LAB ASSIGNMENT- 2

NAME: M Gyanada Chowdary

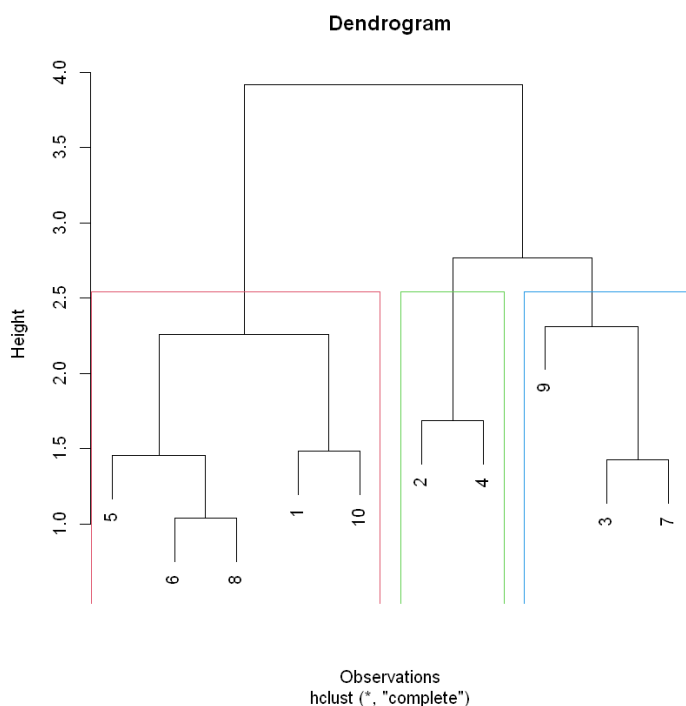**REG NO:** 21BCE7727

SLOT: L39+L40

1.Construction of Dendrogram

Code:

```
set.seed(10)
random_data <- matrix(rnorm(10 * 5), ncol = 5)
hc_random <- hclust(dist(random_data), method = "complete")

dendrogram <- cutree(hc_random, k = 3)
plot(hc_random, main = "Dendrogram", xlab = "Observations", sub = NULL)
rect.hclust(hc_random, k = 3, border = 2:4)
dendrogram
```

1. 1
2. 2
3. 3
4. 2
5. 1
6. 1
7. 3
8. 1
9. 3
10. 1

Dendrogram

2.Create dendrogram that shows the distribution of characters in a string.

Code:

```
dendrogram <- function(input_string) {
  char_count <- table(strsplit(input_string, NULL)[[1]])
  char_df <- data.frame(char = names(char_count), count = as.numeric(char_count))
  dist_matrix <- dist(char_df$count)
  hclust_result <- hclust(dist_matrix)
  plot(hclust_result, main = "Character Distribution Dendrogram", xlab = "Characters",
sub = "")
}
str <- "Gyanada"
dendrogram(str)
```

Character Distribution Dendrogram
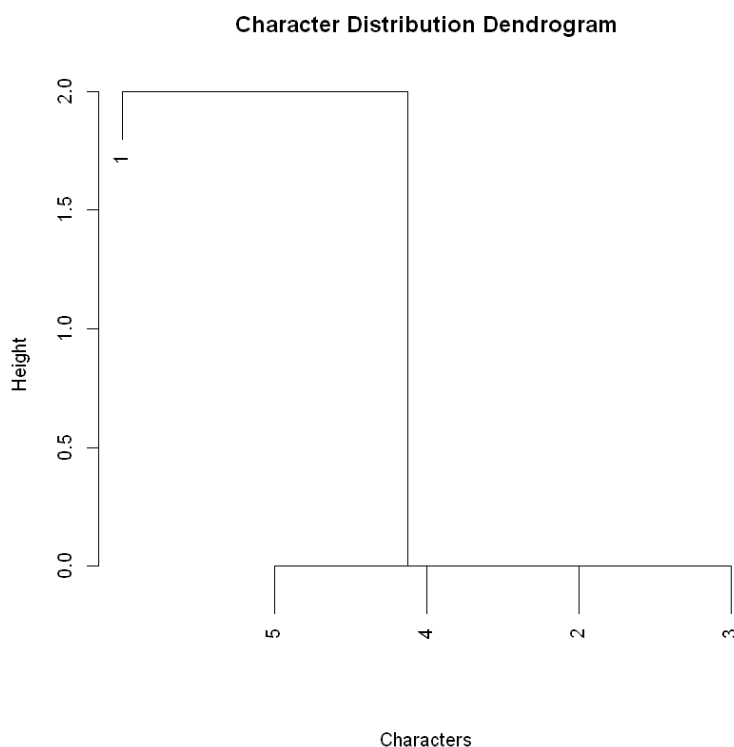
3. Let consider a hypothetical dataset, where we have a set of points in a 2D space and a similarity(symmetric)matrix indicating the pairwise similarities between these points.The choice of linkage method (single, complete, average, etc.)

