PROBLEM STATEMENT:

Perform Clustering for the crime data and identify the number of clusters formed and draw inferences.

THE HEIRACHICAL CLUSTERRING ALGORITHM:

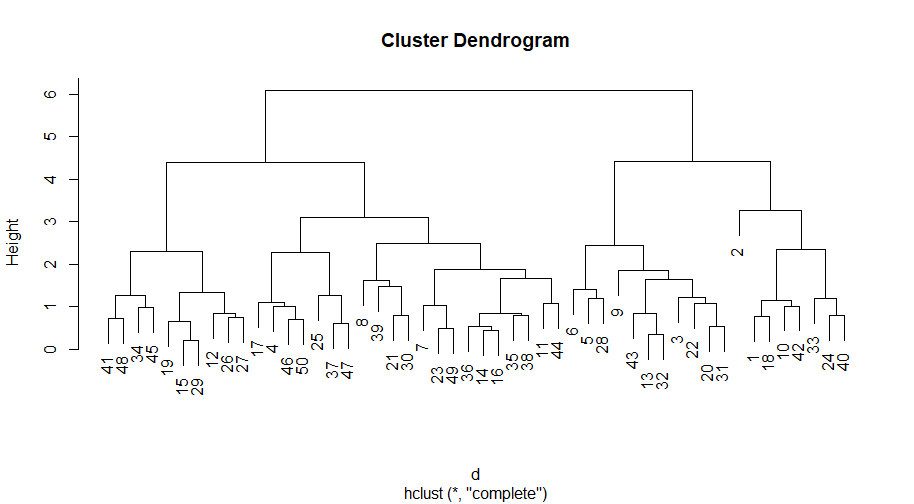
* Start with n clusters(record=clusters)
* Step 1: two closest records are merged into one cluster.
* At every step pair of clusters with smallest distance will be merged.

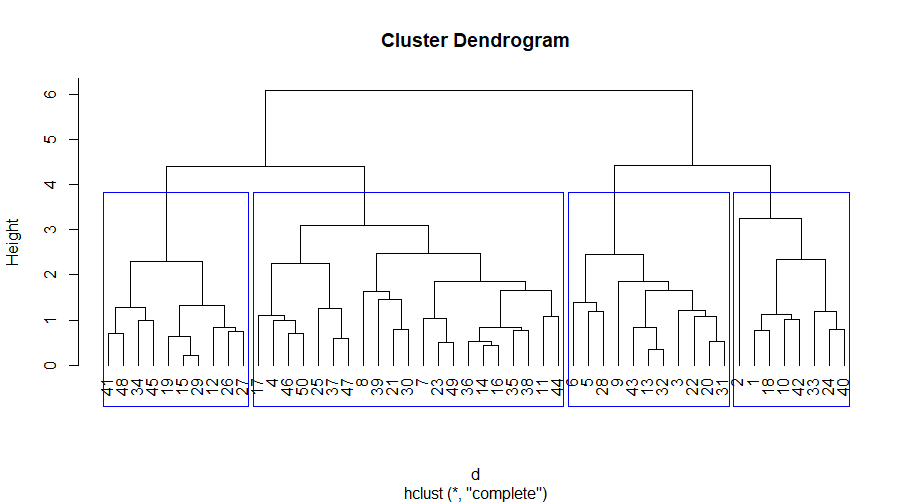
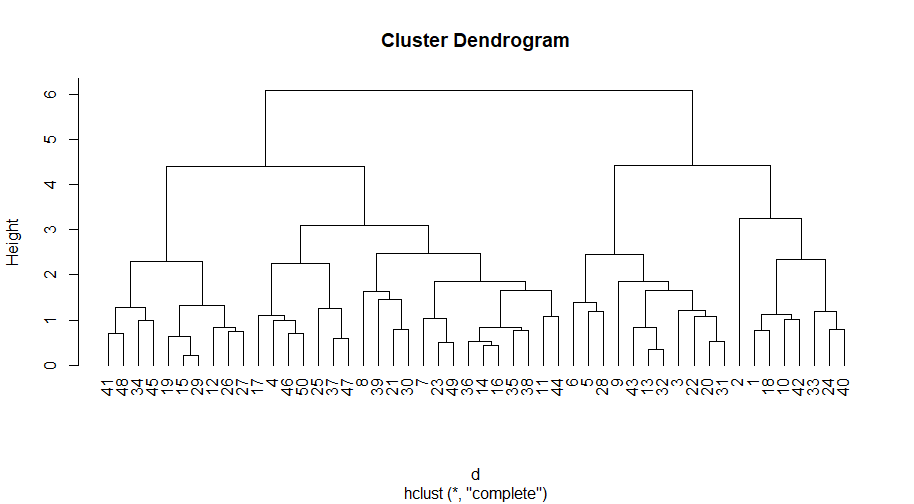
-two records are merged.

-or single record added to an existing cluster.

-or two existing clusters are combined.

1. Load the data frame and remove the column with the names.
2. Convert the data to Standardize values (z values).
3. Use Euclidean method for finding the distance between the records and merge the nearest neighbours.
4. Measure distance between clusters using linkage functions.

DENDROGRAMS: 



* Based on the dendrogram 4 number of clusters formed.
* Based on the 4 number of clusters created, we can say that

Fourth cluster is best when compared with the crime rate of other clusters.

* Fourth cluster consists of low crime rate, while other clusters contain high crime rate than the 4th cluster.
* 1st cluster consists of high crime rate .

ORDER OF CLUSTERS BASED ON THE CRIME RATE HIGH TO LOW:

* 1
* 3
* 2
* 4