|  |
| --- |
| > ## Load package 'readxl' to load data from xlsx file  > library(readxl)  > airline\_data <- read.csv(file.choose())  > View(airline\_data)  > normalized\_data <- scale(airline\_data[,2:12])  > d <- dist(normalized\_data,method = "euclidean") ## Distance matrix  > fit <- hclust(d,method = "complete")  > plot(fit) #### display Dendrogram  > plot(fit,hang = -1)  > groups <- cutree(fit,k=100)  > rect.hclust(fit,k=100,border="red")  > membership <- as.matrix(groups)  > final <- data.frame(airline\_data,membership)  > View(final)  > final1 <- final[,c(ncol(final),1:(ncol(final)-1))]  > View(final1)  > #install.packages("plyr")  > library(plyr)  > airline\_data <- read.csv(file.choose())  > View(airline\_data)  > plot(airline\_data)  > km <- kmeans(airline\_data,8) ## kmeans clustering - 4 clusters;k ~ sqrt(n/2)  > km  K-means clustering with 8 clusters of sizes 510, 1665, 19, 142, 954, 291, 68, 350  Cluster means:  ID. Balance Qual\_miles cc1\_miles cc2\_miles cc3\_miles Bonus\_miles Bonus\_trans  1 1766.8843 105856.16 208.90588 2.378431 1.005882 1.003922 18009.220 13.672549  2 2309.9195 16182.02 81.26006 1.308709 1.015616 1.000000 5189.429 7.157357  3 814.0000 922162.53 564.73684 3.210526 1.000000 1.157895 58492.053 20.894737  4 1381.5352 266577.35 319.03521 3.077465 1.000000 1.000000 39022.451 17.676056  5 2070.7170 52093.84 125.32285 1.853249 1.019916 1.000000 11048.897 11.131027  6 1737.6873 67002.09 169.42955 4.274914 1.010309 1.116838 59668.763 21.391753  7 971.8824 444918.54 418.25000 3.235294 1.044118 1.000000 51954.397 20.455882  8 1575.0657 167363.80 231.82286 3.182857 1.011429 1.028571 36135.557 18.182857  Flight\_miles\_12mo Flight\_trans\_12 Days\_since\_enroll Award.  1 582.6608 1.7941176 4584.751 0.3431373  2 187.8408 0.6114114 3558.107 0.2816817  3 1607.5263 6.2631579 6642.316 0.8947368  4 1013.8944 3.0492958 5254.739 0.5492958  5 378.4161 1.1153040 4038.305 0.3123690  6 779.2921 2.2852234 4610.959 0.7663230  7 1758.3971 5.4852941 6113.985 0.7352941  8 994.2314 2.5885714 4929.109 0.4885714  Clustering vector:  [1] 2 2 5 2 1 2 1 