Technological Institute of Tijuana

Academic Subdirectorate

Systems and Computing Department

SEMESTER: August - December 2021

CAREER: Computer Systems Engineer

MATTER: Data Mining

JOB NAME: Unit 4 - Exam

STUDENT NAME AND CONTROL NUMBER:

Castro Cebreros Alejandro - 16211341

Márquez Millán Seashell Vanessa - 17212153

TEACHER NAME: Jose Christian Romero Hernandez

DATE OF DELIVERY: Dicember 12, 2021

Developement

The firs part is decide where we want work and we put the direction

```
getwd()
setwd("C:/Users/vanem/OneDrive/Documentos/9 SEMESTRE/Mineria/Repo
mineria/DataMining/MachineLearning/LogisticRegression")
getwd()
```

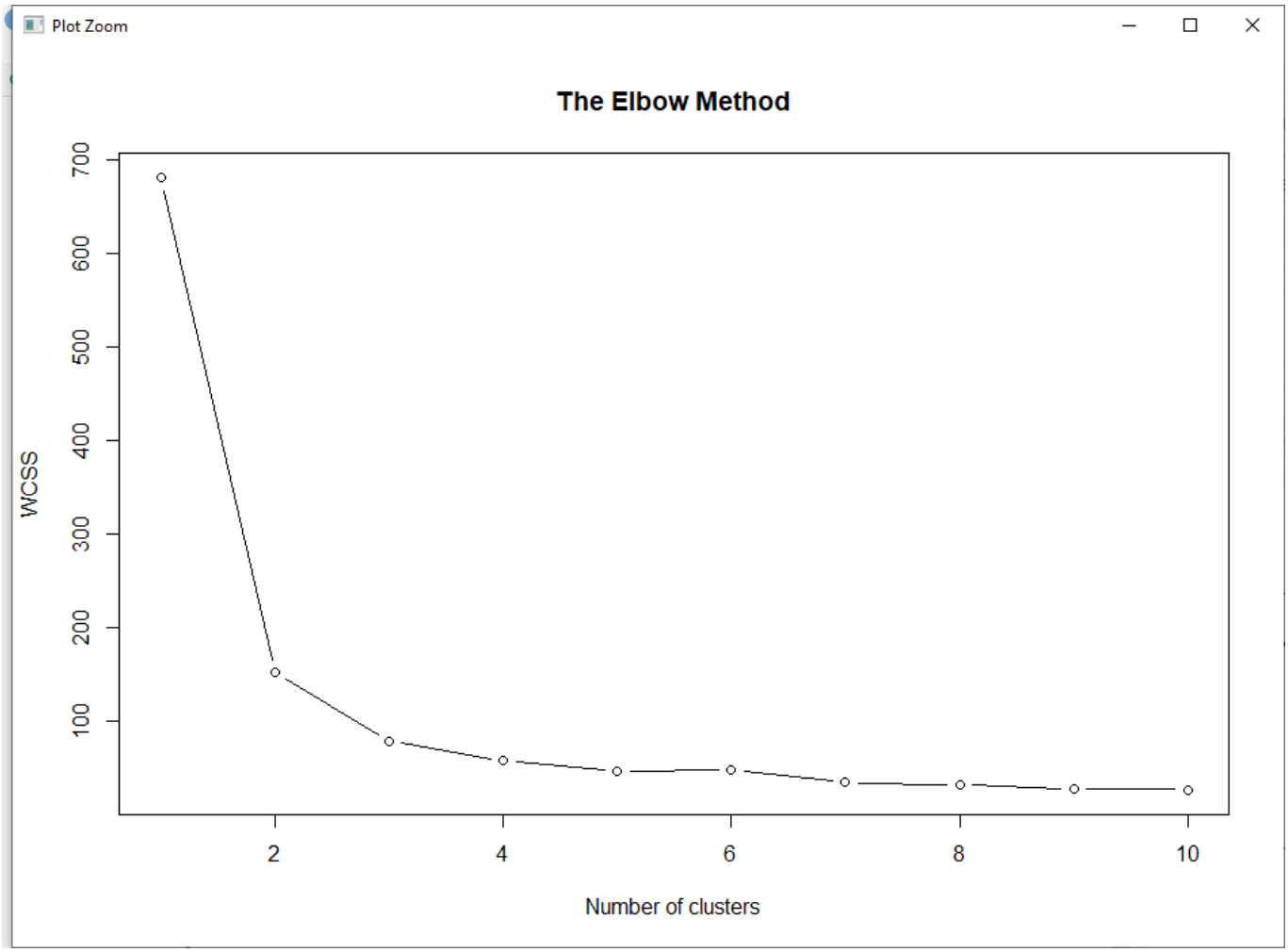
Now we import the data set and select the columns

