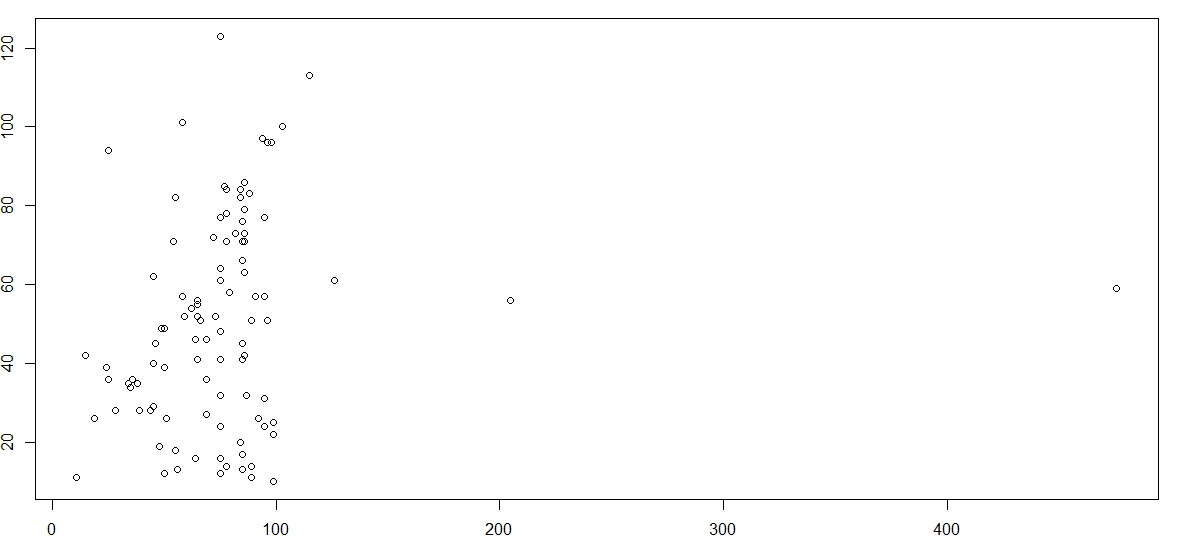


**After**

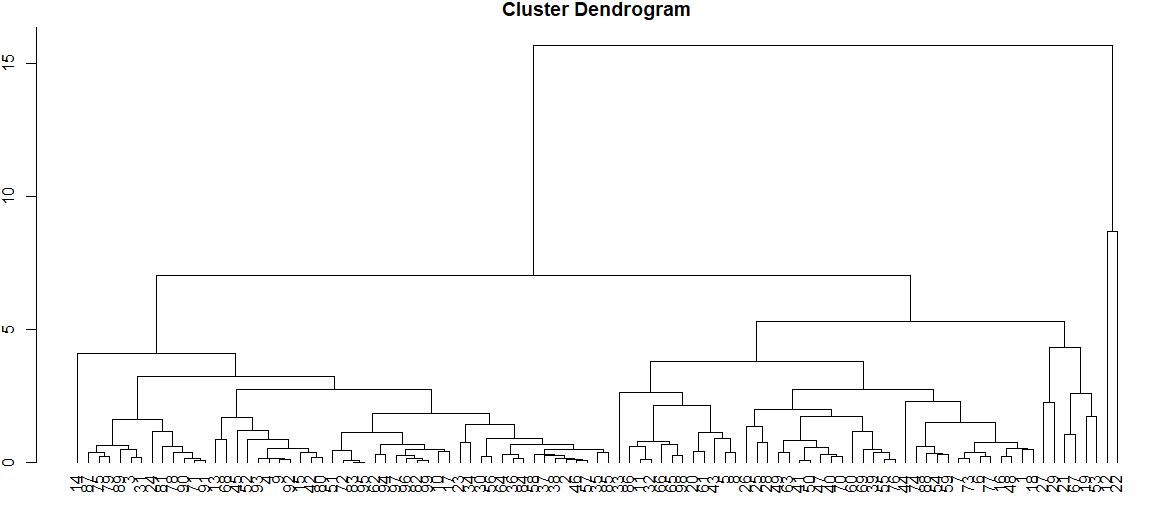
****

**#Clustering**

**Plot**

****

**#Cluster Dendrogram with complete linkage**



**#Cluster Dendrogram with Average linkage**

