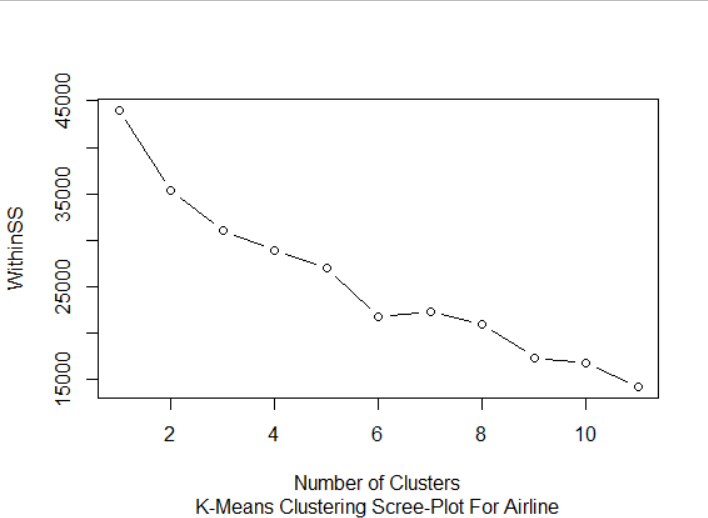


**HCLUST DENDOGRAM FOR AIRLINE: HIERARCHICAL CLUSTERING**

INFERENCES : Hclust doesn’t work or help In when drawing plots of large dataset .Hclust is useful for relatively small dataset.

**KMEANS CLUSTER FOR AIRLINE DATASET:**

* After finding the twss (total within sum of squares) ie adding the withins SS of each clusters.
* Within sum of squares the less the better and changing the number of clusters will be make the difference
* 43978.00 35400.65 31001.31 29035.77 24877.43 24402.81 22034.19 19477.13 16060.60 20097.96
* [11] 15935.97 . These are the TOTAL WITHIN SS of each Col .(11 cols)
* Below Elbow curve has a start of descent from 4 onwards .
* So we change the no. of clusters to 4 .
* fit <- kmeans(normalized\_data, 4) #trying with 4 clusters to get minimum withinss.
* str(fit)
* fit$cluster
* View(final <- data.frame(fit$cluster, airline))



List of 9

$ cluster : int [1:3999] 3 3 3 3 1 3 1 3 2 1 ...

$ centers : num [1:4, 1:11] 0.445 1.158 -0.293 0.457 -0.112 ...

..- attr(\*, "dimnames")=List of 2

.. ..$ : chr [1:4] "1" "2" "3" "4"

.. ..$ : chr [1:11] "Balance" "Qual\_miles" "cc1\_miles" "cc2\_miles" ...

$ totss : num 43978

$ withinss : num [1:4] 11727 4383 11267 901

$ tot.withinss: num 28279

$ betweenss : num 15699

$ size : int [1:4] 1237 147 2554 61

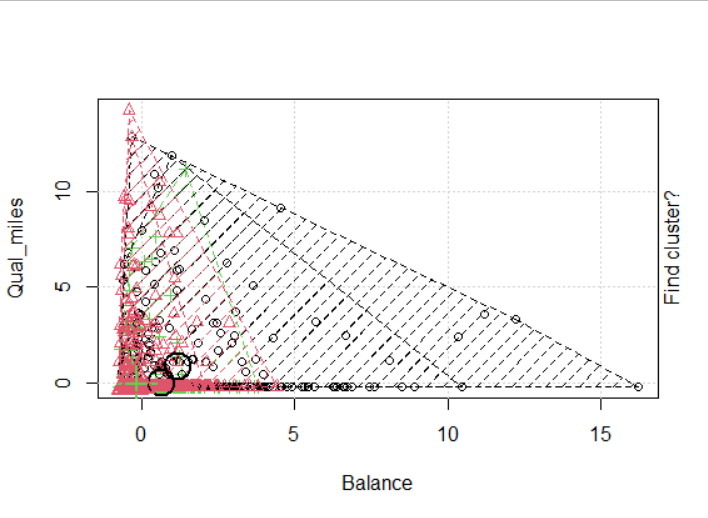
$ iter : int 4

$ ifault : int 0

- attr(\*, "class")= chr "kmeans"