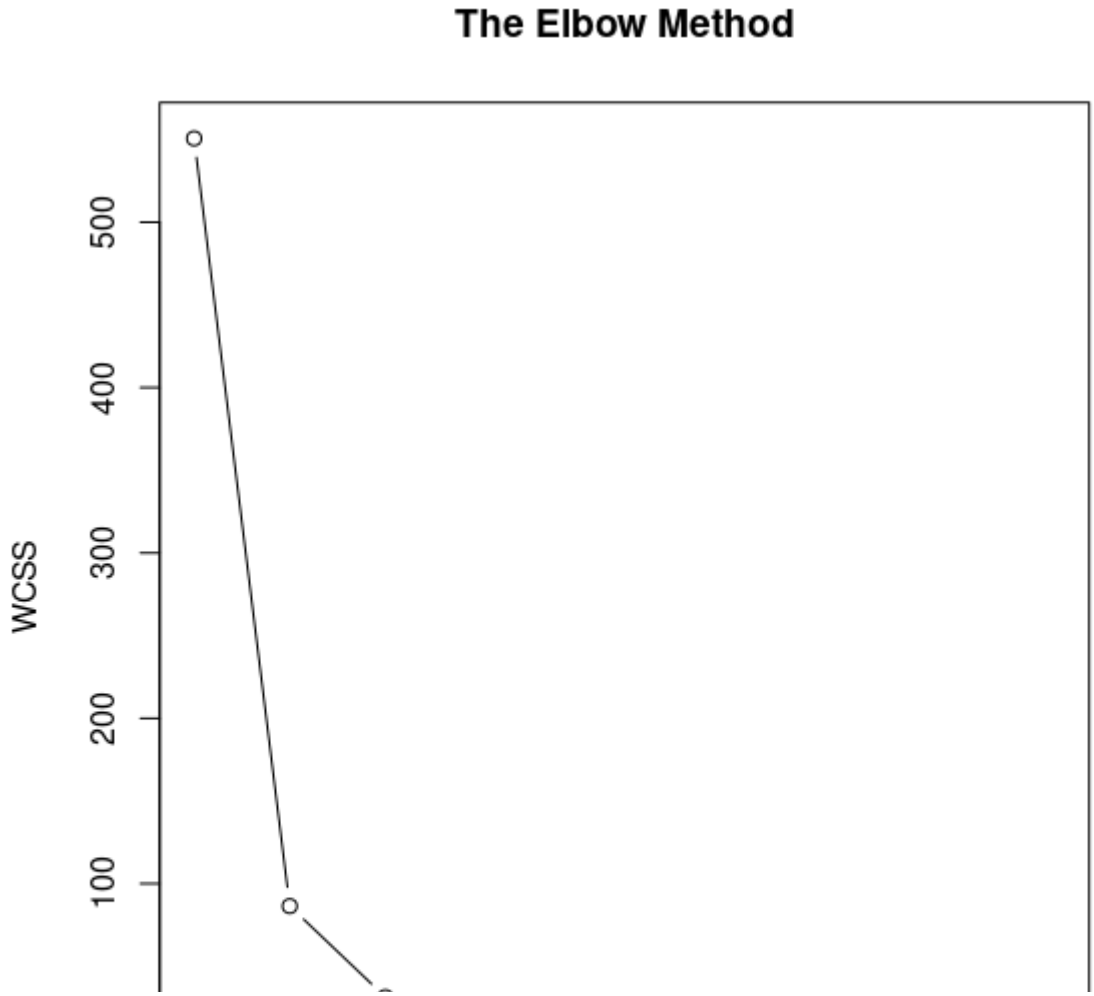
```
dataset <- read.csv(file.choose())
dat1 = dataset[1:4]
dat2 = dataset[3:4]
```

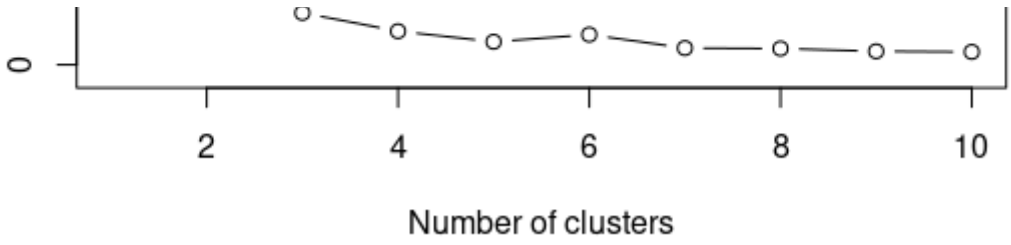
Well here we used the the elbow method to find the optimal number of cluster, but these method is on the funcion because is more easy to visualizing the graphic, and after we call the function and send like param the data set and the resul is here.

```
TEM <- function(dataset){
        set.seed(6)
        wcss = vector()
        for (i in 1:10) wcss[i] = sum(kmeans(dataset, i)$withinss)
        plot(1:10,
             wcss,
             type = 'b',
             main = paste('The Elbow Method'),
             xlab = 'Number of clusters',
             ylab = 'WCSS')
}

TEM(dat1)
TEM(dat2)
```


Plot Zoom
The Elbow Method
WCSS
100
200
300
400
500
600
700
2
4
6
8
10
Number of clusters


The Elbow Method
WCSS
100
200
300
400
500



Like we can see the breaking point is in the point number 3 so is these the number we used more later.