2 7 1 5 6 5 5 2 6 6 2 1 2 1 8 2 5 8 2 2 2 5 2 8 2 4 2 5 1 2 5 5 2 2 2 5 7  [45] 1 4 1 5 6 2 1 2 1 1 5 2 6 5 5 1 1 2 2 7 5 5 2 1 4 5 1 4 7 2 1 1 1 1 5 6 2 1 2 5 2 6 5 7  [89] 5 5 2 2 6 8 8 2 5 1 2 5 6 2 2 2 2 2 6 5 6 2 8 4 1 5 4 5 8 3 1 2 5 8 6 5 6 8 7 8 5 8 2 2  [133] 8 1 1 4 1 1 8 2 5 2 2 1 5 1 2 6 1 2 6 3 4 1 5 4 8 5 4 5 5 1 8 2 1 5 1 7 7 2 6 5 4 8 1 7  [177] 1 2 5 2 5 1 2 2 5 1 3 8 7 5 6 8 7 5 2 1 6 1 2 1 5 2 2 8 5 5 6 2 2 2 8 2 8 2 2 1 2 5 8 8  [221] 3 4 5 2 8 1 1 1 5 8 1 6 2 2 2 8 5 4 6 2 8 2 4 1 7 2 8 2 6 1 5 2 1 5 1 4 7 2 8 7 8 5 8 5  [265] 5 1 5 5 6 6 8 1 6 2 5 3 2 1 2 6 8 5 8 8 6 1 6 5 8 5 7 8 4 2 2 8 6 4 8 5 2 2 4 7 1 6 1 7  [309] 1 5 8 2 5 3 8 1 2 5 3 5 1 8 2 4 5 7 5 7 8 5 8 5 1 5 8 4 1 2 5 5 8 2 5 1 5 6 1 8 8 1 2 1  [353] 4 2 5 2 8 4 5 1 2 2 4 2 5 5 5 2 5 8 1 1 6 8 2 2 2 8 2 6 7 8 6 7 6 5 5 5 8 1 1 5 2 2 1 2  [397] 1 2 2 2 8 5 1 2 7 8 2 4 2 2 7 6 6 4 5 1 5 5 1 8 7 2 8 6 1 1 5 7 1 6 4 2 5 8 5 2 1 7 6 2  [441] 2 1 2 2 8 1 4 2 7 2 8 5 8 8 5 2 5 8 5 5 5 8 2 2 8 7 3 7 1 1 2 3 6 5 5 8 4 8 4 1 5 3 2 1  [485] 5 5 6 1 3 1 2 5 6 2 1 2 2 2 2 8 8 1 2 4 2 8 8 2 1 2 3 2 8 2 5 4 2 6 8 2 2 8 6 2 1 5 5 2  [529] 2 2 7 6 6 2 4 2 2 6 2 2 2 5 2 2 2 8 8 5 2 4 1 1 2 2 8 2 2 8 2 2 5 7 8 6 6 8 5 5 2 5 2 5  [573] 1 2 6 5 6 2 2 2 1 5 5 2 2 2 5 1 2 1 1 2 5 5 6 2 2 2 6 4 8 5 6 1 1 1 2 2 7 2 5 8 5 5 5 1  [617] 1 8 4 6 2 1 5 8 7 1 5 5 8 3 7 2 5 5 2 2 5 2 1 6 2 5 2 5 2 2 2 2 2 4 8 8 4 5 2 2 5 1 4 2  [661] 5 2 6 1 5 1 2 5 2 8 6 2 1 5 8 2 2 2 4 8 2 1 4 7 5 8 2 6 1 6 2 1 4 7 2 5 2 7 5 1 5 7 4 2  [705] 4 8 2 2 4 2 2 6 2 2 1 5 8 2 6 5 5 5 2 7 2 8 2 5 5 2 1 2 2 2 2 1 6 5 5 8 1 6 5 7 5 4 5 1  [749] 2 2 8 1 1 5 8 2 4 6 2 5 1 2 2 1 6 4 2 2 2 8 2 8 1 5 6 2 5 4 5 2 8 4 2 2 5 8 2 4 8 1 2 1  [793] 8 1 5 5 6 2 8 2 6 2 2 5 2 5 6 8 5 2 5 5 6 1 8 2 5 2 2 5 7 8 4 3 5 5 1 5 6 2 5 6 4 1 2 6  [837] 8 8 2 1 8 6 8 5 1 2 5 6 2 5 6 6 2 1 6 1 5 5 6 1 1 8 2 1 6 5 2 5 8 4 5 2 2 5 2 5 2 2 2 2  [881] 4 4 5 1 8 2 2 1 7 1 1 8 1 5 2 5 4 2 2 1 8 8 1 1 6 4 2 2 2 1 2 4 2 2 8 5 2 8 4 6 2 2 4 8  [925] 5 2 2 2 8 2 1 2 4 2 1 8 1 5 2 1 5 2 5 2 5 1 1 4 5 2 5 8 5 1 6 6 2 2 2 1 2 2 2 4 2 1 2 2  [969] 1 2 2 2 8 5 2 5 