

Tanmay Yogesh Raykar  
2193266  
BDA Lab Assignment 3

Code:

1. Installation and cluster plotting:

```
# Installing Packages
```

```
install.packages("ClusterR")
```

```
install.packages("cluster")
```

```
# 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 = 2, nstart = 20)
```

```
kmeans.re
```

```
# Cluster identification for
```

```
# each observation
```

```
kmeans.re$cluster
```

```
# Confusion Matrix
```

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

```
# Model Evaluation and visualization
```

```
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")
```

```
## Plotting cluster centers
```

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

```
# cex is font size, pch is symbol
```

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

```
## 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 = TRUE,  
span = TRUE,  
main = paste("Cluster iris"),  
xlab = 'Sepal.Length',  
ylab = 'Sepal.Width')
```

## 2. Elbow Method

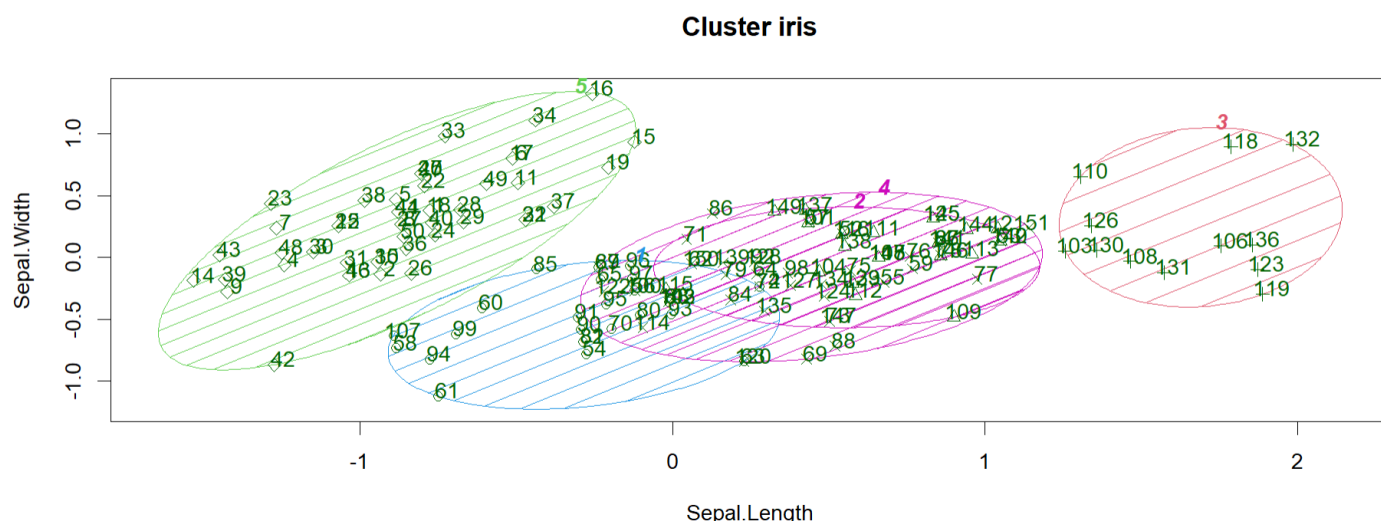
```
#Elbow Method for finding the optimal number of clusters  
set.seed(123)  
# Compute and plot wss for k = 2 to k = 15.  
k.max <- 15  
data <- iris_1  
wss <- sapply(1:k.max,  
              function(k){kmeans(data, k, nstart=50,iter.max = 15  
)$tot.withinss})  
wss  
plot(1:k.max, wss,  
     type="b", pch = 19, frame = FALSE,  
     xlab="Number of clusters K",  
     ylab="Total within-clusters sum of squares")
```

