NAME: *Hariprasad K K*

REG NO: *19BCE7079*

COURSE : *DATA_ANALYTICS*

DATE : *13-11-2021*

TOPIC : K-means clustering on iris dataset

loading data

#structure

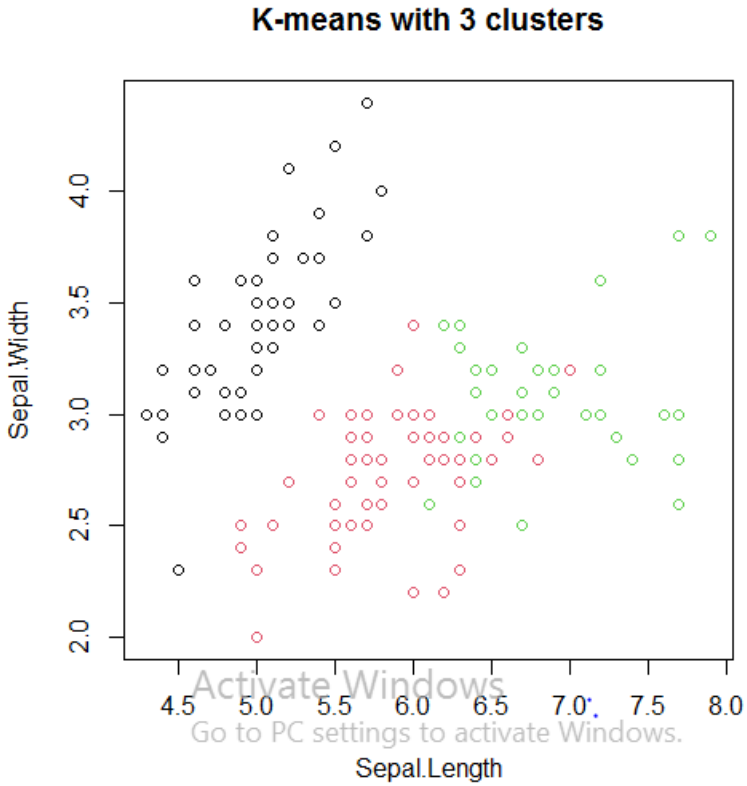
```
>
> data(iris)
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1
 1 ...
>
```

```
The downloaded binary packages are in
        C:\Users\lenovo\AppData\Local\Temp\RtmpA5tLAb\downloaded_packages
> library(ClusterR)
Loading required package: gtools
> library(cluster)
>
> iris_1 <- iris[, -5]
>
> set.seed(240)
>
> kmeans.re <- kmeans(iris_1, centers = 3, nstart = 20)
> kmeans.re
K-means clustering with 3 clusters of sizes 50, 62, 38

Cluster means:
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1     5.006000    3.428000     1.462000    0.246000
2     5.901613    2.748387     4.393548    1.433871
3     6.850000    3.073684     5.742105    2.071053

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
 [28] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 2
 [55] 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
 [82] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 3 3 3 3 2 3
[109] 3 3 3 3 3 2 2 3 3 3 3 2 3 2 3 2 3 3 2 2 3 3 3 3 3 2 3
[136] 3 3 3 2 3 3 3 2 3 3 3 2 3 3 2

Within cluster sum of squares by cluster:
[1] 15.15100 39.82097 23.87947
 (between_SS / total_SS =  88.4 %)

Available components:

[1] "cluster"      "centers"      "totss"
[4] "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"
> |
```

