# IRIS Example

**1) Importing libraries**

```
library(datasets) ## Loads some popular datasets
library(ggplot2) ## For visualization
```

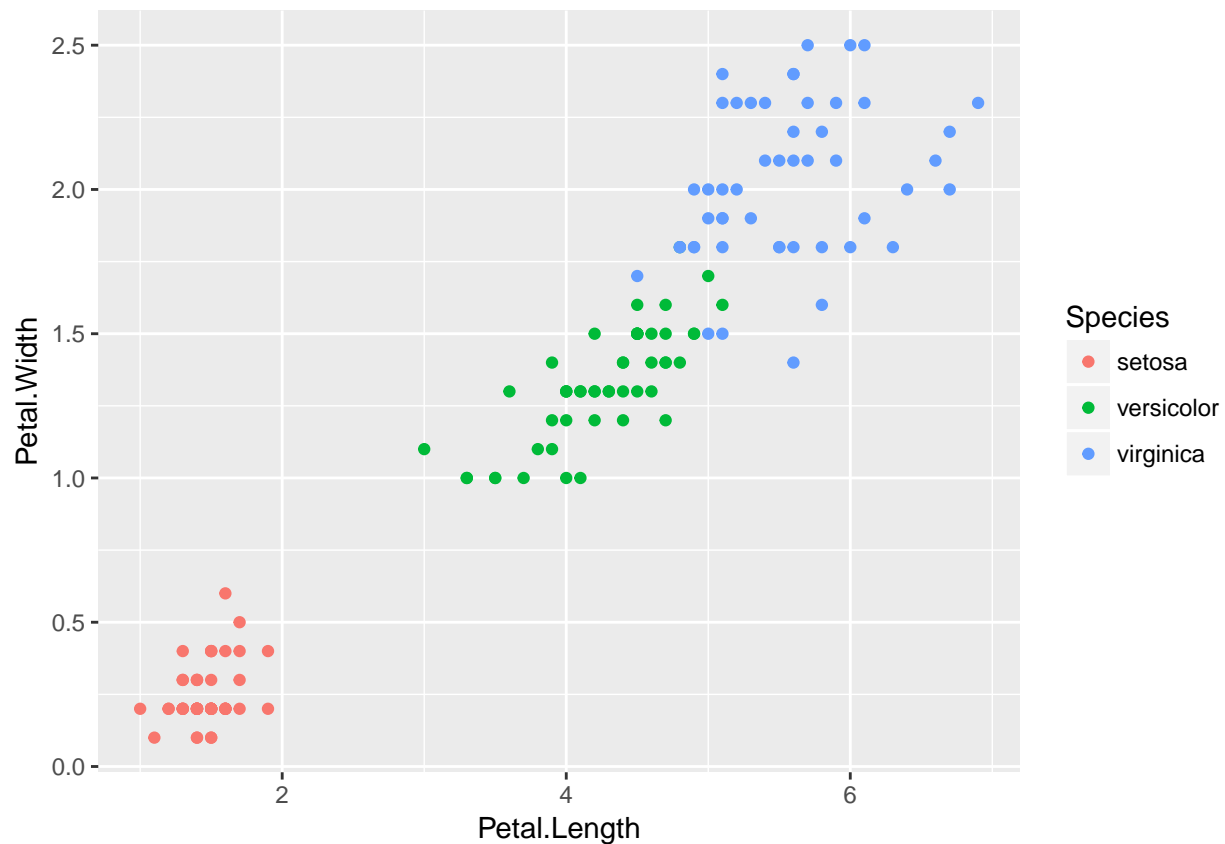**2) Looking at the data**

Having a sample look at few rows in data

```
head(iris)
```

```
##   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
```
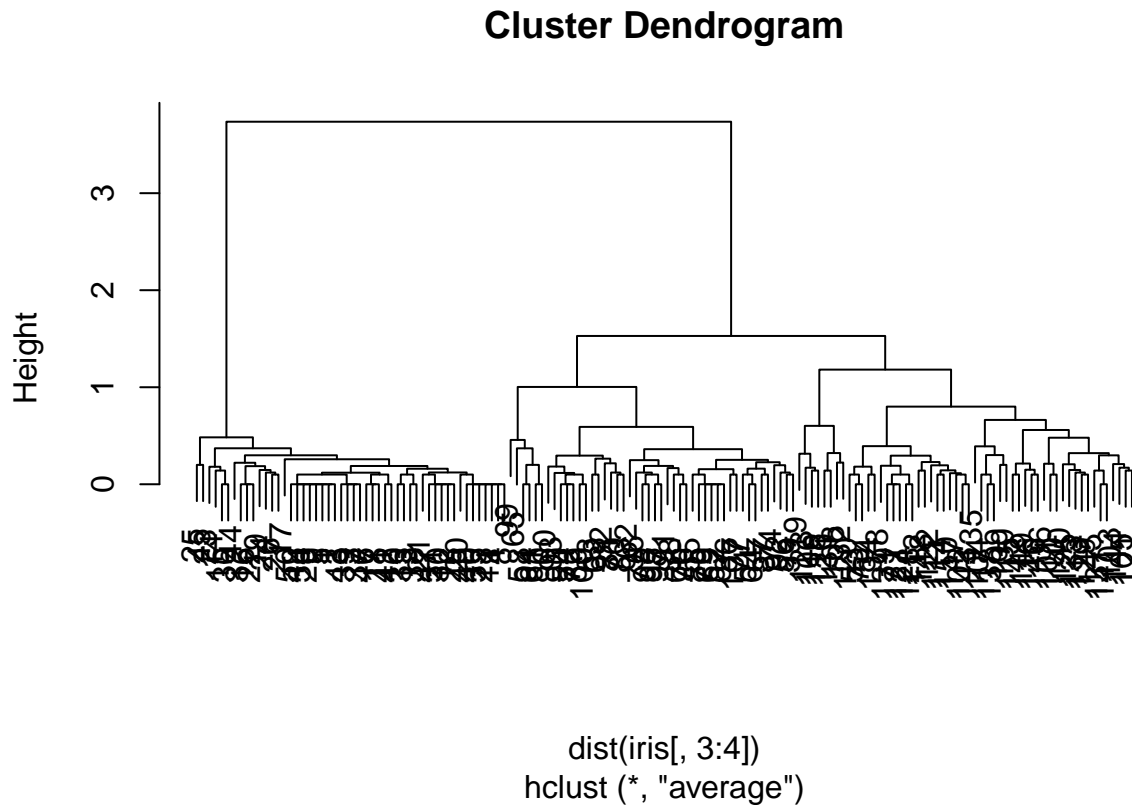
Plotting data on Petal length and width with labels to see the structure

```
ggplot(iris, aes(Petal.Length, Petal.Width, color = Species)) + geom_point()
```