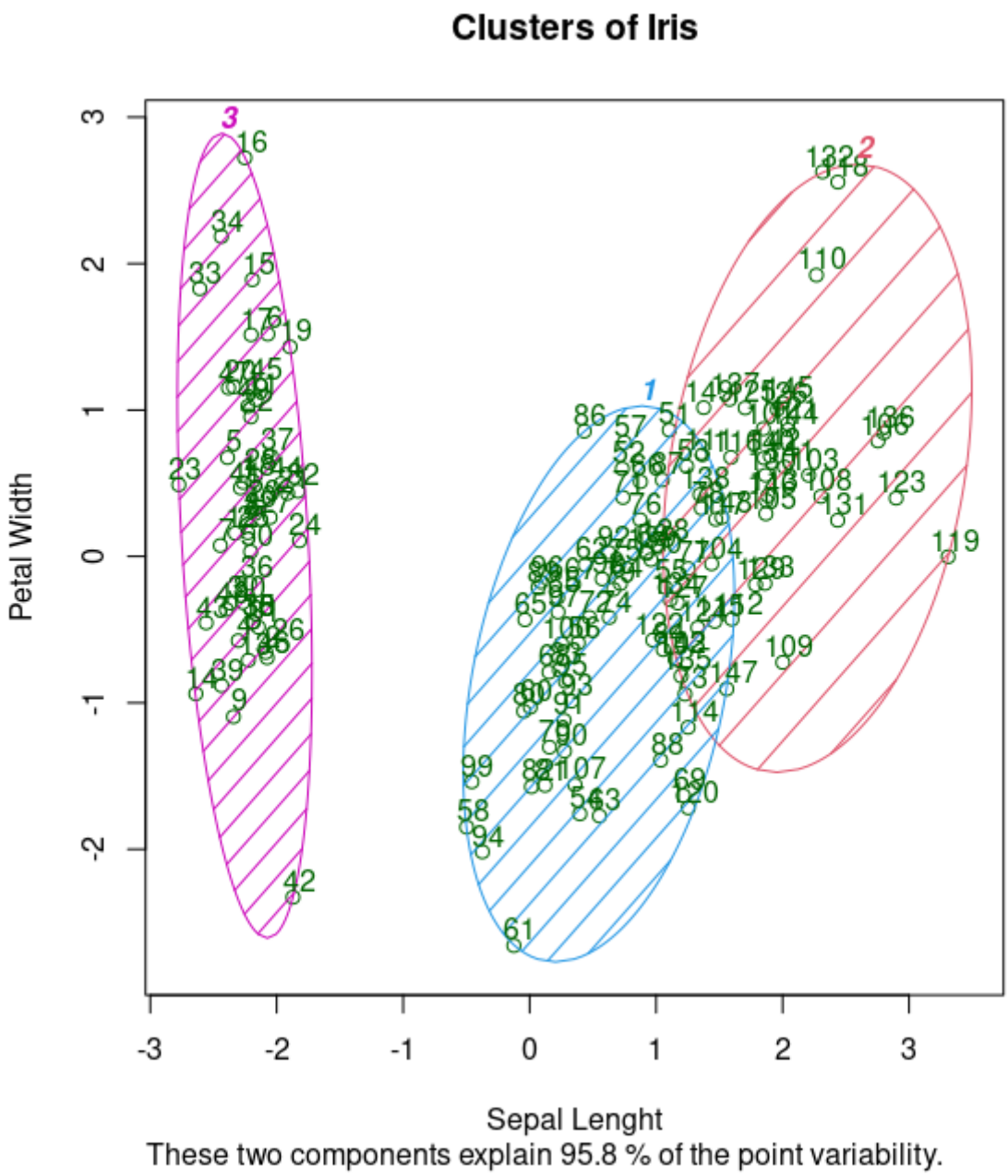
Here we used the model and fitting the dataset

```
set.seed(29)
kmeans = kmeans(x = dataset, centers = 3)
y_kmeans = kmeans$cluster
```

