**#-----------importing data**

**iris <- read.csv("Iris.csv")**

**#-----------VIsualising the dataset**

**library(ggplot2)**

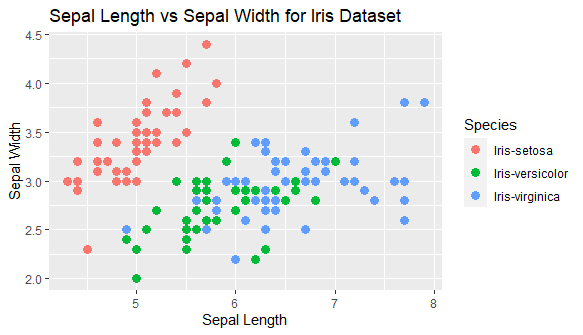
**#for sepal length and width**

**ggplot(data = iris)+**

**geom\_point(aes(x=SepalLengthCm, y= SepalWidthCm, col= Species), size= 2.8)+**

**xlab("Sepal Length")+ ylab("Sepal Width")+**

**ggtitle("Sepal Length vs Sepal Width for Iris Dataset")**

****

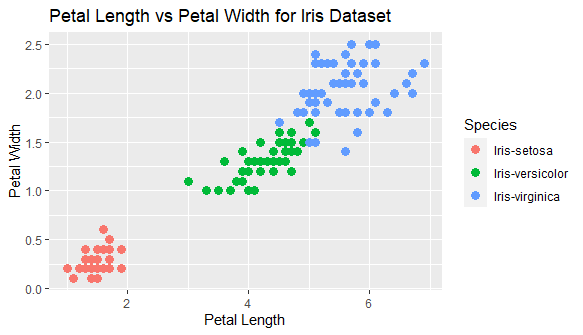
**#for petal length and width**

**ggplot(data = iris)+**

**geom\_point(aes(x=PetalLengthCm, y= PetalWidthCm, col= Species), size=2.8)+**

**xlab("Petal Length")+ ylab("Petal Width")+**

**ggtitle("Petal Length vs Petal Width for Iris Dataset")**

****

**#----------Taking only datas of petal length and width for clustering**

**data<- iris[4:5]**

**#---------Using Elbow Method to find the optimal number of clusters**

**set.seed(78)**

**wcss<- vector()**

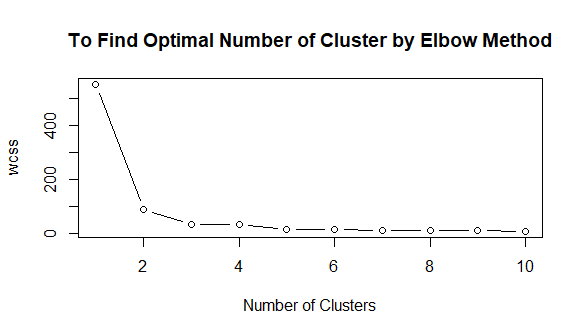
**for(i in 1:10) wcss[i]<- sum(kmeans(data,i)$withinss)**

**plot(1:10, wcss, type = "b",**

**main = paste("To Find Optimal Number of Cluster by Elbow Method"),**

**xlab = "Number of Clusters",**

**ylab = "wcss")**

****

**#----------Applying k-means to the Iris data**

**set.seed(100)**

**kmeans<- kmeans(data, 3, iter.max = 400, nstart = 10)**

**#------------Visualising the clusters**

**library(cluster)**

**clusplot(data, kmeans$cluster,**

**lines = 0, shade = TRUE,**

**color=TRUE, labels=2,**

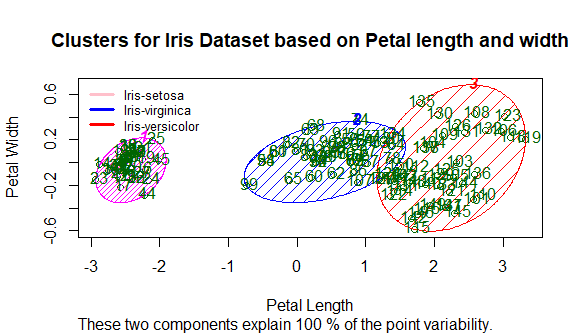
**plotchar= FALSE, span = TRUE,**

**main = paste("Clusters for Iris Dataset based on Petal length and width"),**

**xlab = "Petal Length", ylab="Petal Width")**

**legend("topleft", legend = c("Iris-setosa", "Iris-virginica", "Iris-versicolor"),**

**col=c("pink", "blue", "red"), lty = 1.2, lwd = 3.7, bty = "n", cex = 0.8)**

****