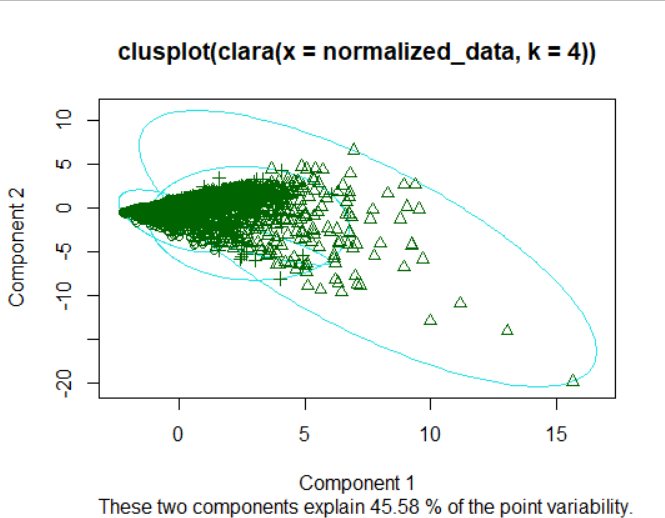
> fit$cluster

Kmeans.ani(normalized\_Data, 4)



* View(aggregate(airline, by = list(fit$cluster), FUN = mean)) .

Aggregating normal airline data with where its respective cluster Number as Column and along with it the Mean of those entries.



Above clusplot to plot the cluster on normalized data on Airline Dataset.

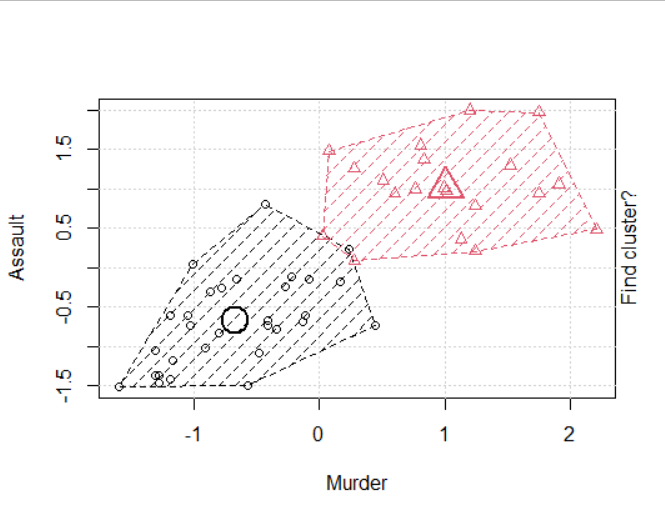
**Crime Dataset Clustering Kmeans :**

**> twss**

**[1] 196.00000 102.86240 78.32327 56.40317**

Total within ss for the 4 cols on crime dataset

Kmeans cluster with K =2



fit$cluster

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

[47] 1 1 1 1

> fit$centers

Murder Assault UrbanPop Rape

[1,] -0.669956 -0.6758849 -0.1317235 -0.5646433

[2,] 1.004934 1.0138274 0.1975853 0.8469650

Above are centroids K= 2 clusters w.r.t to each cols [ Murder , Assault , UrbanPop ,Rape]

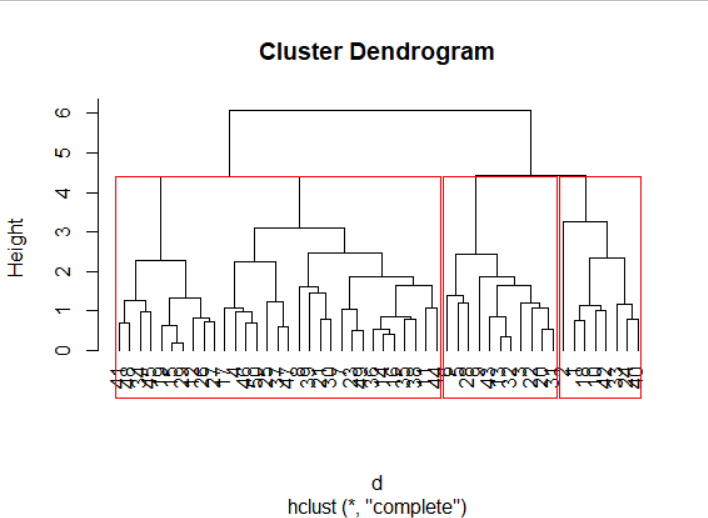
After aggregating the dataset with their mean value.