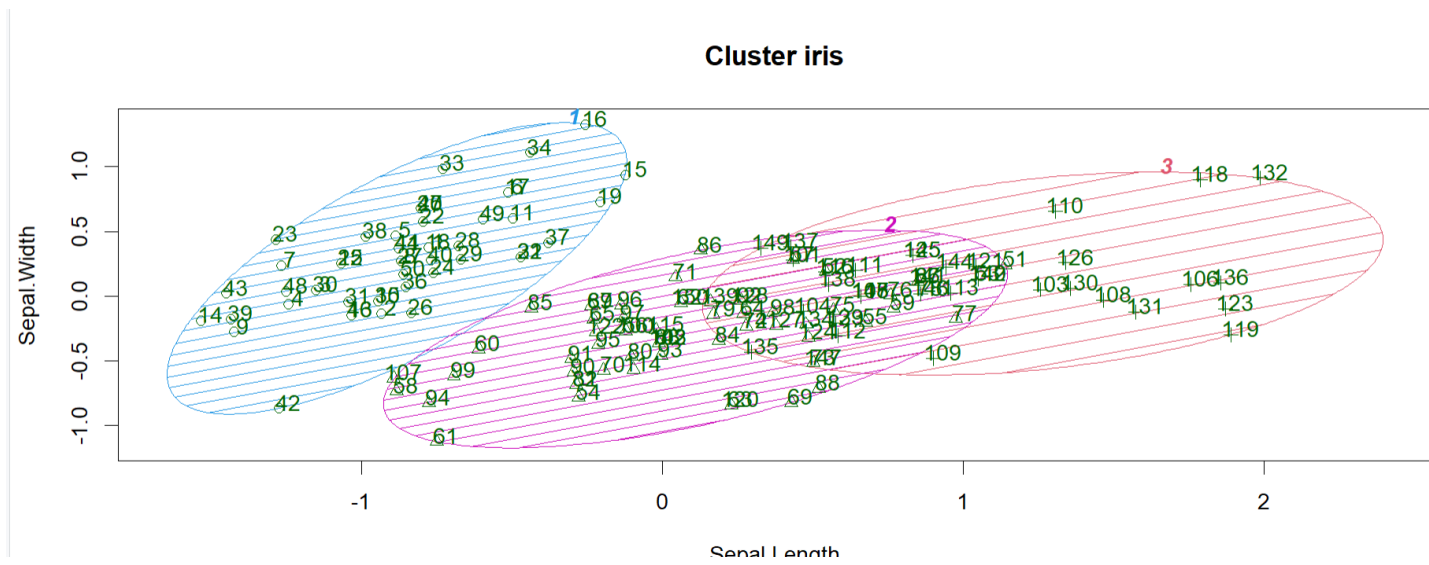
## Output:

```
R 4.2.1 · ~/
> points(kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")],
+        col = 1:3, pch = 8, cex = 3)
>
> ## visualizing clusters
> y_kmeans <- kmeans.re$cluster
> clusplot(iris_1[, c("Sepal.Length", "Sepal.Width")],
+          y_kmeans,
+          lines = 0,
+          shade = TRUE,
+          color = TRUE,
+          labels = 2,
+          plotchar = TRUE,
+          span = TRUE,
+          main = paste("Cluster iris"),
+          xlab = 'Sepal.Length',
+          ylab = 'Sepal.Width')
> #Elbow Method for finding the optimal number of clusters
> set.seed(123)
> # Compute and plot wss for k = 2 to k = 15.
> k.max <- 15
> data <- iris_1
> wss <- sapply(1:k.max,
+              function(k){kmeans(data, k, nstart=50, iter.max = 15 )$tot.withinss})
> wss
[1] 681.37060 152.34795 78.85144 57.22847 46.44618 39.03999
[7] 34.29823 29.98894 27.78609 25.83405 24.03282 22.39425
[13] 21.20467 20.07429 18.78016
> plot(1:k.max, wss,
```

```

> #Elbow Method for finding the optimal number of clusters
> set.seed(123)
> # Compute and plot wss for k = 2 to k = 15.
> k.max <- 15
> data <- iris_1
> wss <- sapply(1:k.max,
+              function(k){kmeans(data, k, nstart=50, iter.max = 15 )$tot.withinss})
> wss
[1] 681.37060 152.34795 78.85144 57.22847 46.44618 39.03999
[7] 34.29823 29.98894 27.78609 25.83405 24.03282 22.39425
[13] 21.20467 20.07429 18.78016
> plot(1:k.max, wss,
+      type="b", pch = 19, frame = FALSE,
+      xlab="Number of clusters K",
+      ylab="Total within-clusters sum of squares")
> [1] 2005.0000 1635.8573 1416.7041 1253.9959 1115.4657 1026.0506 952.4835 887.7202
Error: unexpected '[' in "["
> |
```





Elbow Method graph:

