****

**ABHINAV VIJAYAKUMAR**

**19BCE1311**

**CSE3506 – ESSENTIALS OF DATA ANALYTICS LAB-8**

**DR. LAKSHMI PATHI JAKKAMPUTI (L21 + L22)**

**--------------------------------------------------------------------------------------------**

**Tasks for Week-8: Hierarchical Clustering**

**Aim:** To understand the following operations/functions on ‘USArrests’ data and perform similar operations on ‘iris’ dataset based on given instructions.

**Algorithm:**

**1.** Removing all the values from the global environment

**2.** Set the working directory to the dataset where we store by using setwd().

**3.** To see the dataset use view() function.

**4.** By using scale function, we scale the data and store it in another variable.

**5.** Using dist function we find the Euclidean distances for the scaled data.

**6.** By using the Euclidean distances and hclust function we can create and then plot the hierarchical clustering dendogram.

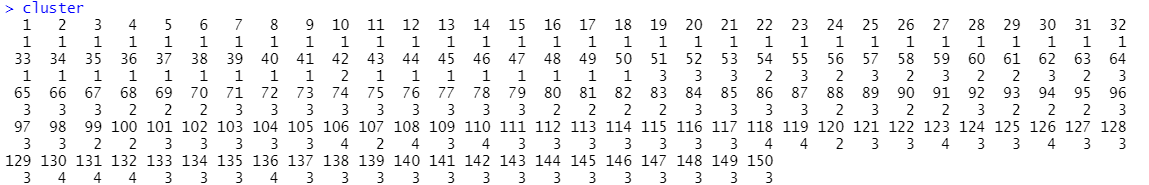
**7.** By using cutree we divide the elements of the dendogram into k number of clusters (k=4 in our case).

**8.** Then, by using rect.hclust function we can divide the dendogram into k clusters (k=4 in our case) , i.e. create k rectangular divisions/borders in the dendogram

