EX.NO: 10	K-MEANS ALGORITHM
DATE: 20-10-2022	

Aim:

To implement k-means Algorithm using R programming.

Description:

ALGORITHM:

Step 1: Choose groups in the feature plan randomly

Step 2: Minimize the distance between the cluster center and the different observations (**centroid**). It results in groups with observations

Step 3: Shift the initial centroid to the mean of the coordinates within a group.

Step 4: Minimize the distance according to the new centroids. New boundaries are created. Thus, observations will move from one group to another

Step 5:Repeat until no observation changes groups

K-means usually takes the Euclidean distance between the feature and feature :

$$distance(x, y) = \sum_{i}^{n} (x_i - y_i)^2$$

PROGRAM:

```
library(ClusterR)
library(cluster)
iris_1 <- iris[,-5]
View(iris_1)
```

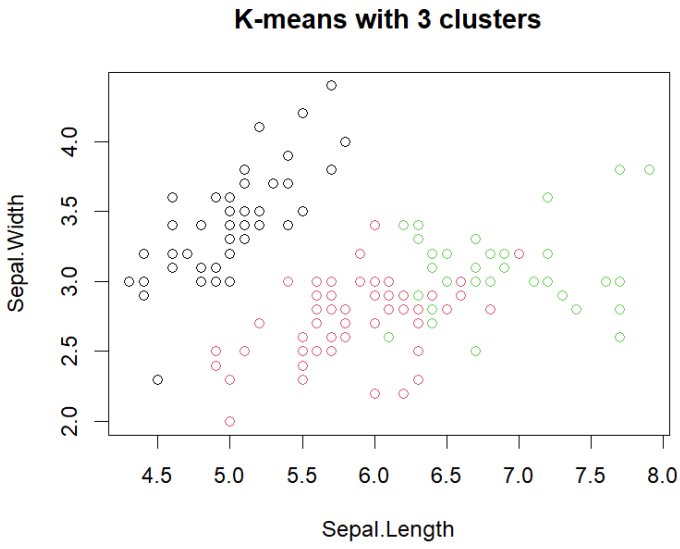
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1

```r
set.seed(240)
kmeans.re <- kmeans(iris_1, centers = 3, nstart = 20)
kmeans.re
```

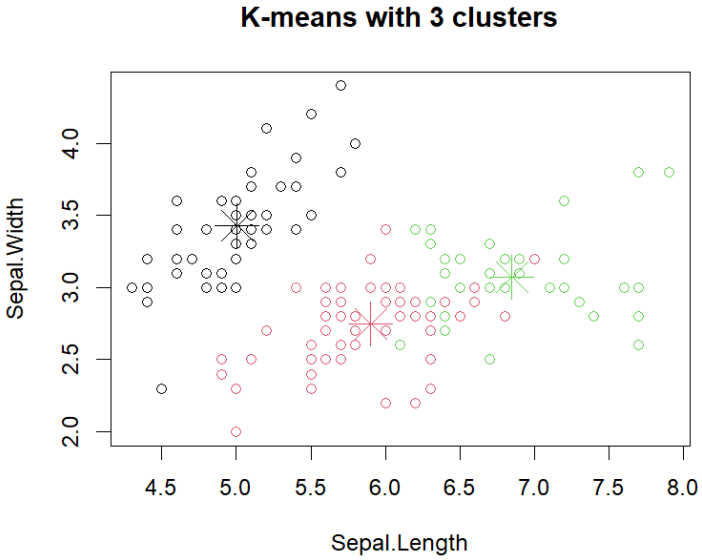
```
Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

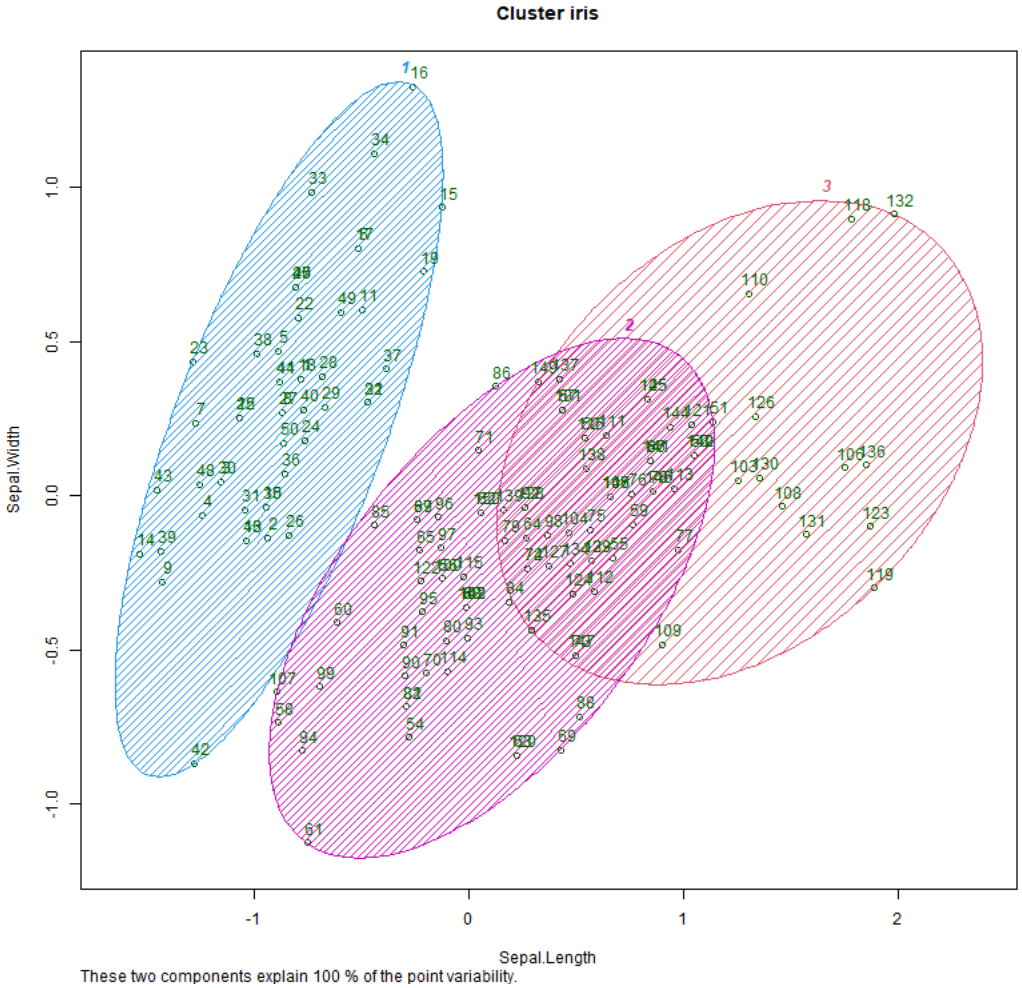
```r
plot(iris_1[c("Sepal.Length", "Sepal.Width")],
     col = kmeans.re$cluster,
     main = "K-means with 3 clusters")
```




```r
kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")]
points(kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")],
       col = 1:3, pch = 8, cex = 3)
```



```
y_kmeans <- kmeans.re$cluster
clusplot(iris_1[, c("Sepal.Length", "Sepal.Width")],
         y_kmeans,
         lines = 0,
         shade = TRUE,
         color = TRUE,
         labels = 2,
         plotchar = FALSE,
         span = TRUE,
         main = paste("Cluster iris"),
         xlab = 'Sepal.Length',
         ylab = 'Sepal.Width')
```



RESULT:

Thus, successfully completed K-means algorithm using r programming