

Code and Output

```
> rm(list=ls()) #Clear the environment
> setwd("E:/Dropbox/RU DataScience/MSDS650/week7/Assignment") #Set working directory for the assignment
> getwd() #Check working directory
[1] "E:/Dropbox/RU DataScience/MSDS650/week7/Assignment"
> data(iris) #include iris dataset
> head(iris) #view the first few rows of data
  Sepal.Length Sepal.width Petal.Length Petal.width Species
1          5.1         3.5         1.4         0.2  setosa
2          4.9         3.0         1.4         0.2  setosa
3          4.7         3.2         1.3         0.2  setosa
4          4.6         3.1         1.5         0.2  setosa
5          5.0         3.6         1.4         0.2  setosa
6          5.4         3.9         1.7         0.4  setosa
> set.seed(42) #set seed for the algorithm to ensure reproducible results
> km <-kmeans(iris[,1:4], 3, nstart=25) #chose 3 clusters as there are three species in the dataset
> km #output results
K-means clustering with 3 clusters of sizes 38, 62, 50
```

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

Clustering vector:

```
[1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[69] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 2 1 2 1 2 1 1 2 2 1 1 1 1 1 2 1 1
[137] 1 1 2 1 1 1 2 1 1 1 2 1 1 2
```

within cluster sum of squares by cluster:

```
[1] 23.87947 39.82097 15.15100
(between_SS / total_SS = 88.4 %)
```

Available components:

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

>

```
> #Compare the clusters with the species and plot results
> table(km$cluster, iris$Species)
```

	setosa	versicolor	virginica
1	0	2	36
2	0	48	14
3	50	0	0

```
> table(km$cluster, iris$Species)
```

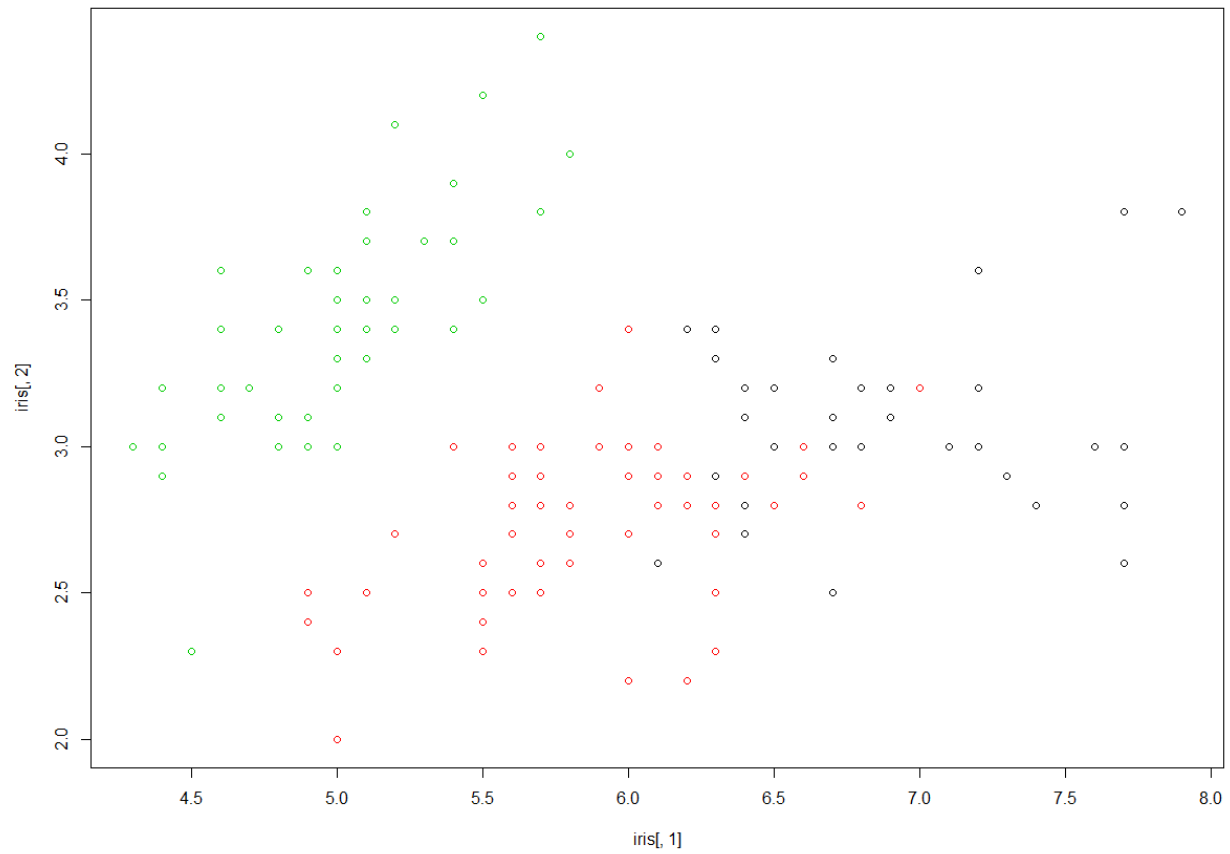
	setosa	versicolor	virginica
1	0	2	36
2	0	48	14
3	50	0	0

```
> |
```