2 8 5 6 8 8 2 1 4 5 2 1 1 5 2 2 1 6 2 1 2 4 8 7  [ reached getOption("max.print") -- omitted 2999 entries ]  Within cluster sum of squares by cluster:  [1] 2.504496e+11 2.891750e+11 1.303138e+12 3.843647e+11 2.548610e+11 3.995391e+11 6.087753e+11  [8] 4.696673e+11  (between\_SS / total\_SS = 90.8 %)  Available components:  [1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss"  [7] "size" "iter" "ifault"  > library(animation)  > km <- kmeans.ani(airline\_data,8)  > km$cluster  [1] 7 7 3 7 2 7 2 7 1 2 3 5 3 3 7 5 5 7 2 7 2 4 7 3 4 7 7 7 3 7 4 7 8 7 3 2 7 3 3 7 7 7 3 1  [45] 2 8 2 3 5 7 2 7 2 2 3 7 5 3 3 2 2 7 7 1 3 3 7 2 8 3 2 8 1 7 2 2 2 2 3 5 7 2 7 3 7 5 3 1  [89] 3 3 7 7 5 4 4 7 3 2 7 3 5 7 7 7 7 7 5 3 5 7 4 8 2 3 8 3 4 6 2 7 3 4 5 3 5 4 1 4 3 4 7 7  [133] 4 2 2 8 2 2 4 7 3 7 7 2 3 2 7 5 2 7 5 6 8 2 3 8 4 3 8 3 3 2 4 7 2 3 2 1 1 7 5 3 8 4 2 1  [177] 2 7 3 7 3 2 7 7 3 2 6 4 1 3 5 4 1 3 7 2 5 2 7 2 3 7 7 4 3 3 5 7 7 7 4 7 4 7 7 2 7 3 4 4  [221] 6 8 3 7 4 2 2 2 3 4 2 5 7 7 7 4 3 8 5 7 4 7 8 2 1 7 4 7 5 2 3 7 2 3 2 8 1 7 4 1 4 3 4 3  [265] 3 2 3 3 5 5 4 2 5 7 3 6 7 2 7 5 4 3 4 4 5 2 5 3 4 3 1 4 8 7 7 4 5 8 4 3 7 7 8 1 2 5 2 1  [309] 2 3 4 7 3 6 4 2 7 3 6 3 2 4 7 8 3 1 3 1 4 3 4 3 2 3 4 8 2 7 3 3 4 7 3 2 3 5 2 4 4 2 7 2  [353] 8 7 3 7 4 8 3 2 7 7 8 7 3 3 3 7 3 4 2 2 5 4 7 7 7 4 7 5 1 4 5 1 5 3 3 3 4 2 2 3 7 7 2 7  [397] 2 7 7 7 4 3 2 7 1 4 7 8 7 7 1 5 5 8 3 2 3 3 2 4 1 7 4 5 2 2 3 1 2 5 8 7 3 4 3 7 2 1 5 7  [441] 7 2 7 7 4 2 8 7 1 7 4 3 4 4 3 7 3 4 3 3 3 4 7 7 4 1 6 1 2 2 7 6 5 3 3 4 8 4 8 2 3 6 7 2  [485] 3 3 5 2 6 2 7 3 5 7 2 7 7 7 7 4 4 2 7 8 7 4 4 7 2 7 6 7 4 7 3 8 7 5 4 7 7 4 5 7 2 3 3 7  [529] 7 7 1 5 5 7 8 7 7 5 7 7 7 3 7 7 7 4 4 3 7 8 2 2 7 7 4 7 7 4 7 7 3 1 4 5 5 4 3 3 7 3 7 3  [573] 2 7 5 3 5 7 7 7 2 3 3 7 7 7 3 2 7 2 2 7 3 3 5 7 7 7 5 8 4 3 5 2 2 2 7 7 1 7 3 4 3 3 3 2  [617] 2 4 8 5 7 2 3 4 1 2 3 3 4 6 1 7 3 3 7 7 3 7 2 5 7 3 7 3 7 7 7 7 7 8 4 4 8 3 7 7 3 2 8 7  [661] 3 7 5 2 3 2 7 3 7 4 5 7 2 3 4 7 7 7 8 4 7 2 8 1 3 4 7 5 2 5 7 2 8 1 7 3 7 1 3 2 3 1 8 7  [705] 8 4 7 7 8 7 7 5 7 7 2 3 4 7 5 3 3 3 7 1 7 4 7 3 3 7 2 7 7 7 7 2 5 3 3 4 2 5 3 1 3 8 