VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Essentials of Data Analytics - (CSE3506)

Faculty – Lakshmi Pathi Jakkamputi **Sir**

Lab-8

Harshanth k Prakash

19BCE1293

L21-22 **Slot**

Tasks for Week-8: Hierarchical Clustering

Understand the following operations/functions on 'USArrests' data and perform similar operations on 'iris' dataset based on given instructions.

AIM

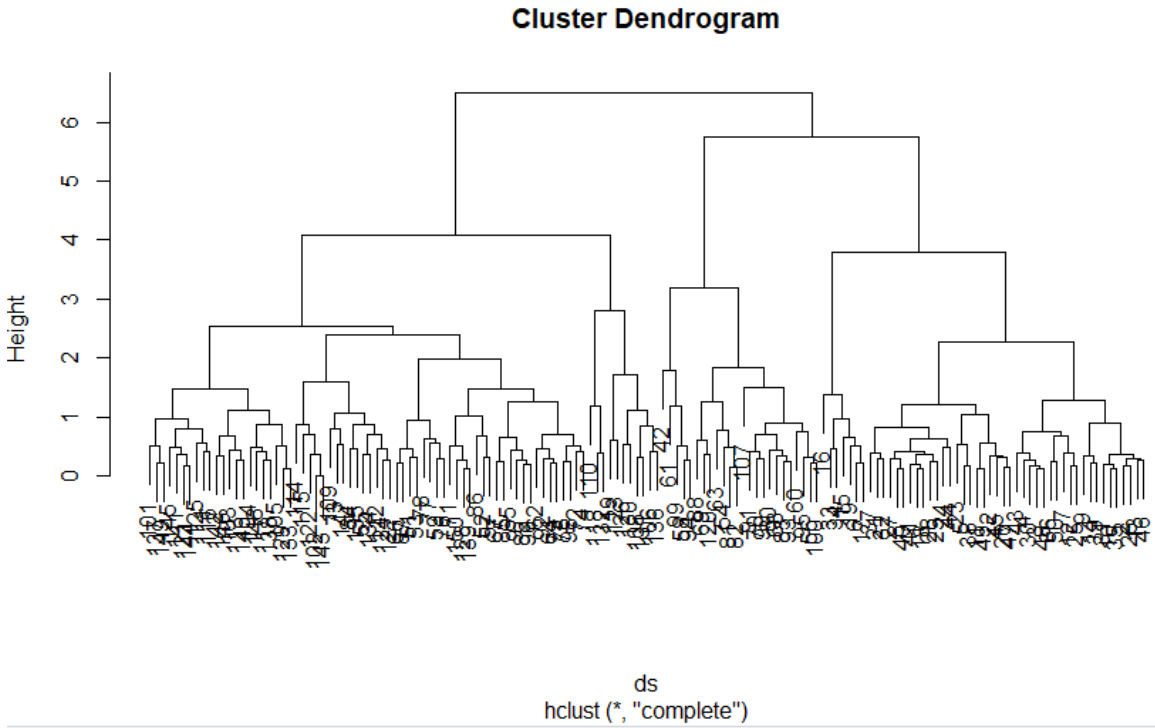
To Understand the following operations/functions on 'USArrests' data and perform similar operations on 'iris' dataset based on given instructions.

