

R Notebook

```
# Loading data
data(iris)

# Structure
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 ...

# Installing Packages
install.packages("ClusterR")

## Installing package into '/home/6b/R/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)

install.packages("cluster")

## Installing package into '/home/6b/R/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)

# Loading package
library(ClusterR)
library(cluster)

# Removing initial label of
# Species from original dataset

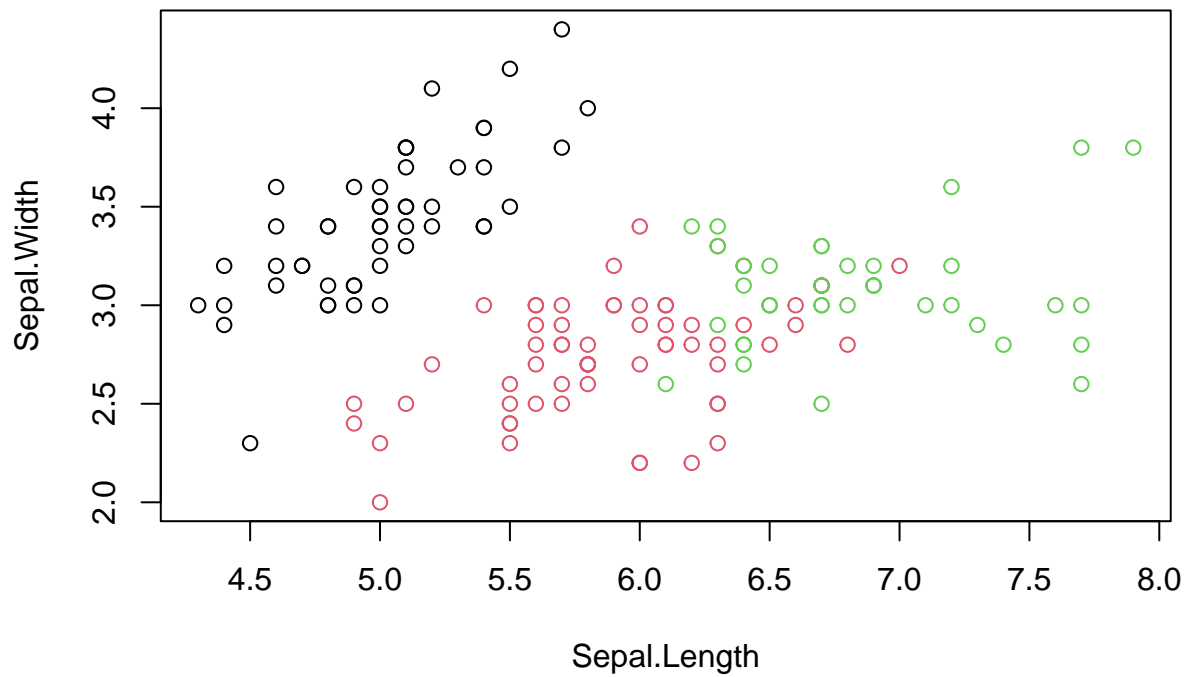
iris_1 <- iris[, -5]

# Fitting K-Means clustering Model
# to training dataset
set.seed(240) # Setting seed
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 1 1 1 1 1 1
##  [38] 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
```

