

Q4

Omkar Sinha

12/15/2021

#####

University : Stevens Institute of Technology

Project : Final Exam

Purpose : CS513A KDD - Final Exam Question 3

First Name : Omkar

Last Name : Sinha

CWId : 10468312

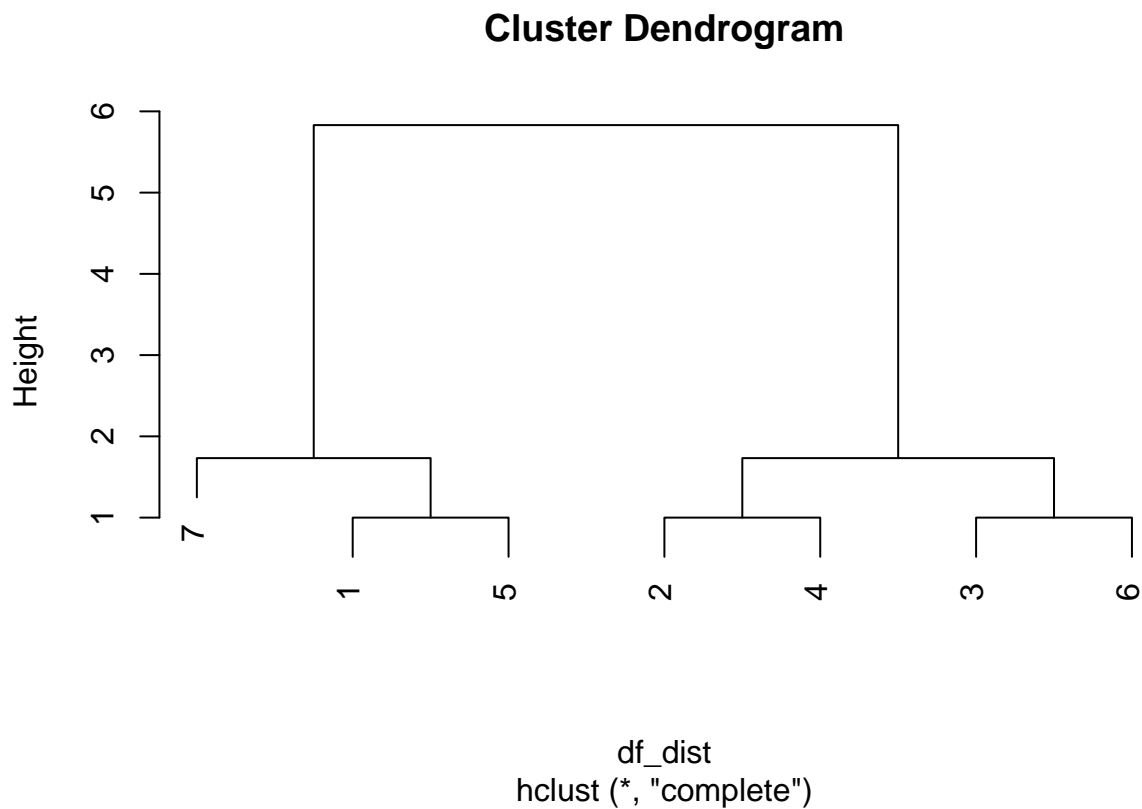
Date : 12/14/2021

```
#clear the environment
```

```
rm(list=ls())
X <- c(1,5,4,4,1,4,2)
Y <- c(1,3,4,3,2,4,1)
Z <- c(1,4,5,4,1,4,2)
df <- data.frame(X,Y,Z)
df
```

```
##   X Y Z
## 1 1 1 1
## 2 5 3 4
## 3 4 4 5
## 4 4 3 4
## 5 1 2 1
## 6 4 4 4
## 7 2 1 2
```

```
#1->a,2->b,3->c,4->d,5->e,6->f,7->g
#Hierarchial Clustering
df_dist<-dist(df)
hclust_results<-hclust(df_dist)
plot(hclust_results)
```



```
#### 4.a. Cluster members ####
#Cluster 1 has members: A, E and G
#Cluster 2 has members: B, C, D and F

#### 4.b Cluster center co-ordinates ####
clusters <- cutree(hclust_results, 2)
clusters
```

```
## [1] 1 2 2 2 1 2 1
```

```
cluster_center = aggregate(df,list(cluster=clusters), mean)
cluster_center
```

```
##   cluster      X      Y      Z
## 1      1 1.333333 1.333333 1.333333
## 2      2 4.250000 3.500000 4.250000
```

```
plot(cluster_center[2:4])
```

