**Assignment on Clustering**

1. Perform Clustering for the crime data and identify the number of clusters formed and draw inferences.

Data Description:

Murder -- Muder rates in different places of United States

Assualt- Assualt rate in different places of United States

UrbanPop - urban population in different places of United States

Rape - Rape rate in different places of United States

Solution:

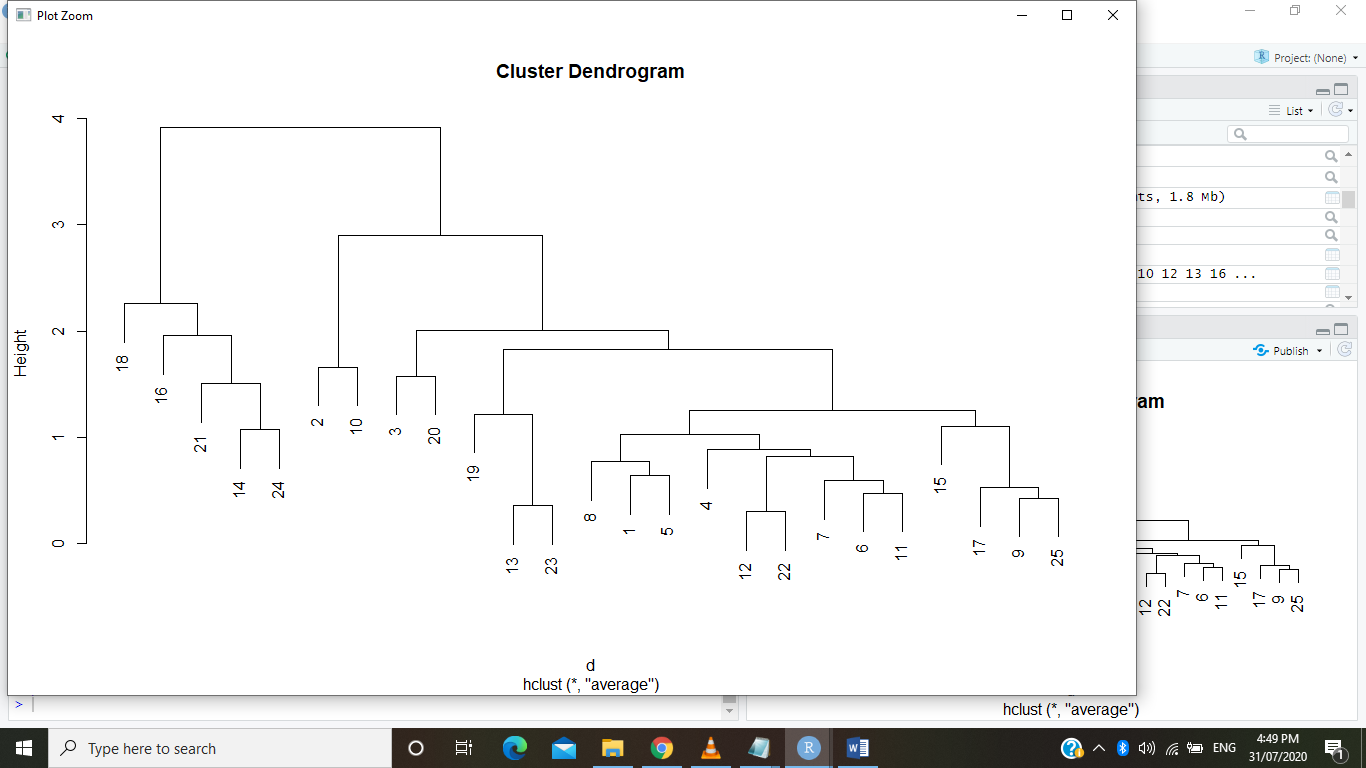
mydata<-crime\_data

mydata1<-scale(mydata[,2:5])

d<-dist(mydata1,method = "euclidean")

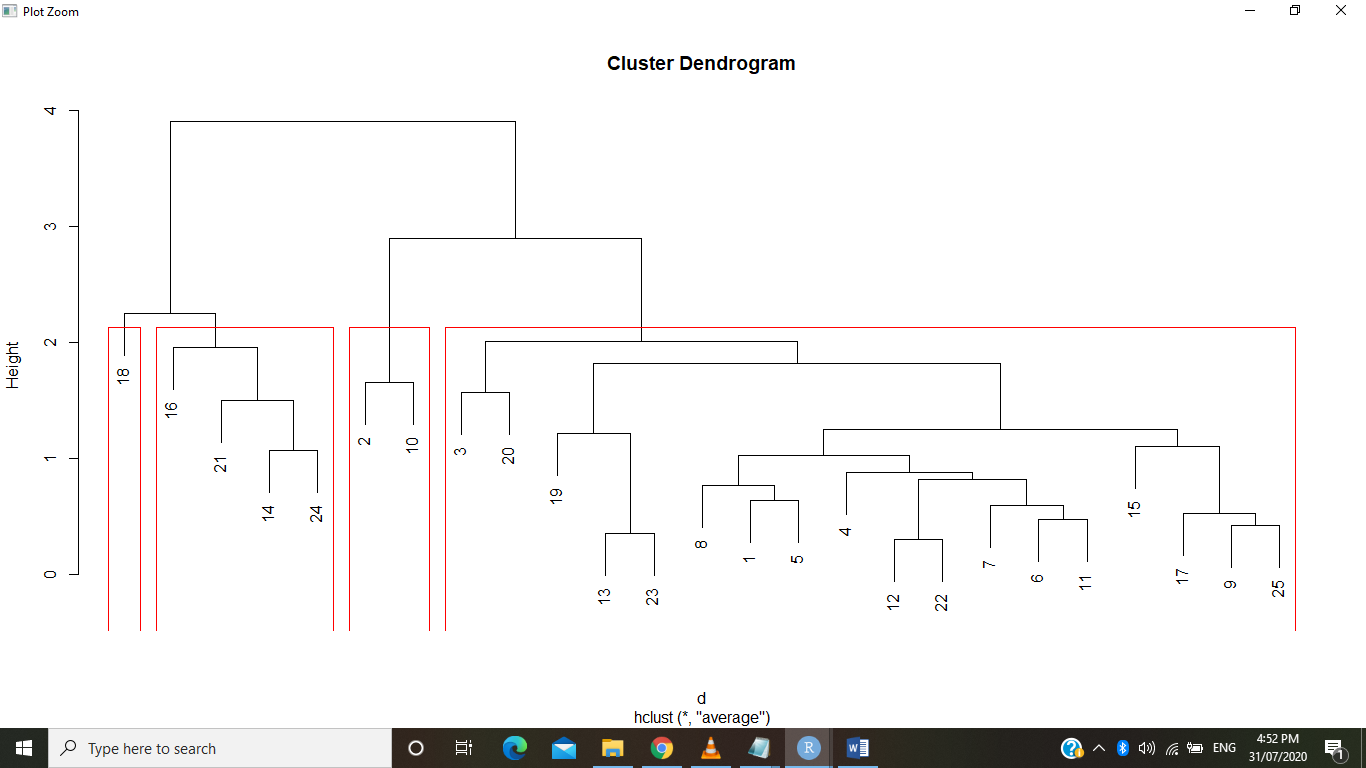
fit<-hclust(d,method = "average")

plot(fit)



groups<-cutree(fit,k=4)

rect.hclust(fit,k=4,border ="red" )



clusters=data.frame('crime'=mydata[,1],'cluster'=groups)

clusters

crime cluster

1 0.40199420 1

2 1.37098850 2

3 -0.05943165 1

4 0.40199420 1

5 0.12513869 1

6 0.67884972 1

7 0.44813679 1

8 -0.10557424 1

9 1.23256074 1

10 0.35585162 2

11 1.04799040 1

12 -0.05943165 1

13 -0.10557424 1

14 -1.71133621 3

15 1.00184782 1

16 -2.41270351 3

17 0.86342006 1

18 -1.76670731 4

19 -0.24400199 1

20 0.21742386 1

21 -0.79771302 3

22 0.17128128 1

23 -0.38242975 1

24 -1.67442214 3

25 1.00184782 1

Inferences: Data reduced to 4 clusters since number of clusters are formed more than 20.

1. Perform clustering (Both hierarchical and K means clustering) for the airlines data to obtain optimum number of clusters. Draw the inferences from the clusters obtained.