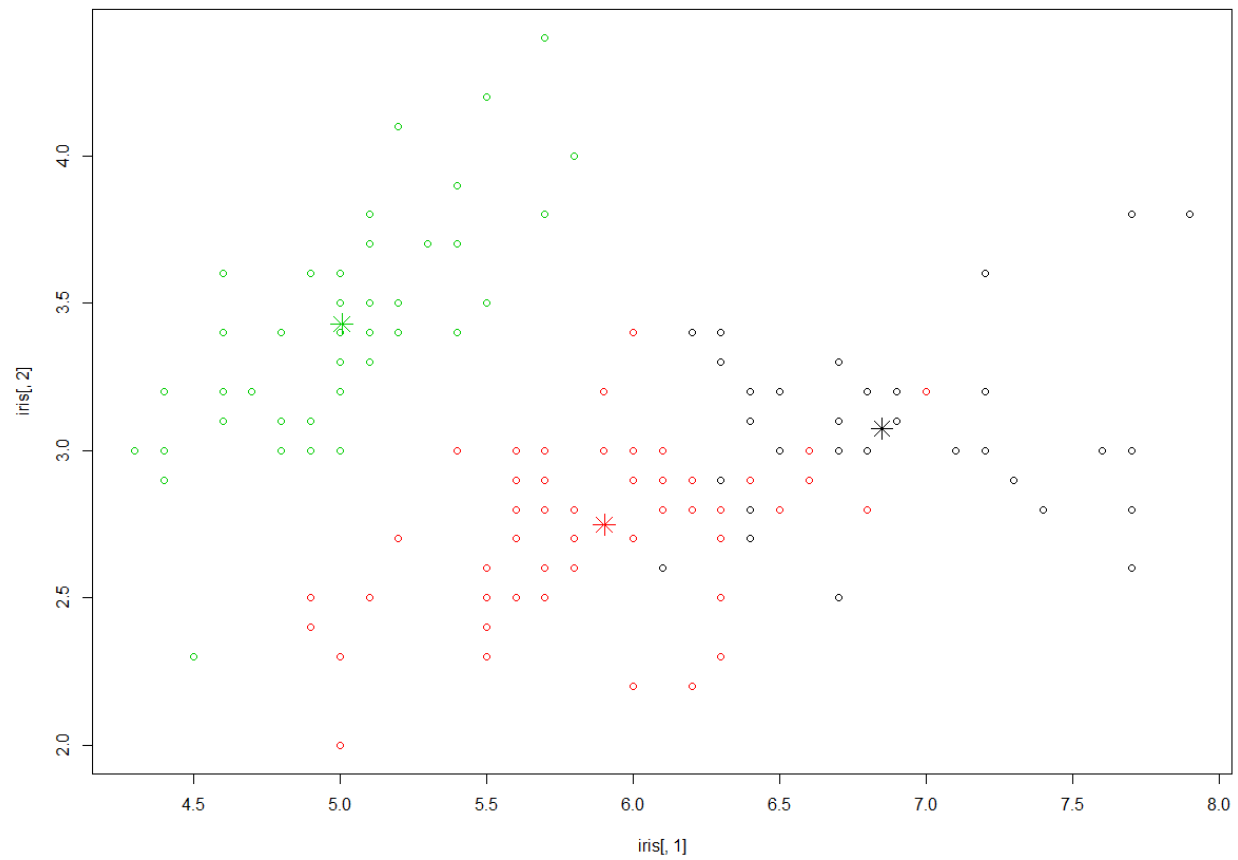
```
> #Plot the results
```

```
> plot(iris[,1], iris[,2], col=km$cluster) #Plot cluster centers by Sepal Length by Sepal width
```

>



```
> #add center points  
> points(km$centers[,c(1,2)], col=1:3, pch=8, cex=2)  
>
```



```
#Plot Petal length and width
> plot(iris[,3], iris[,4], col=km$cluster)
> #add center points
> points(km$centers[,c(3,4)], col=1:3, pch=8, cex=2)
```

