Perform Clustering for the crime data and identify the number of clusters formed and draw inferences.

Data Description:

Murder -- Muder rates in different places of United States

Assualt- Assualt rate in different places of United States

UrbanPop - urban population in different places of United States

Rape - Rape rate in different places of United States

> crime <- read.csv(file.choose())

> View(crime)

> norm\_crime <- scale(crime[,-1])

> View(norm\_crime)

> install.packages("animation")

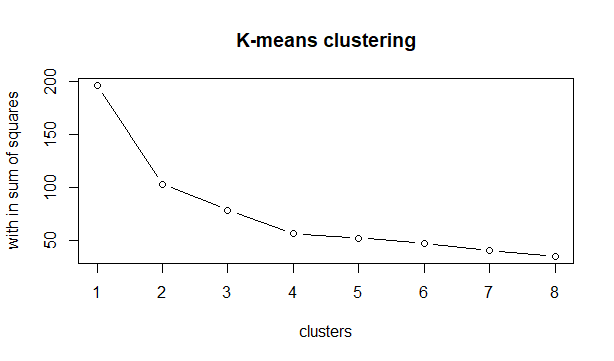
>install.packages(“cluster”)

#finding number of clusters

> wss=(nrow(norm\_crime)-1)\*sum(apply(norm\_crime,2,var))

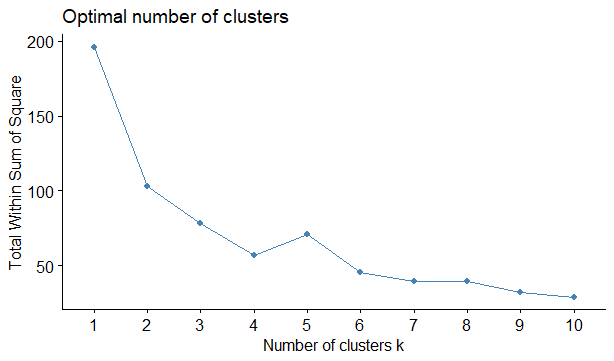
> for(i in 2:8) wss[i]=sum(kmeans(norm\_crime,centers=i)$withinss)

> plot(1:8,wss,type = "b",xlab = "clusters",ylab = "with in sum of squares",main = "K-means clustering")

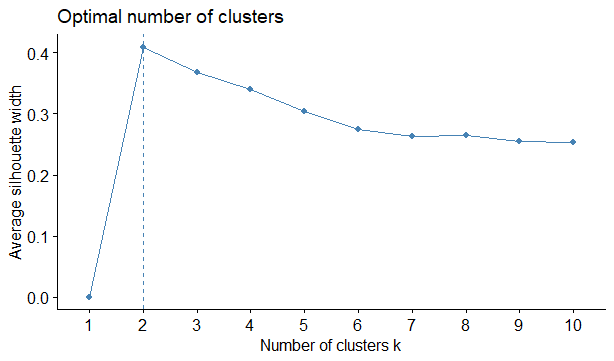
\

>install.packages(“factoextra”)

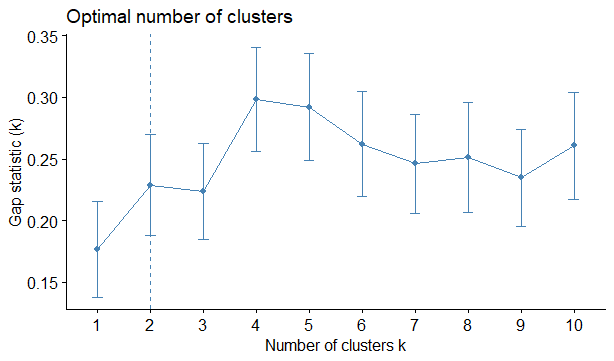
> fviz\_nbclust(norm\_crime,method = 'wss',FUNcluster = kmeans)



> fviz\_nbclust(norm\_crime,method = 'silhouette',FUNcluster = kmeans)

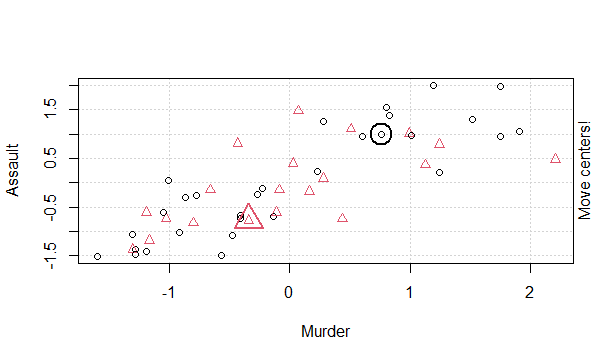


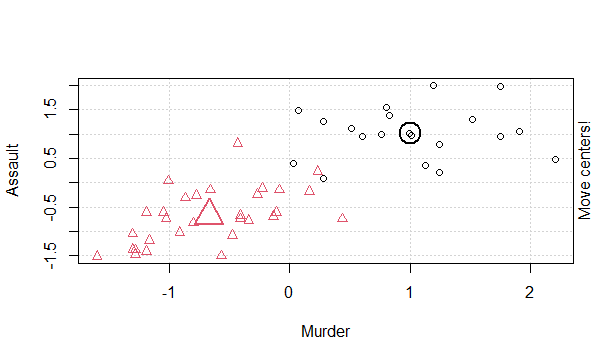
> fviz\_nbclust(norm\_crime,method = 'gap\_stat',FUNcluster = kmeans)

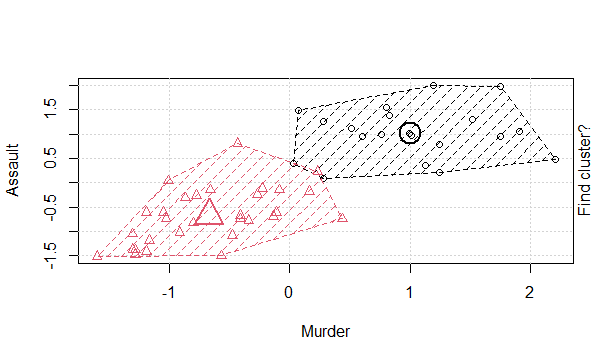
 from all the elbow plots it is clear that optimal number of clusters is 2

final <- kmeans(norm\_crime,2)

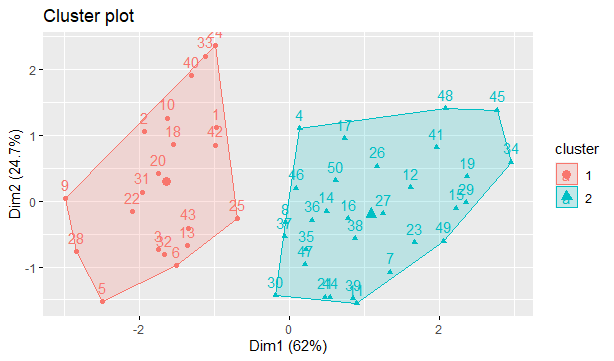
finalanim <- kmeans.ani(norm\_crime,2)







fviz\_cluster(final,data = crime[-1])



> aggregate(crime[,-1],by=list(final$cluster),FUN = mean)

Group.1 Murder Assault UrbanPop Rape

1 1 12.165 255.2500 68.40000 29.16500

2 2 4.870 114.4333 63.63333 15.94333