Code:

```
library(cluster)

matrix <- matrix(c(0, 0.6, 0.2, 0.4, 0.7, 0.3,
                   0.6, 0, 0.5, 0.3, 0.8, 0.1,
                   0.2, 0.5, 0, 0.7, 0.4, 0.6,
                   0.4, 0.3, 0.7, 0, 0.2, 0.5,
                   0.7, 0.8, 0.4, 0.2, 0, 0.9,
                   0.3, 0.1, 0.6, 0.5, 0.9, 0),
                 nrow = 6, byrow = TRUE)
dist_matrix <- as.dist(matrix)
hc_single <- hclust(dist_matrix, method = "single")
hc_complete <- hclust(dist_matrix, method = "complete")
hc_average <- hclust(dist_matrix, method = "average")
par(mfrow=c(1,3))
```
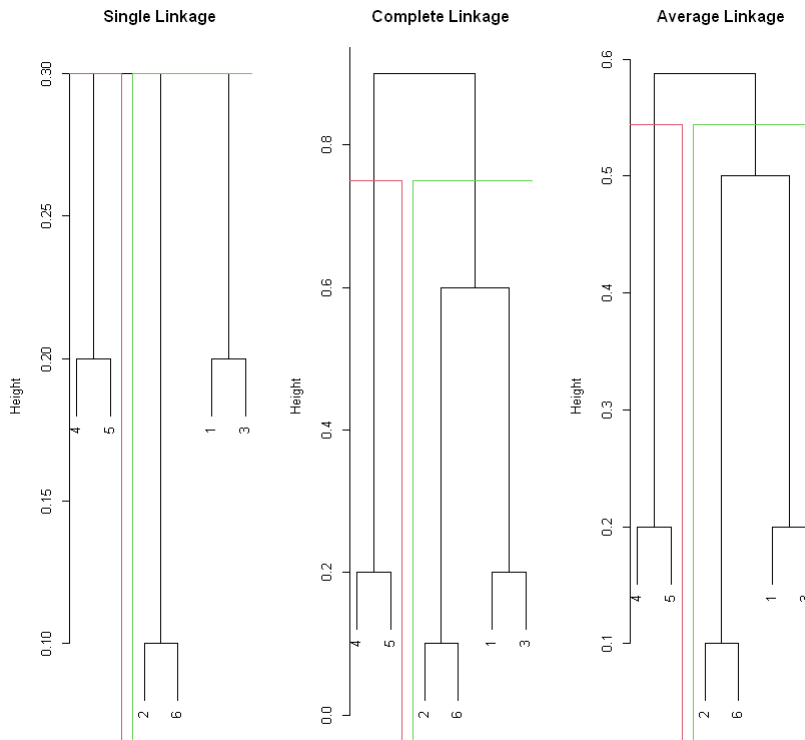
```
plot(hc_single, main = "Single Linkage", xlab = "", sub = "")
rect.hclust(hc_single, k = 2, border = 2:3)

plot(hc_complete, main = "Complete Linkage", xlab = "", sub = "")
rect.hclust(hc_complete, k = 2, border = 2:3)

plot(hc_average, main = "Average Linkage", xlab = "", sub = "")
rect.hclust(hc_average, k = 2, border = 2:3)
```


1. Minimum (Single) Linkage Dendrogram:

1. **Start:** Treat each data point as a separate cluster.

2. **Merge:** Continuously merge the two closest clusters until all points are in one cluster.

3. **Draw:** Represent each data point as a leaf and join the clusters with lines based on the order of merging.

2. Maximum (Complete) Linkage Dendrogram:

1. **Start:** Treat each data point as a separate cluster.

2. **Merge:** Continuously merge the two clusters with the farthest points until all points are in one cluster.

3. **Draw:** Represent each data point as a leaf and join the clusters with lines based on the order of merging.

3. Average Linkage Dendrogram:

1. **Start:** Treat each data point as a separate cluster.

2. **Merge:** Continuously merge the two clusters with the smallest average distance between points until all points are in one cluster.

3. **Draw:** Represent each data point as a leaf and join the clusters with lines based on the order of merging.

24th Jan', 24

Data Visualization Lab 2

Name: M. Gyanada Chowdary

Reg.No: 21BCE7727

Slot : L39+L40

3) Let us consider a hypothetical dataset, where we have a set of points in a 2D space and a similarity (symmetric) matrix indicating the pairwise similarities between these points.