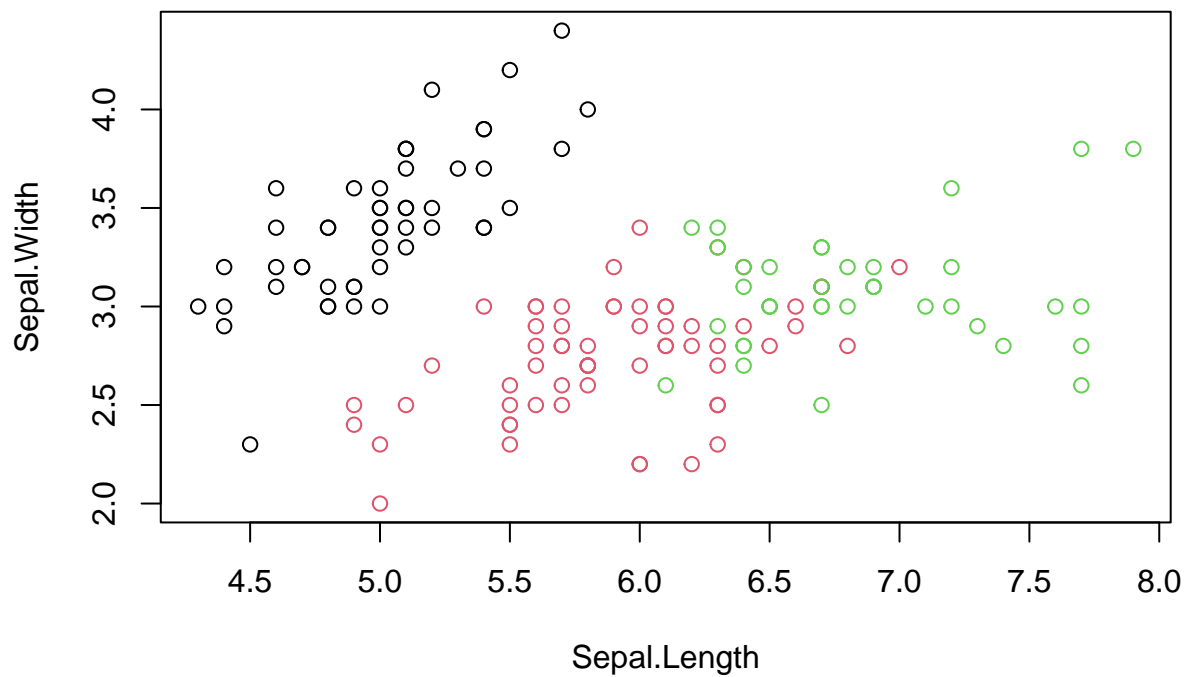


```
plot(iris_1[c("Sepal.Length", "Sepal.Width")],
     col = kmeans.re$cluster)
```



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

K-means with 3 clusters



```
## Plotting cluster centers
```

```
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
```

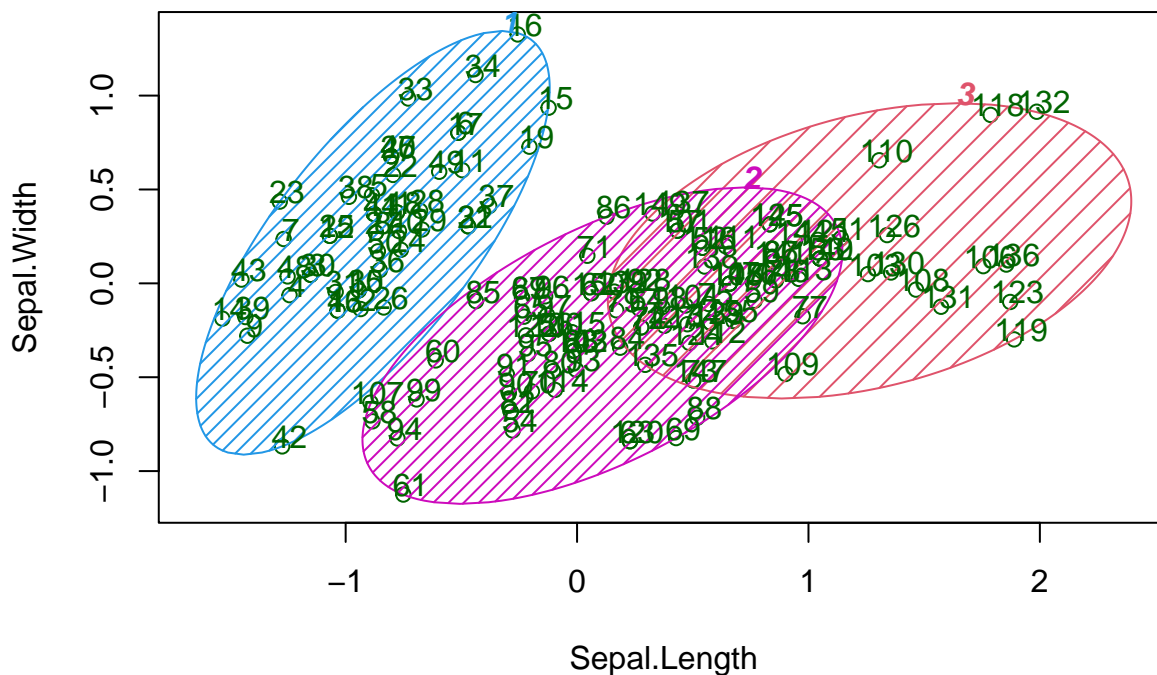
```
kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")]
```

```
##   Sepal.Length Sepal.Width
## 1      5.006000      3.428000
## 2      5.901613      2.748387
## 3      6.850000      3.073684
```

```
## Visualizing clusters
```

```
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



These two components explain 100 % of the point variability.

Code contributed by Manja