Task 2

Hrithik Agarwal

09/01/2021

The Sparks Foundation

Task 2 (Prediction using Unsupervised ML)

Loading the required libraries

```
library(factoextra)
library(NbClust)
library(cluster)
```

Importing the data

```
data <- read.csv('Iris.csv')
```

```
str(data)
```

```
## 'data.frame':    150 obs. of  6 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ SepalLengthCm: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ SepalWidthCm : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ PetalLengthCm: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ PetalWidthCm : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species      : chr  "Iris-setosa" "Iris-setosa" "Iris-setosa" "Iris-setosa" ...
```

Now we have to do **K-means Clustering** so we will remove the *Species* column.
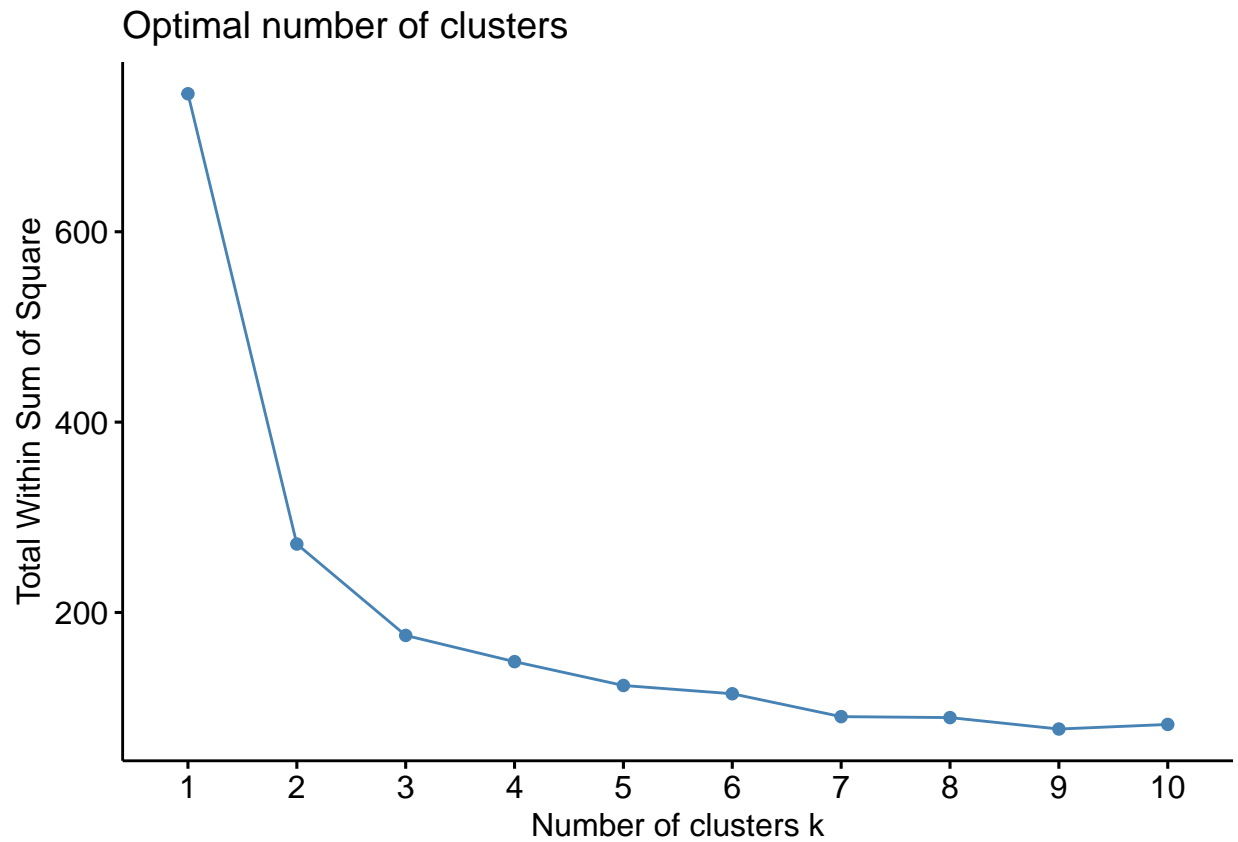
```
d1 <- data[,-6]
```

Scaling the data as the different columns has data at different levels.

```
d1 <- scale(d1)
```

Finding Optimum number of clusters.