Group.1 Murder Assault UrbanPop Rape

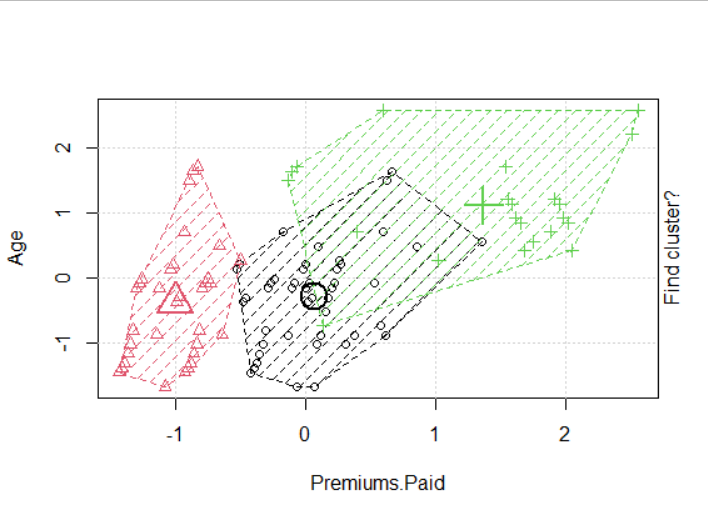
1 1 4.870 114.4333 63.63333 15.94333

2 2 12.165 255.2500 68.40000 29.16500

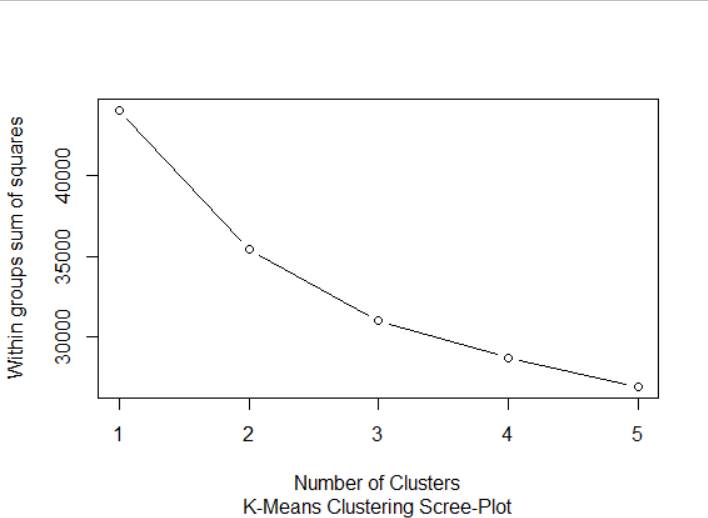
**Crime Dataset using Hierarchical Clustering:**



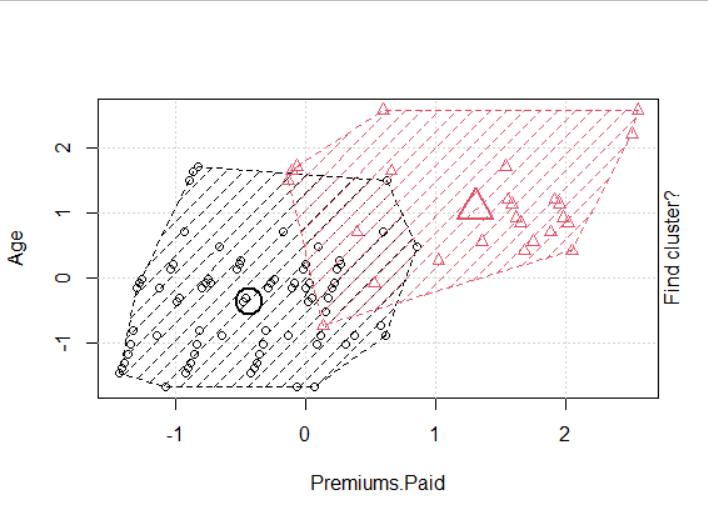
Clusters k=3 are bordered with red color .

**INSURANCE DATASET CLUSTERING USING** K-El:

ELBOW CURVE ON THE SCREE PLOT SHOWS THAT CURVE STARTED THE DESCENT ON K=2 ONWARDS.



SO WE TAKE K=2 AND TAKE KMEAN AGAIN .THE CLUSTER PLOT



**OBSERVATION on Inurance Dataset: :**

* Majority clusters fall under 2nd cluster .
* The average age is 61 .Assuming senior citizens have a larger Premium paid and their income /pension is higher.
* Whereas the younger age group including average age as 41 has paid lesser Premium .
* The population of younger people outweigh the senior citizens but still senior citizens have higher values in their details.