**3) Running Hierarcical clustering**

```
clusters <- hclust(dist(iris[, 3:4]),method = 'average') ## Hclust
plot(clusters) ## Looking at the dendogram output
```

# Cluster Dendrogram

dist(iris[, 3:4])
hclust (*, "average")

Looks like 3 or 5 cluster solution will work in this case. Because we know there are only 3 species let's go by a 3 cluster solution

```
## Looks like 3 or 4 clusters looks fine
clusterCut <- cutree(clusters, 3) ## Getting 3 cluster solution
```
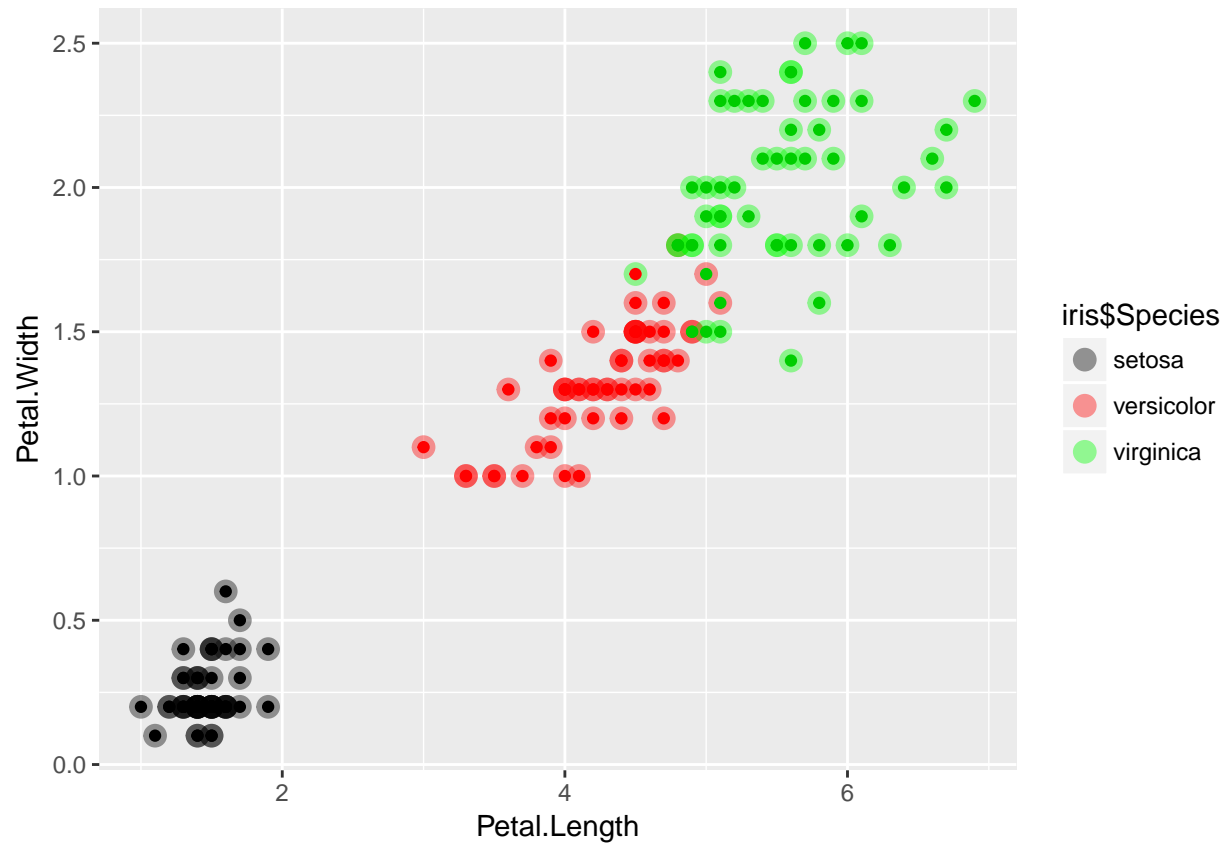
**4) Checking accuracy of the clusters**

```
## Checking accuracy of cluster
table(clusterCut, iris$Species)

##
## clusterCut setosa versicolor virginica
##          1     50          0         0
##          2      0         45         1
##          3      0          5        49
## Visualizing clusters
ggplot(iris, aes(Petal.Length, Petal.Width, color = iris$Species)) +
  geom_point(alpha = 0.4, size = 3.5) + geom_point(col = clusterCut) +
  scale_color_manual(values = c('black', 'red', 'green'))
```

Looks like the clustering is doing a pretty good job in determining the flower specie except of some borderline cases.