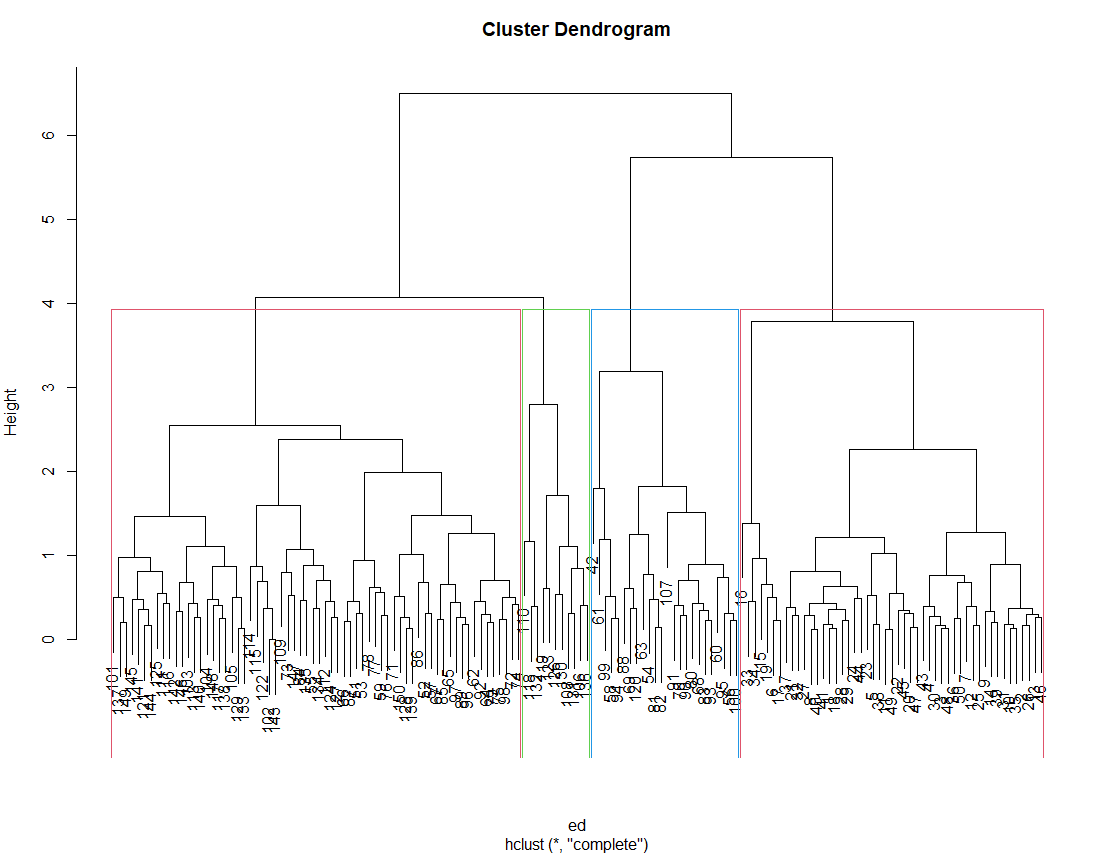
**Result:**

***For iris.csv***

**Cluster:**

****

**Cluster Dendogram:**

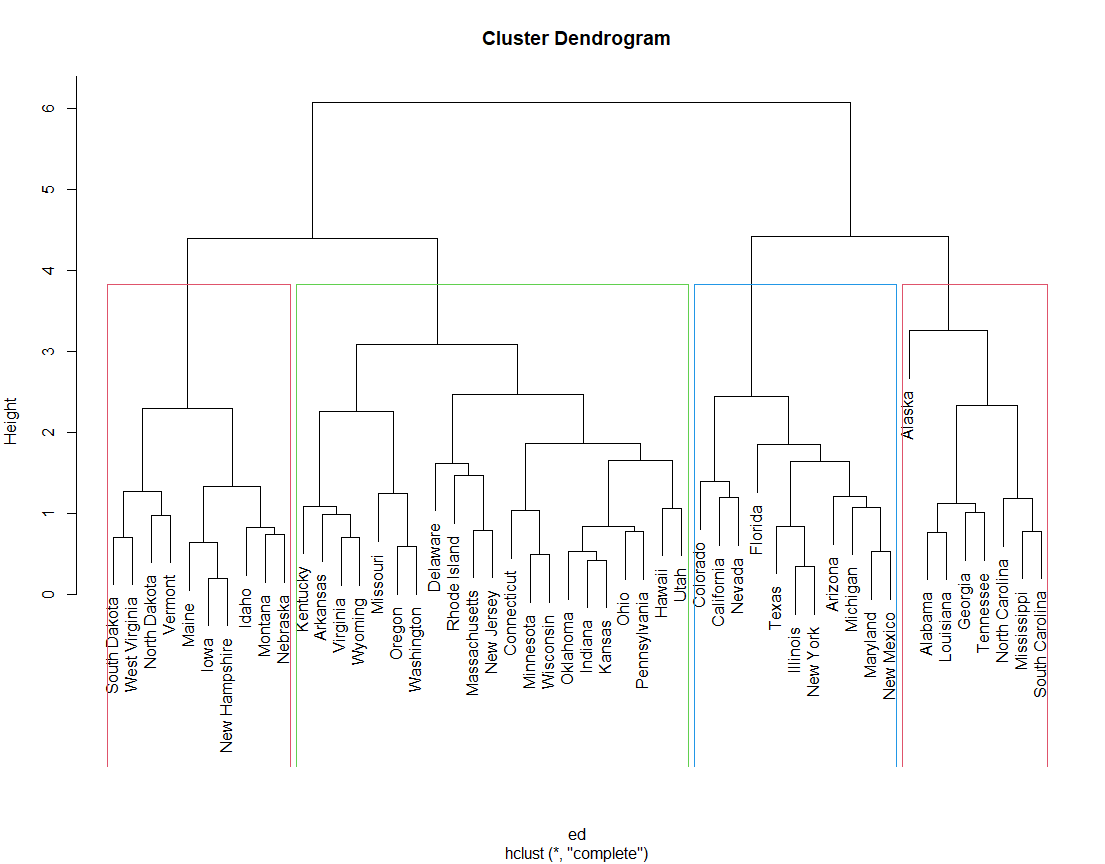
****

***For USArrests.csv***

**Cluster:**

****

**Cluster Dendogram:**

****

**Program:**

***For iris.csv***

rm(list=ls())

setwd("C:/Users/Abhinav Vijayakumar/Desktop/VIT Academics/Sem 6/EDA/LAB/LAB 8")

data <- read.csv("iris.csv",row.names=1)

View(data)

df <- scale(data)

View(df)

ed <- dist(df, method = 'euclidean')

hierClust <- hclust(ed, method = 'complete')

plot(hierClust)

cluster <- cutree(hierClust, k = 4)

cluster

rect.hclust(hierClust, k = 4, border = 2:4)

***For USArrests.csv***

rm(list=ls())

setwd("C:/Users/Abhinav Vijayakumar/Desktop/VIT Academics/Sem 6/EDA/LAB/LAB 8")

data <- read.csv("USArrests.csv",row.names=1)

View(data)

df <- scale(data)

View(df)

ed <- dist(df, method = 'euclidean')

hierClust <- hclust(ed, method = 'complete')

plot(hierClust)

cluster <- cutree(hierClust, k = 4)

cluster

rect.hclust(hierClust, k = 4, border = 2:4)