For this we will use th elbow method and use **fviz_nbclust** function.

```
fviz_nbclust(d1, kmeans, method = "wss")
```

Optimal number of clusters

From the Graph obtained we see the elbow is at $\mathbf{k = 3}$, so we will take number of centroids as 3.
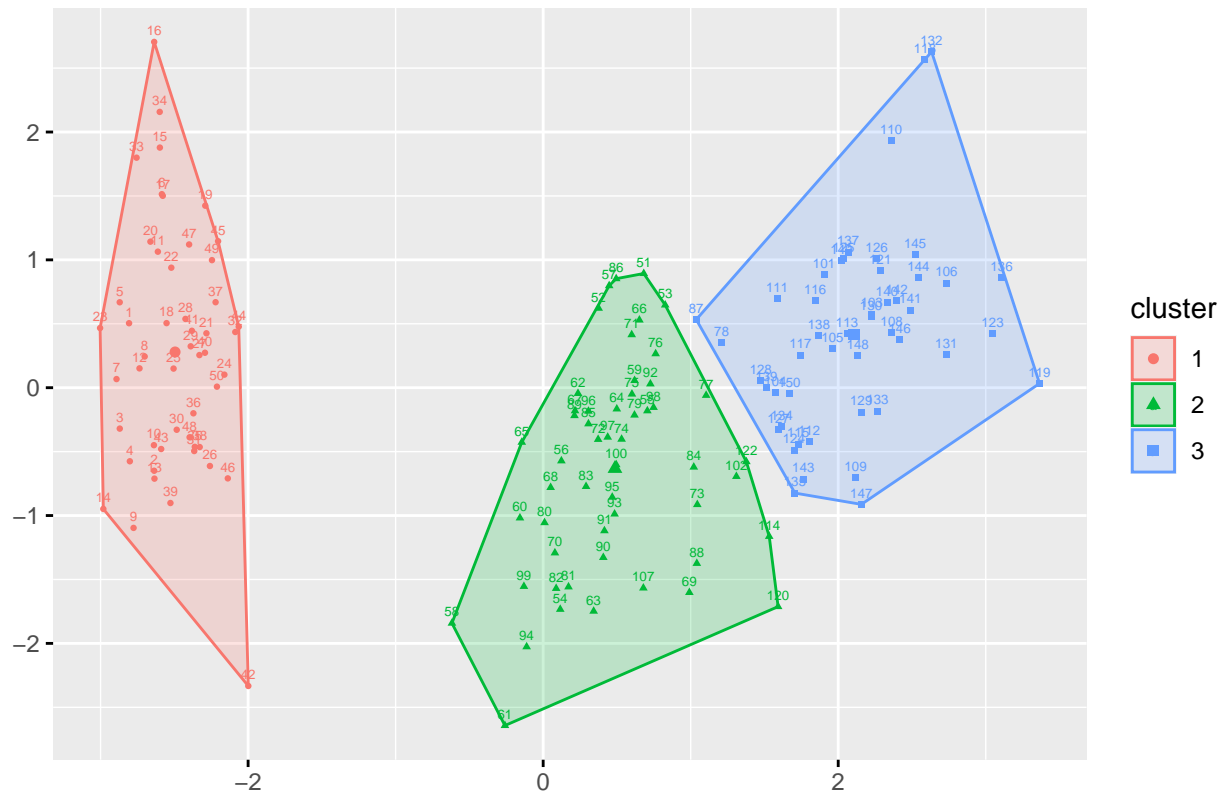
Clustering

```
m <- kmeans(d1 , centers = 3 , nstart = 200 , iter.max = 5)
```

Plotting

```
fviz_cluster(m, d1,pointsize = 0.8 , labelsize = 5 , outlier.color = "black" , xlab = F , ylab = F, shou
```

Cluster plot

This is our Cluster plot.