K-Means Clustering

install.packages('factoextra')

library(factoextra)

install.packages("readxl")

library("readxl")

airline<-read.csv("EastWestAirlines.csv")

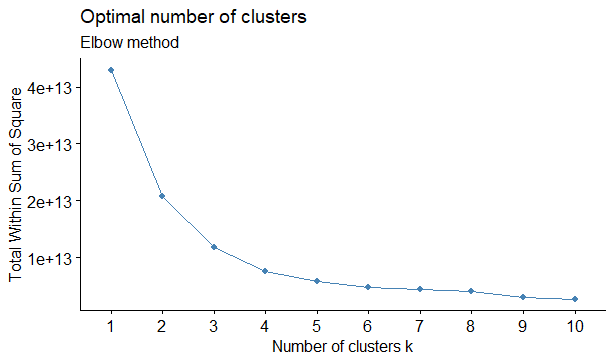
install.packages("fviz\_nbclust")

# Elbow method

summary(airline)

fviz\_nbclust(airline[,-1], kmeans, method = "wss") +

labs(subtitle = "Elbow method")



km <- kmeans(airline[,-1],4)

km$centers

km$cluster

Balance Qual\_miles cc1\_miles cc2\_miles cc3\_miles Bonus\_miles Bonus\_trans

1 30832.31 101.9886 1.665714 1.016429 1.005000 10077.51 9.314643

2 323009.85 337.2486 3.145946 1.016216 1.000000 43644.12 18.583784

3 127878.64 217.6366 2.938259 1.009109 1.032389 31142.03 16.483806

4 842320.23 512.6923 3.346154 1.000000 1.115385 57797.19 22.730769

Flight\_miles\_12mo Flight\_trans\_12 Days\_since\_enroll Award.

1 290.8486 0.8846429 3785.444 0.3239286

2 1283.3351 3.8810811 5532.584 0.6270270

3 748.9646 2.1427126 4736.144 0.4412955

4 1845.8846 6.9615385 6463.000 0.8461538

> km$cluster

[1] 1 1 1 1 3 1 3 1 2 3 1 3 1 1 1 1 1 1 3 1 3 3 1 1 3 1 1 1 1 1 3 1 2 1 1 3 1 1 1

[40] 1 1 1 1 4 3 2 3 1 1 1 3 1 3 3 1 1 1 1 1 3 3 1 1 2 1 1 1 3 2 1 3 2 2 1 3 3 3 3

[79] 1 1 1 3 1 1 1 1 1 4 1 1 1 1 3 3 3 1 1 3 1 1 1 1 1 1 1 1 3 1 3 1 3 2 3 1 2 1 3

[118] 4 3 1 1 3 3 1 3 3 4 3 1 3 1 1 3 3 3 2 3 3 3 1 1 1 1 3 1 3 1 1 3 1 3 4 2 3 1 3

[157] 3 1 2 1 1 3 3 1 1 1 3 2 2 1 3 1 2 3 3 2 3 1 1 1 1 3 1 1 1 3 4 3 2 1 3 3 2 1 1

[196] 3 1 3 1 3 1 1 1 3 1 1 3 1 1 1 3 1 3 1 1 3 1 1 3 3 4 2 1 1 3 3 3 3 1 3 3 3 1 1

[235] 1 3 1 2 1 1 3 1 2 3 2 1 3 1 3 3 1 1 3 1 3 2 2 1 3 2 3 1 3 1 1 3 1 1 1 3 3 3 3

[274] 1 1 4 1 3 1 1 3 1 3 3 3 3 1 1 3 1 2 3 2 1 1 3 1 2 3 1 1 1 3 2 3 3 3 2 3 1 3 1

[313] 1 4 3 3 1 1 4 1 3 3 1 2 1 2 1 2 3 1 3 1 3 1 3 2 3 1 1 1 3 1 1 3 1 1 3 3 3 3 1

[352] 3 2 1 1 1 3 3 1 3 1 1 3 1 1 1 1 1 1 3 3 3 3 3 1 1 1 3 1 3 2 3 3 2 1 1 1 1 3 3

[391] 3 1 1 1 3 1 3 1 1 1 3 1 3 1 2 3 1 2 1 1 4 1 3 2 1 3 1 1 3 3 2 1 3 1 3 3 1 2 3

[430] 3 2 1 1 3 1 1 3 2 1 1 1 3 1 1 3 3 2 1 2 1 3 1 3 3 1 1 1 3 1 1 1 3 1 1 3 2 4 2

[469] 3 1 1 4 1 1 1 3 2 3 2 3 1 4 1 3 1 1 3 3 4 3 1 1 1 1 3 1 1 1 1 3 3 3 1 3 1 3 3

[508] 1 3 1 4 1 3 1 1 2 1 1 3 1 1 3 1 1 3 1 1 1 1 1 2 3 3 1 2 1 1 1 1 1 1 1 1 1 1 3

[547] 3 1 1 2 3 1 1 1 3 1 1 3 1 1 1 2 3 1 3 3 1 1 1 1 1 1 3 1 3 1 1 1 1 1 3 1 1 1 1

[586] 1 1 3 1 3 3 1 1 1 1 1 1 1 1 2 3 1 3 3 3 3 1 1 2 1 1 3 1 1 1 3 3 3 2 1 1 3 1 3

[625] 2 3 1 1 3 4 2 1 1 1 1 1 1 1 3 3 1 1 1 1 1 1 1 1 1 2 3 3 2 1 1 1 1 3 2 1 1 1 3

[664] 3 1 3 1 1 1 3 1 1 3 1 3 1 1 1 2 3 1 3 2 2 1 3 1 1 3 3 1 3 2 2 1 1 1 2 1 3 1 2

[703] 2 1 2 3 1 1 2 1 1 1 1 1 3 1 3 1 1 1 1 1 1 2 1 3 1 1 1 1 3 1 1 1 1 3 3 1 1 3 3

[742] 1 1 2 1 2 1 3 1 1 3 3 3 1 3 1 2 3 1 1 3 1 1 3 1 2 1 1 1 3 1 3 3 1 1 1 1 2 1 1

[781] 3 2 1 1 1 3 1 2 3 3 1 3 3 3 1 1 1 1 3 1 3 1 1 1 1 1 3 3 1 1 1 1 1 3 3 1 1 1 1

[820] 1 2 3 3 4 1 1 3 1 3 1 1 1 2 3 1 1 3 3 1 3 3 3 3 1 3 1 1 3 1 1 1 1 1 3 1 3 1 1

[859] 3 3 3 3 1 3 1 1 1 1 3 2 1 1 1 1 1 1 1 1 1 1 2 2 1 3 3 1 1 3 2 3 3 3 3 1 1 1 3

[898] 1 1 3 3 3 3 3 1 2 1 1 1 3 1 2 1 1 3 1 1 3 2 1 1 1 3 3 1 1 1 1 3 1 3 1 2 1 3 3

[937] 3 1 1 3 1 1 1 1 1 3 3 2 1 1 1 3 1 3 1 1 1 1 1 3 1 1 1 2 1 3 1 1 3 1 1 1 3 1 1

[976] 1 1 3 1 3 3 3 1 3 2 1 1 3 3 1 1 1 3 1 1 3 1 2 3 2

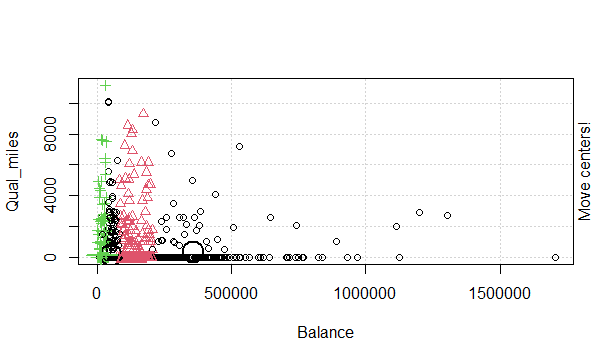
clust<-data.frame("Airlines"=airline[,1],"cluster"=km$cluster)

##Animation

install.packages("animation")

library(animation)

km <- kmeans.ani(airline[,-c(1)], 4)



Conclusion:

Cluster =4

Hierarchical Clustering

#Data load

mydata1<-read.csv("EastWestAirlines.csv")

str(mydata1)

##data standardization

mydata <- scale(mydata1[2:11])

d <- dist(mydata, method = "euclidean") #Computing the distance natrix

as.matrix(d)[1:6, 1:6]

fit <- hclust(d, method="ward") # Building the algorithm # try with 'centroid'

plot(fit) # display dendogram

clusters <- cutree(fit, k=4) # cut tree into 4 clusters

table(clusters)

# draw dendogram with red borders around the 4 clusters

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

#Attach the cluster numbers to Uni

Final\_output=data.frame('Uni'=mydata1[,1],'Cluster' =clusters)

View(Final\_output)

1 2 3 4 5 6

1 0.0000000 0.1374149 0.3765259 0.1353371 4.301453 0.1587920

2 0.1374149 0.0000000 0.3445425 0.1140485 4.253731 0.2106926

3 0.3765259 0.3445425 0.0000000 0.4372821 4.031616 0.5155770

4 0.1353371 0.1140485 0.4372821 0.0000000 4.319071 0.1075187

5 4.3014526 4.2537307 4.0316155 4.3190715 0.000000 4.3879926

6 0.1587920 0.2106926 0.5155770 0.1075187 4.387993 0.0000000

> fit <- hclust(d, method="ward") # Building the algorithm # try with 'centroid'

The "ward" method has been renamed to "ward.D"; note new "ward.D2"

> clusters <- cutree(fit, k=3) # cut tree into 4 clusters

> table(clusters)

clusters

1 2 3

1146 1605 1248

> # draw dendogram with red borders around the 4 clusters

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

> #Attach the cluster numbers to Uni

> Final\_output=data.frame('Uni'=mydata1[,1],'Cluster' =clusters)

> View(Final\_output)

> plot(fit) # display dendogram

> clusters <- cutree(fit, k=4) # cut tree into 4 clusters

> table(clusters)

clusters

1 2 3 4

1146 1195 410 1248

