**K-Means Clustering**

install.packages('factoextra')

library(factoextra)

install.packages("readxl")

library("readxl")

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

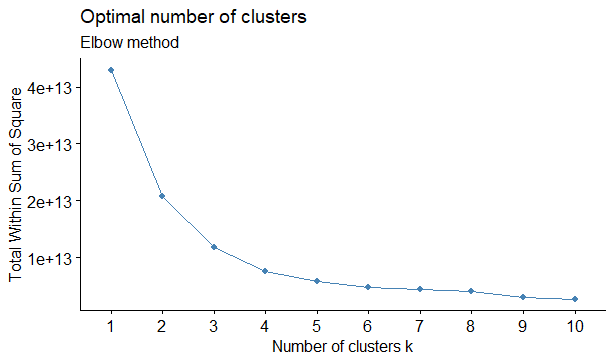
install.packages("fviz\_nbclust")

# Elbow method

summary(airline)

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

labs(subtitle = "Elbow method")



km <- kmeans(airline[,-1],3)

km$centers

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

1 36759.97 104.6905 1.785439 1.015394 1.007697 11901.00 9.926876

2 563233.22 451.9259 3.222222 1.037037 1.037037 53100.85 20.444444

3 167615.27 266.6037 3.010000 1.008750 1.027500 33942.18 17.235000

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

1 309.5686 0.9278384 3854.871 0.3316228

2 1758.1235 5.7037037 6203.753 0.7901235

3 915.1500 2.6725000 4935.159 0.4787500

km$cluster

> km$cluster

km$cluster

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

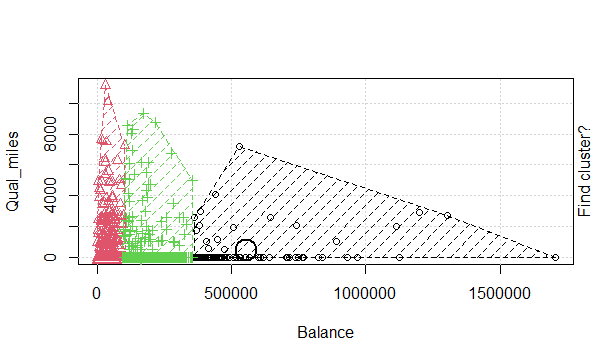
clust<-data.frame("Airlines"=airline[,1],"cluster"=km$cluster)

##Animation

install.packages("animation")

library(animation)

km <- kmeans.ani(airline[,-c(1)], 3)



Conclusion:

Cluster =3

**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)

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

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

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"

> plot(fit) # display dendogram

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

> table(clusters)

clusters

1 2 3

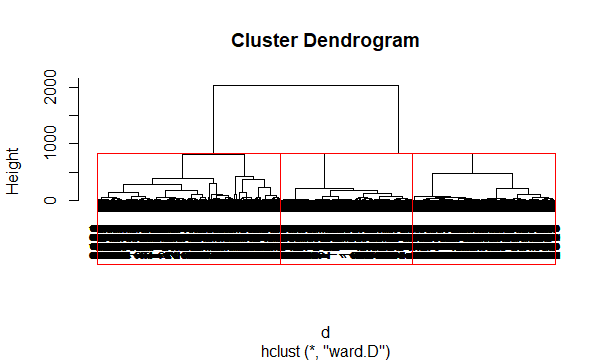
1146 1605 1248

> # draw dendogram with red borders around the 4 clusters

> rect.hclust(fit, k=3, border="red")

> #Attach the cluster numbers to Uni

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



Conclusion:

3 clusters are formed. Here, ward is used in hierarchical clustering