The choice of linkage method (single, complete, average)

Ans:- ① Minimum Distance

	A	B	C	D	E	F
A	0	0.6	0.2	0.4	0.7	0.3
B	0.6	0	0.5	0.3	0.8	0.1
C	0.2	0.5	0	0.07	0.4	0.6
D	0.4	0.3	0.7	0	0.2	0.5
E	0.7	0.8	0.4	0.2	0	0.9
F	0.3	0.1	0.6	0.5	0.9	0

◯ → Min

☐ → Max

Merge (B, F) together

Step 2:

	A	BF	C	D	E
A	0	0.3	0.2	0.4	0.7
BF	0.3	0	0.5	0.3	0.8
C	(0.2)	0.5	0	0.4	0.6
D	0.4	0.3	0.4	0	0.2
E	0.7	0.8	0.6	0.2	0

Step 3: Merge A and C

	AC	BF	D	E
AC	0	0.3	0.4	0.6
BF	0.3	0	0.3	0.8
D	0.4	0.3	0	0.2
E	0.6	0.8	(0.2)	0

Step 4: Merge D and E

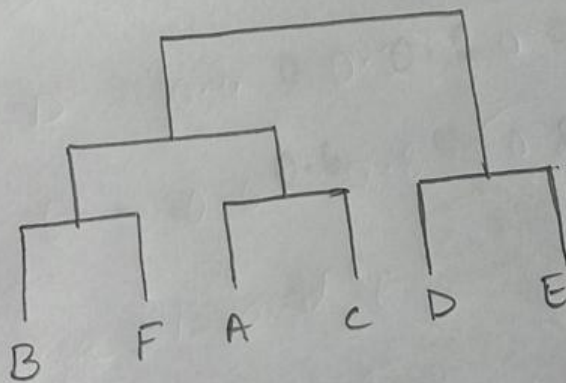|	AC	BF	DE
AC	0	0.2	0.4
BF	(0.2)	0	0.3
DE	0.4	0.3	0

Step 5: Merge AC and BF

	ACBF	DE
ACBF	0	0.3
DE	0.3	0

Step 6: Merge ACBF and DE

ACBFDE

	ACBFDE
ACBFDE	0

Dendrogram:-

Minimum Distance (Simple)

B   F   A   C   D   E

② Maximum Distance (Complete)

Step 2: Merge E and F

	A	B	C	D	EF
A	0	0.6	0.2	0.4	0.7
B	0.6	0	0.5	0.3	0.8
C	0.2	0.5	0	0.7	0.6
D	0.4	0.3	0.7	0	0.5
EF	0.7	(0.8)	0.6	0.5	0

Step 3: Merge B and EF

	A	BEF	C	D
A	0	0.7	0.2	0.4
BEF	0.7	0	0.6	0.5
C	0.2	0.6	0	0.7
D	0.4	0.5	(0.7)	0

Step 4: Merge C and D

	A	BEF	CD
A	0	0.7	0.4
BEF	(0.7)	0	0.6
CD	0.4	0.6	0

Step 5: Merge A and BEF
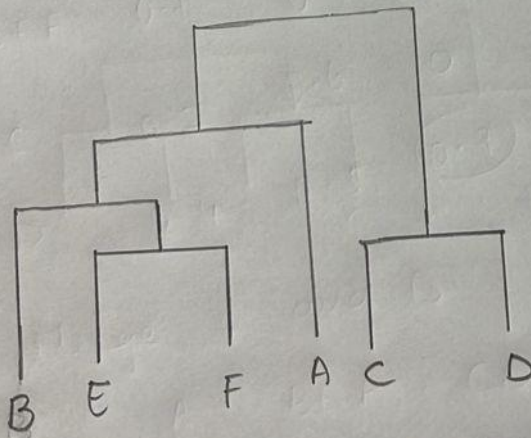
	ABEF	CD
ABEF	0	0.6
CD	0.6	0

Step 6: Merge ABEF and CD

	ABEFCD
ABEFCD	0

Dendrogram :

Maximum Distance [Complete]

Average Distance

Step 2: Merge B,F [Consider Min, smallest Avg
                                    Distance value]

	A	BF	C	D	E
A	0	0.45	0.2	0.4	0.7
BF	0.45	0	0.55	0.4	0.85
C	(0.2)	0.55	0	0.7	0.4
D	0.4	0.4	0.7	0	0.2
E	0.7	0.85	0.4	0.2	0

Step 3: Merge A and C

	AC	BF	D	E
AC	0	0.5	0.55	0.55
BF	0.5	0	0.4	0.85
D	0.55	0.4	0	0.2
E	0.55	0.85	(0.2)	0

Step 4: Merge D and E

	AC	BF	DE
AC	0	0.5	0.55
BF	(0.5)	0	0.625
DE	0.55	0.625	0

Step 5: Merge AC and BF

	ACBF	DE
ACBF	0	0.5875
DE	(0.5875)	0

Step 6: Merge ACBF and DE
                    (and)

	ACBFDE
ACBFDE	0

Dendrogram:

[Average Distance]

B   F   A   C   D   E