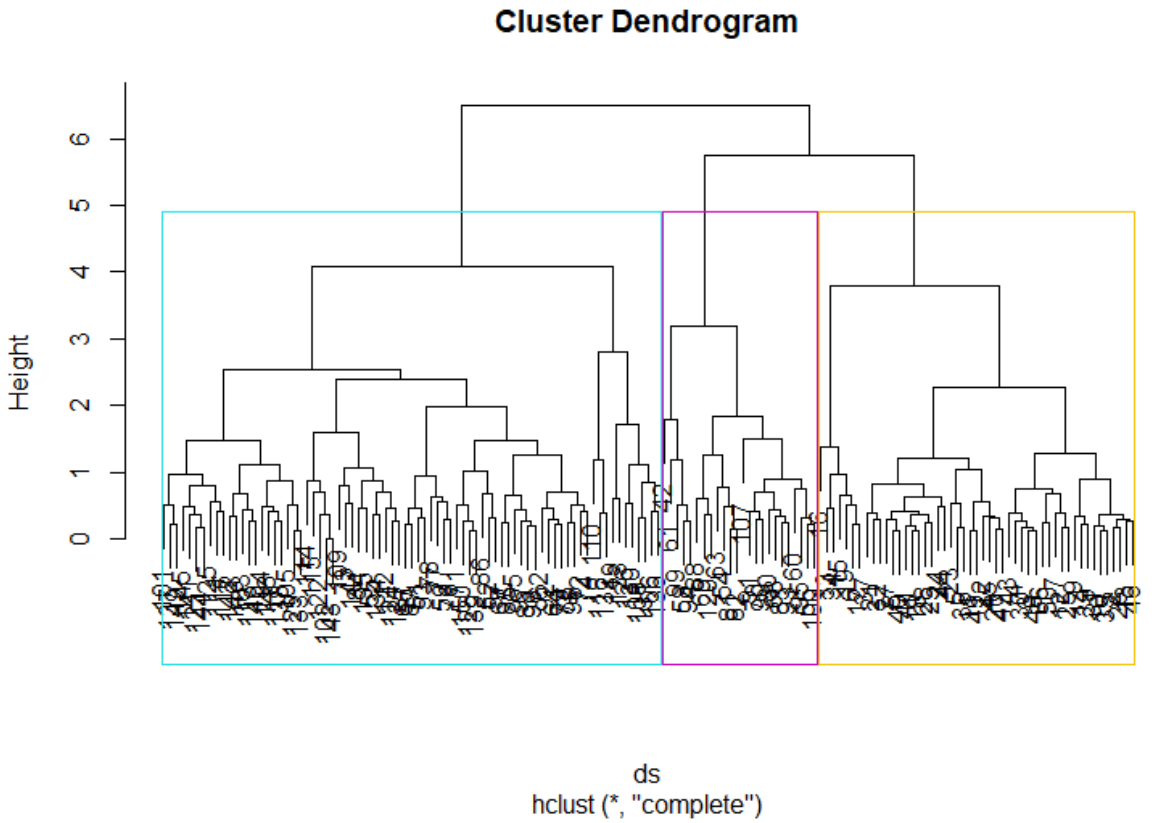
Algorithm

1. Start
2. Use rm(list=ls()) to clear commands.
3. Read the dataset.
4. Make a scaled version of the original data.
5. Find the euclidean distance between each row
6. Perform hierarchical clustering with 'complete' as the method.
7. Perform hierarchical clustering using the hclust method with the 'complete' method.
8. Plot 'heirClust' to get the dendrogram.
9. Cutree cuts the data into several groups either by specifying the desired number(s) of groups or the cut height(s).
10. rect.cluster is used to get the dendrogram
11. Stop.