Soln:

1. H-Clustering:

#Hclustering

airlinedata<-EastWestAirlines.New

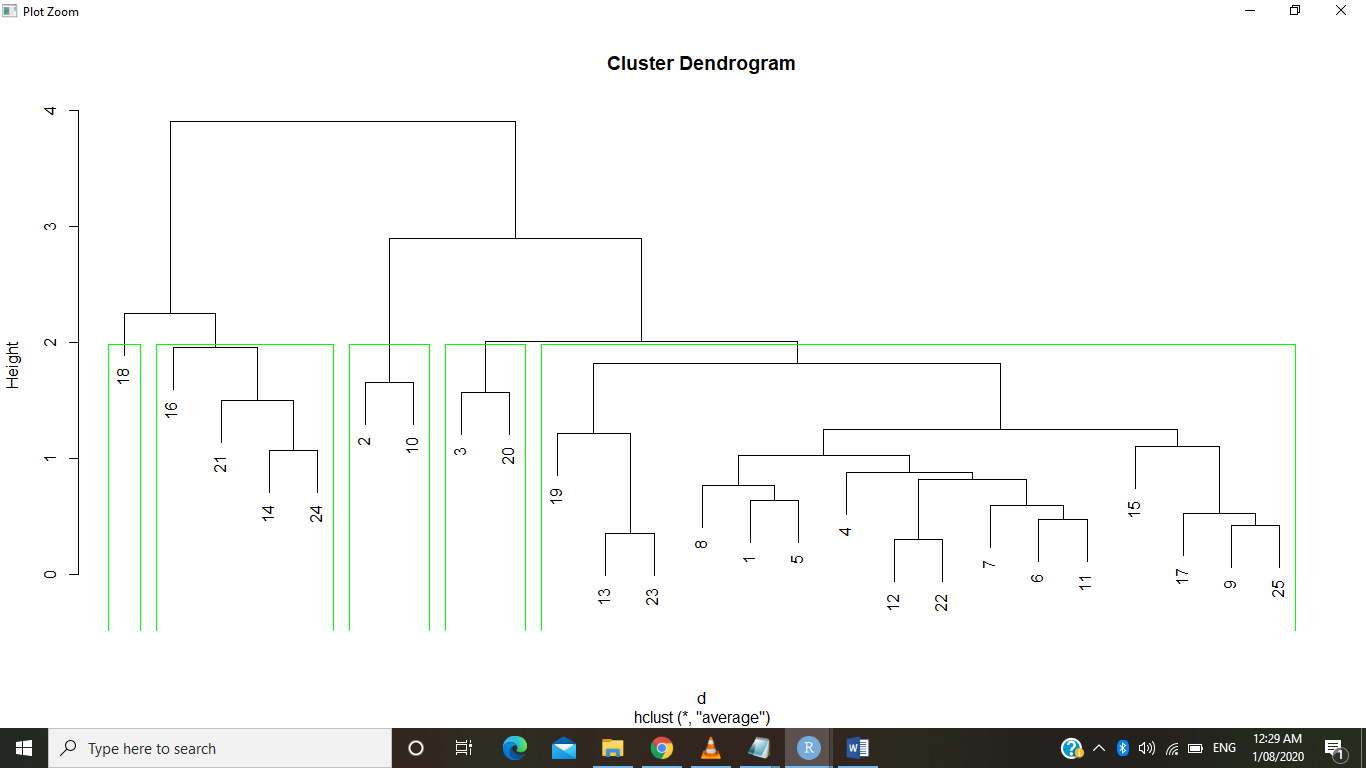
d<-dist(airlinedata,method = "euclidean")

fit<-hclust(d,method = "average")

plot(fit)

groups<-cutree(fit,k=5)

rect.hclust(fit,k=5,border ="green" )



clusters=data.frame('Airline'=airlinedata[,1],'cluster'=groups)

clusters

500 columns displayed

[ reached 'max' / getOption("max.print") -- omitted 3500 rows ]

K-Means clustering:

data<-cbind(EastWestAirlines.New.1)

plot(data)



wss<-c()

for(i in 2:9)wss[i]<-sum(kmeans(data,centers = i)$withinss)

#plot(2:9,wss,type = "b",xlab = "No of Clusters",ylab = "Avg Distance")

km<-kmeans(data,10)

km$centers

km$cluster

library(animation)

windows()

km<-kmeans.ani(data,3)

