#Import the data in R and check the head

> data<-read.csv("C:\\Users\\User\\Desktop\\GRIP\\20201202\_030540.csv")

> iris\_data<-data[,2:5]

> head(iris\_data)

SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm

1 5.1 3.5 1.4 0.2

2 4.9 3.0 1.4 0.2

3 4.7 3.2 1.3 0.2

4 4.6 3.1 1.5 0.2

5 5.0 3.6 1.4 0.2

6 5.4 3.9 1.7 0.4

#Create a function which calculates the differences with its

cluster centers using k-means

> f<-function(k){

+ within\_diff<-kmeans(iris\_data,k)$tot.withinss

+ }

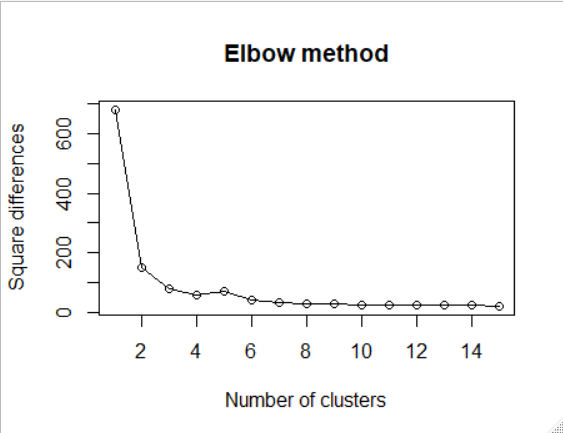
#Plot these differences a to know where there is a kink

> k<-1:15

> plot(k,sapply(k,f),type="o",lwd=1,ylab="Square differences",

+ xlab="Number of clusters",main="Elbow method")

>



#the optimum number if clusters is the point where we can see an elbow shape(kink), which in this case is 3

> iris\_model<-kmeans(iris\_data,3)

> iris\_model

K-means clustering with 3 clusters of sizes 62, 38, 50

Cluster means:

SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm

1 5.901613 2.748387 4.393548 1.433871

2 6.850000 3.073684 5.742105 2.071053

3 5.006000 3.418000 1.464000 0.244000

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

[34] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1

[67] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[100] 1 2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2

[133] 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1

Within cluster sum of squares by cluster:

[1] 39.82097 23.87947 15.24040

(between\_SS / total\_SS = 88.4 %)

Available components:

[1] "cluster" "centers" "totss" "withinss"

[5] "tot.withinss" "betweenss" "size" "iter"

[9] "ifault"

#Make a model using 3 custers

> table(data$Species,iris\_model$cluster)

1 2 3

Iris-setosa 0 0 50

Iris-versicolor 48 2 0

Iris-virginica 14 36 0

> #Plot the clusters

> plot(iris\_data$SepalLengthCm,iris\_data$SepalWidthCm,

+ col=iris\_model$cluster,pch=19,xlab="Sepal Length",

+ ylab="Sepal Width",lwd=2,main="Clusters")

#Plot the centroids

> points(iris\_model$centers[,1:2],col="yellow",pch=19,lwd=2)

#Give appropriate legend

> legend("topright",c("Iris-setosa","Iris-versicolour","Iris-virginica",

+ "Centroids"),col=c(unique(iris\_model$cluster),

+ "yellow"),pch=19)