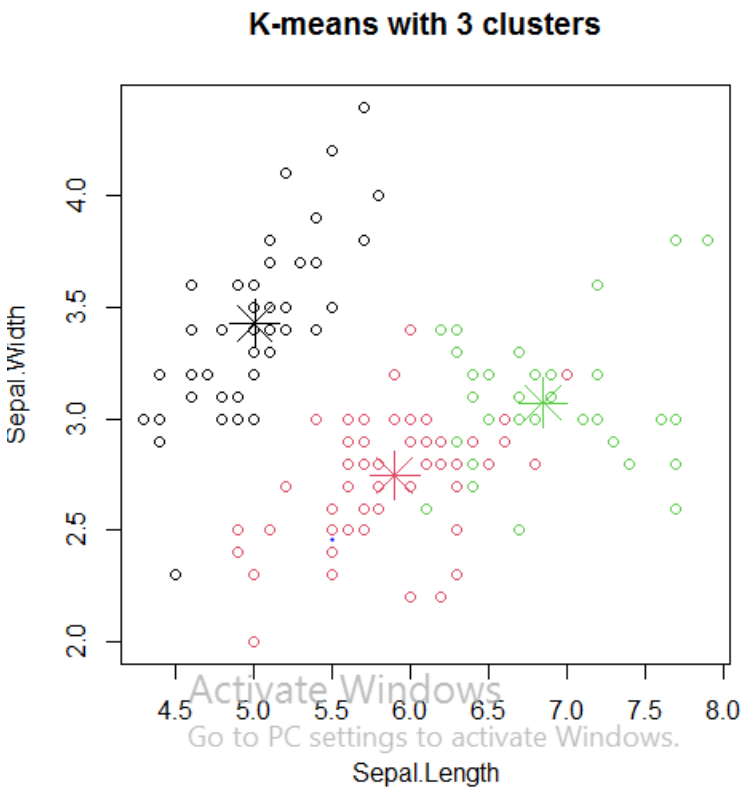
```
> kmeans.re$cluster
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [37] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 [73] 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 3 3 3 3 2 3
[109] 3 3 3 3 3 2 2 3 3 3 3 2 3 2 3 2 3 3 2 2 3 3 3 3 3 2 3 3 3 2 3 3 3 2 3
[145] 3 3 2 3 3 2
>
> |
```

```
> cm <- table(iris$Species, kmeans.re$cluster)
> cm

             1  2  3
  setosa     50  0  0
  versicolor  0 48  2
  virginica   0 14 36
> |
```

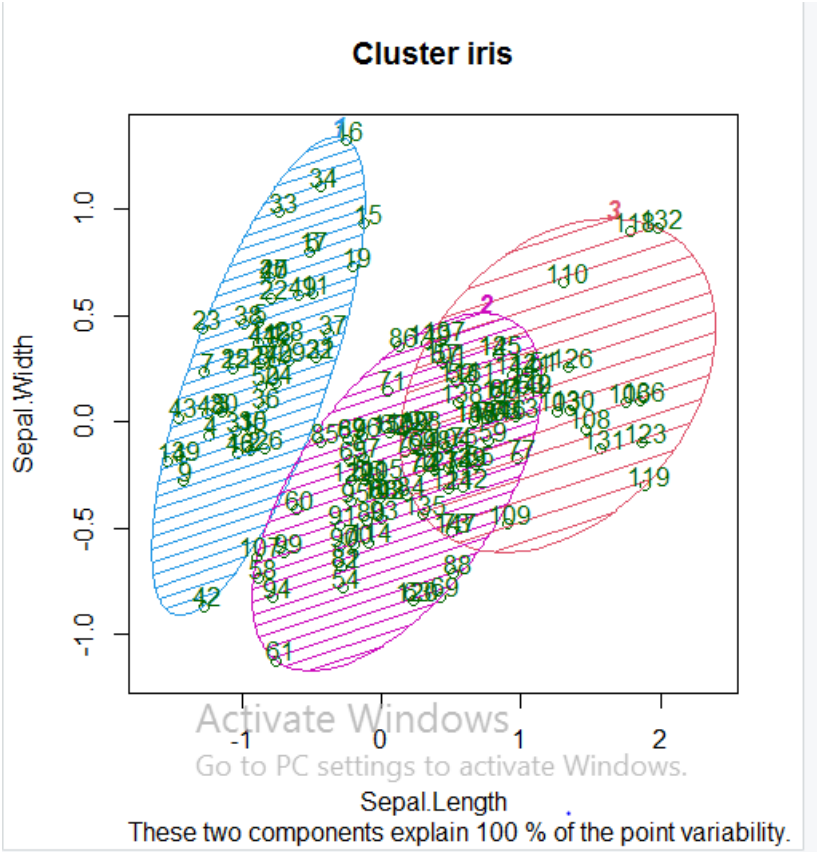
```
>
> plot(iris_1[c("Sepal.Length", "Sepal.Width")])
> plot(iris_1[c("Sepal.Length", "Sepal.Width")],
+      col = kmeans.re$cluster)
> plot(iris_1[c("Sepal.Length", "Sepal.Width")],
+      col = kmeans.re$cluster,
+      main = "K-means with 3 clusters")
>
> |
```




```
> kmeans.re$centers
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1     5.006000    3.428000     1.462000    0.246000
2     5.901613    2.748387     4.393548    1.433871
3     6.850000    3.073684     5.742105    2.071053
> kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")]
  Sepal.Length Sepal.Width
1     5.006000    3.428000
2     5.901613    2.748387
3     6.850000    3.073684
>
```




```
> 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')
>
```


Cluster iris
Sepal.Width
Sepal.Length
1.0
0.5
0.0
-0.5
-1.0
-1
0
1
2
Activate Windows
Go to PC settings to activate Windows.
These two components explain 100 % of the point variability.