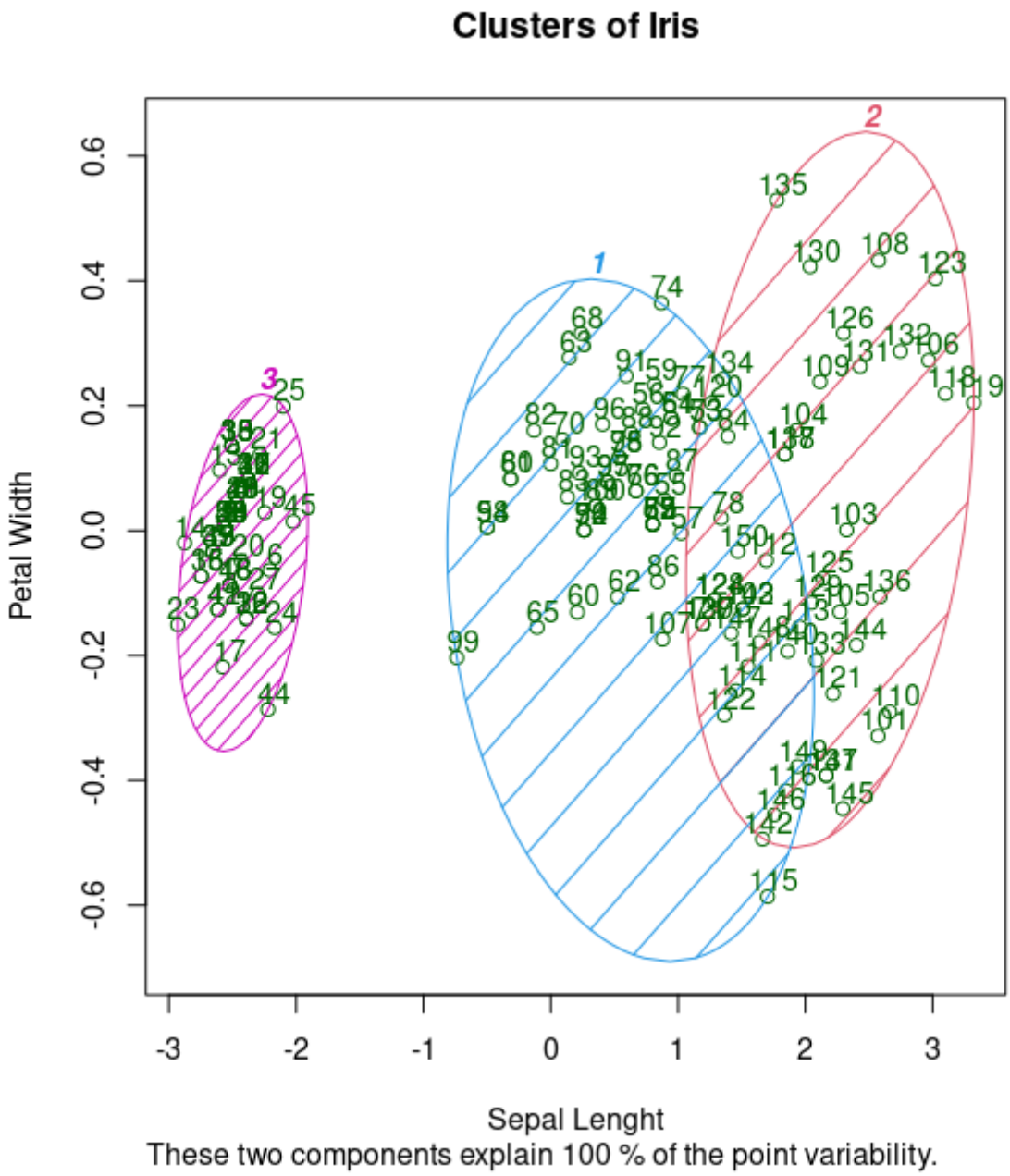
Here we visualize the cluster, but first we need the cluster library in order to execute the code well

```
library(cluster)

#Sepal lenght and Petal width
clusplot(dat1,
         y_kmeans,
         lines = 0,
         shade = TRUE,
         color = TRUE,
         labels = 2,
         plotchar = FALSE,
         span = TRUE,
         main = paste('Clusters of Iris'),
         xlab = 'Sepal lenght',
         ylab = 'Petal width')
```




```
#Petal lenght and Petal width
clusplot(dat2,
         y_kmeans,
         lines = 0,
         shade = TRUE,
         color = TRUE,
         labels = 2,
         plotchar = FALSE,
         span = TRUE,
         main = paste('Clusters of Iris'),
         xlab = 'Petal lenght',
         ylab = 'Petal width')
```


Clusters of Iris
Petal Width
Sepal Lenght
These two components explain 100 % of the point variability.
1
2
3