3 2  [749] 7 7 4 2 2 3 4 7 8 5 7 3 2 7 7 2 5 8 7 7 7 4 7 4 2 3 5 7 3 8 3 7 4 8 7 7 3 4 7 8 4 2 7 2  [793] 4 2 3 3 5 7 4 7 5 7 7 3 7 3 5 4 3 7 3 3 5 2 4 7 3 7 7 3 1 4 8 6 3 3 2 3 5 7 3 5 8 2 7 5  [837] 4 4 7 2 4 5 4 3 2 7 3 5 7 3 5 5 7 2 5 2 3 3 5 2 2 4 7 2 5 3 7 3 4 8 3 7 7 3 7 3 7 7 7 7  [881] 8 8 3 2 4 7 7 2 1 2 2 4 2 3 7 3 8 7 7 2 4 4 2 2 5 8 7 7 7 2 7 8 7 7 4 3 7 4 8 5 7 7 8 4  [925] 3 7 7 7 4 7 2 7 8 7 2 4 2 3 7 2 3 7 3 7 3 2 2 8 3 7 3 4 3 2 5 5 7 7 7 2 7 7 7 8 7 2 7 7  [969] 2 7 7 7 4 3 7 3 7 4 3 5 4 4 7 2 8 3 7 2 2 3 7 7 2 5 7 2 7 8 4 1  [ reached getOption("max.print") -- omitted 2999 entries ]  > km$withinss  NULL  > km$centers  ID. Balance Qual\_miles cc1\_miles cc2\_miles cc3\_miles Bonus\_miles Bonus\_trans  [1,] 971.8824 444918.54 418.25000 3.235294 1.044118 1.000000 51954.397 20.455882  [2,] 1766.1370 105756.72 208.49706 2.369863 1.005871 1.003914 17880.961 13.624266  [3,] 2073.0188 52002.44 125.19162 1.858639 1.019895 1.000000 11111.969 11.139267  [4,] 1575.0657 167363.80 231.82286 3.182857 1.011429 1.028571 36135.557 18.182857  [5,] 1737.6586 67287.09 170.01379 4.275862 1.010345 1.117241 59792.617 21.437931  [6,] 814.0000 922162.53 564.73684 3.210526 1.000000 1.157895 58492.053 20.894737  [7,] 2308.9591 16170.42 81.30889 1.308894 1.015625 1.000000 5192.547 7.161659  [8,] 1381.5352 266577.35 319.03521 3.077465 1.000000 1.000000 39022.451 17.676056  Flight\_miles\_12mo Flight\_trans\_12 Days\_since\_enroll Award.  [1,] 1758.3971 5.4852941 6113.985 0.7352941  [2,] 586.6086 1.8003914 4588.836 0.3424658  [3,] 376.5026 1.1068063 4033.625 0.3141361  [4,] 994.2314 2.5885714 4929.109 0.4885714  [5,] 778.0103 2.3000000 4610.862 0.7620690  [6,] 1607.5263 6.2631579 6642.316 0.8947368  [7,] 187.9537 0.6117788 3559.282 0.2818510  [8,] 1013.8944 3.0492958 5254.739 0.5492958  > km\_8 <- kmeans(airline\_data,10)  > wss<-(nrow(airline\_data)-1)\*sum(apply(airline\_data,2,var))  > for(i in 2:10) wss[i]<-sum(kmeans(airline\_data,centers = i)$withinss)  > plot(1:10,wss,type = "b",xlab = "No of clusters",ylab = "Avg distance") |
|  |
| |  | | --- | | > | |