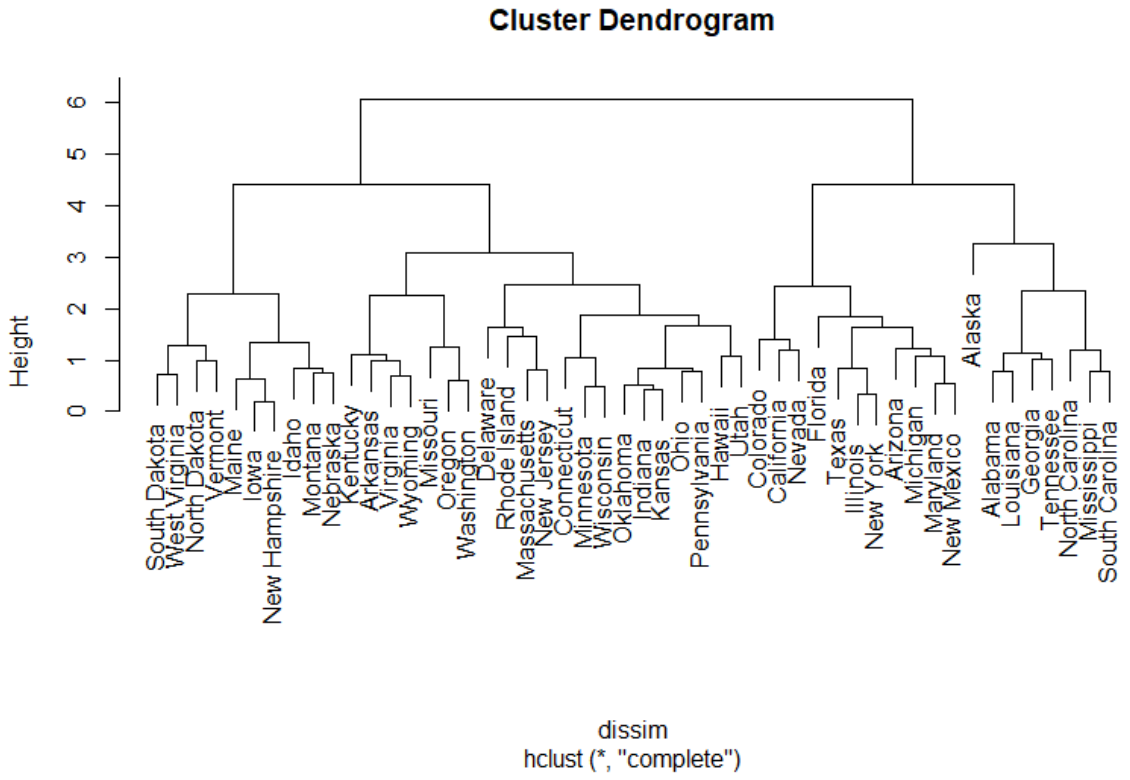
Result

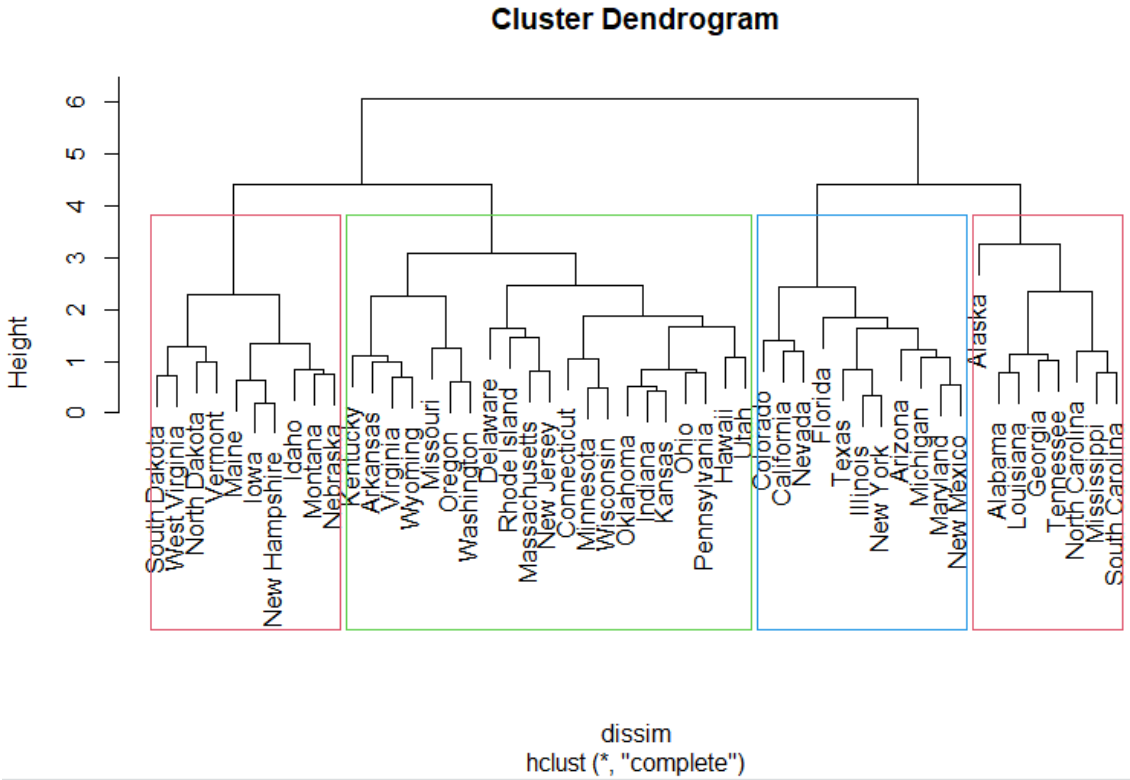
Case 1: Iris dataset





Case II: USArrests Dataset.





Program

Case 1: Iris dataset

```
rm(list=ls())
setwd("D:/data analytics/")
data <- read.csv("D:/6th Sem Works/A2- EDA/LAB/Lab8/iris.csv",row.names=1)
View(data)
df <- scale(data)
View(df)
ds <- dist(df, method = 'euclidean')
ds
hierClust <- hclust(ds, method = 'complete')
plot(hierClust)
cluster <- cutree(hierClust, k = 3)
cluster
rect.hclust(hierClust,k=3,border = 5:7)
```

Case II: USArrests Dataset.

```
rm(list=ls())
data <- read.csv("D:/6th Sem Works/A2- EDA/LAB/Lab8/USArrests.csv",row.names=1)
df <- scale(data)
dissim <- dist(df, method = 'euclidean')
hierClust <- hclust(dissim, method = 'complete')
plot(hierClust)
cluster <- cutree(hierClust, k = 4)
rect.hclust(hierClust, k = 4, border = 2:4)
```