**Business Problem:**

Perform clustering for the crime data and identify the number of clusters formed and

**Data:**

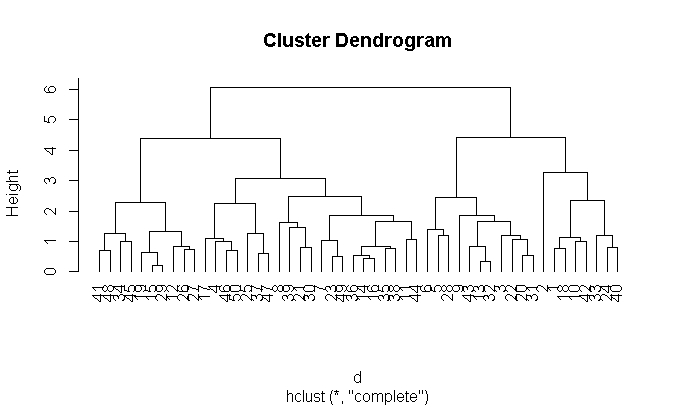
Data in the form of numeric data. It contains the numerical data (Murder, Assault, UrbanPop and Rape)

**Pre-processing Data:**

All the features are not in same scale. So first of all convert them all into single scale. There was no outlier and NA in the data. Delete the unused feature like ‘X’ from the processing.

**Building the Model:**

Build the model using Euclidean distance and complete linkage functions. Please find the Dendrogram.

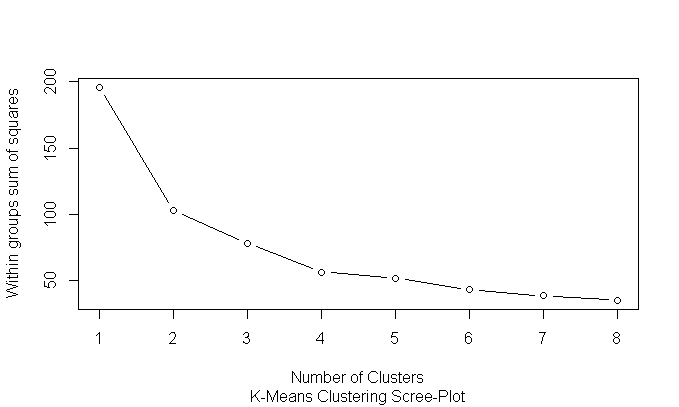


**K-Selection:**

k <- kselection(mydata[,-1], parallel = TRUE, k\_threshold = 0.9, max\_centers=5)

It is **giving k value as 2.**

**Elbow Curve:**



Elbow Curve value is subjective. As per **my analysis 2** will be good value for the K and subjective.

**KMeans:**

Calculate the tot.withinss and betweenss. For a good model tot.withinss should be more and betweenss should be less. If there should be less variation in the difference then we can treat it as final K.

Trails :

K=2 : km <- kmeans(normalized\_data,2)

$ tot.withinss: num 103

$ betweenss : num 93.1

K=3 : km <- kmeans(normalized\_data,3)

$ tot.withinss: num 85.1

$ betweenss : num 111.1

As per standards tot.withinss should be more than betweenss but when we go to cluster 3 then it will be giving lesser value. **So 2 will be good value for k**.

**KMeans Center and Animation diagram:**

km <- kmeans.ani(normalized\_data, 2)

km$centers

##Murder Assault UrbanPop Rape

##[1,] 1.004934 1.0138274 0.1975853 0.8469650

##[2,] -0.669956 -0.6758849 -0.1317235 -0.5646433

km <- kmeans.ani(normalized\_data, 3)

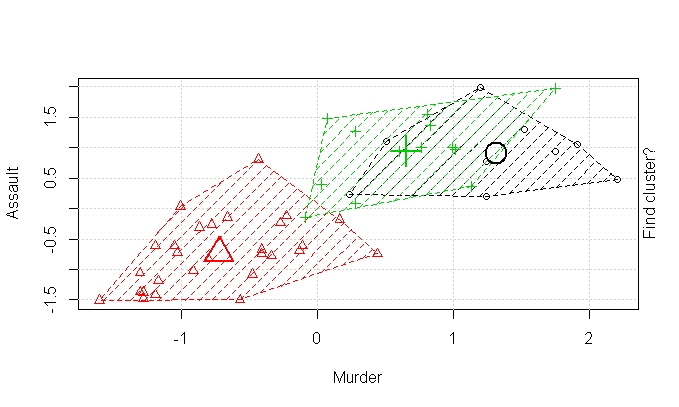
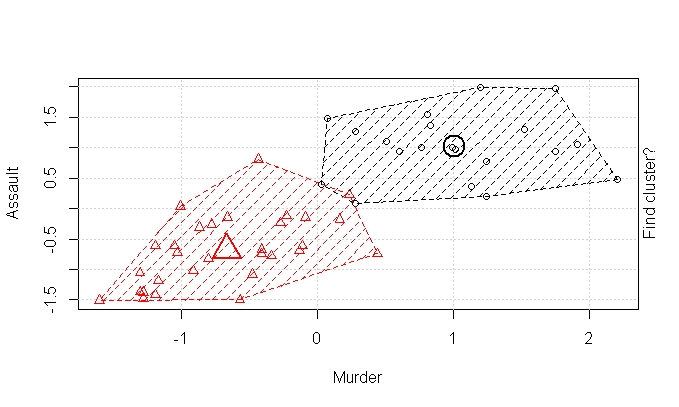
km$centers

##Murder Assault UrbanPop Rape

##[1,] 1.3114424 0.9001666 -0.8586592 0.2931524

##[2,] -0.7229267 -0.7273679 -0.1606737 -0.5890976

##[3,] 0.6491513 0.9434464 0.9405229 1.0658739



By looking both diagrams and centre values, there will be no over lapping for cluster-2 compare than cluster-3.

**Hence K = 2 is good value.**