WEEK 10

PROBLEM DEFINATION:

CLUSTERING MODEL

e. Clustering algorithms for unsupervised classification.

Plot the cluster data using R visualizations

SOURCE CODE:

1. Clustering algorithms for unsupervised classification.

```
library(cluster)

> set.seed(20)
> irisCluster <- kmeans(iris[, 3:4], 3, nstart = 20)
# nstart = 20. This means that R will try 20 different random starting assignments and then select the one
with the lowest within cluster variation.
> irisCluster
```

OUTPUT:

```
  Petal.Length Petal.Width
1     1.462000    0.246000
2     4.269231    1.342308
3     5.595833    2.037500
```

Clustering vector:

```
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [42] 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2
 [83] 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3
[124] 3 3 3 2 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3
```

Within cluster sum of squares by cluster:

```
[1]  2.02200 13.05769 16.29167
 (between_SS / total_SS =  94.3 %)
```

Available components:

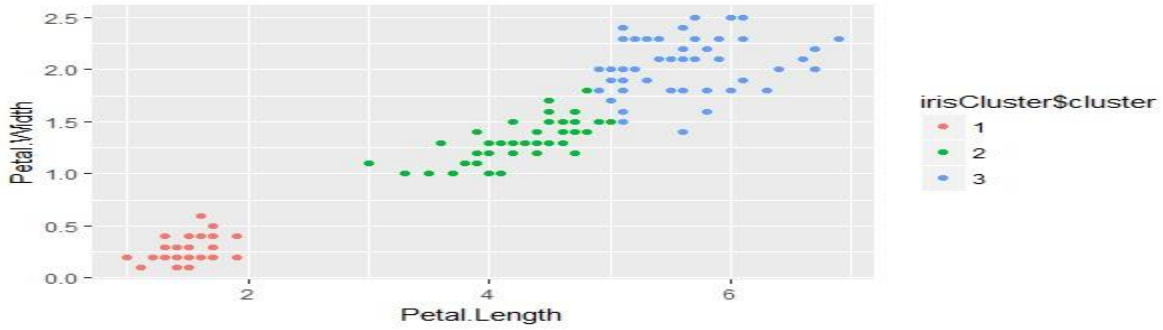
```
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"

[6] "betweenss"    "size"         "iter"         "ifault"
```

SOURCE CODE:

```
> irisCluster$cluster <- as.factor(irisCluster$cluster)
```

> ggplot(iris, aes(Petal.Length, Petal.Width, color = irisCluster$cluster)) + geom_point()

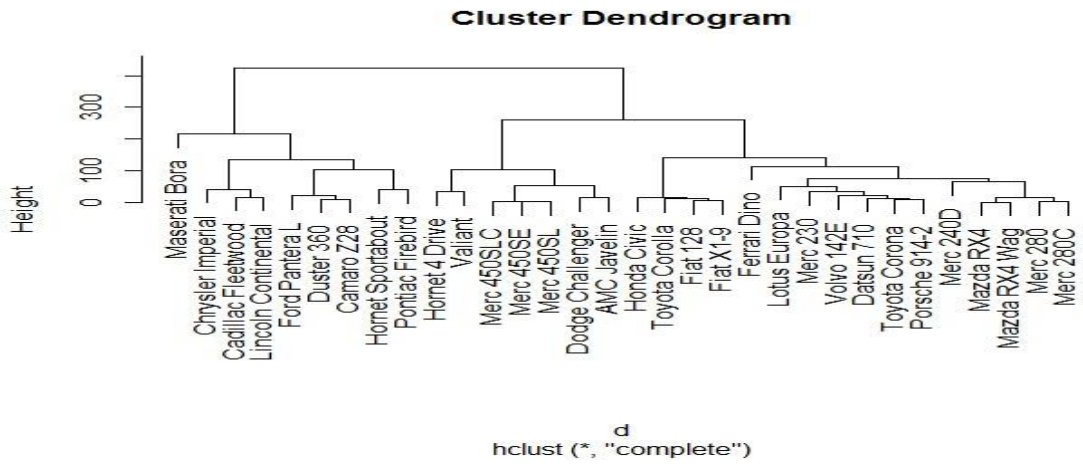
OUTPUT:



SOURCE CODE:

```
> d <- dist(as.matrix(mtcars))   # find distance matrix
> hc <- hclust(d)                # apply hirarchical clustering
> plot(hc)                       # plot the dendrogram
```

OUTPUT:

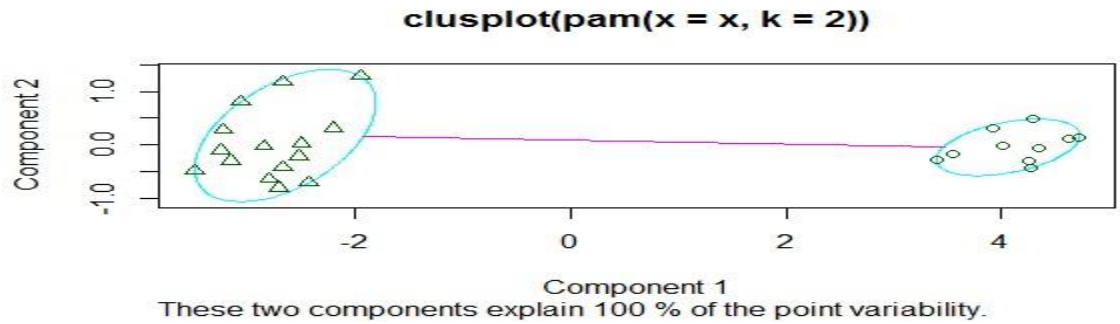


2. Plot the cluster data using R visualizations.

SOURCE CODE:

```
## generate 25 objects, divided into 2 clusters.
x <- rbind(cbind(rnorm(10,0,0.5), rnorm(10,0,0.5)),
cbind(rnorm(15,5,0.5), rnorm(15,5,0.5)))
clusplot(pam(x, 2))
```

OUTPUT:



SOURCE CODE:

```
## add noise, and try again :

x4 <- cbind(x, rnorm(25), rnorm(25))
clusplot(pam(x4, 2))
```

OUTPUT: