Untitled

Ing. Erick Chacón, MAIT | MAF Candidate

2023-07-18

```
# CARGAR LIBRERIAS ####
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(cluster)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(NbClust)
library(tidyr)
library(psych)
```

```
##
## Attaching package: 'psych'
##
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
# 2009 ####
# IMPORTAR DATASET ####
df2 <- read.table("/Users/erickfernandochaconflores/Downloads/GTP2.csv",
                  header=TRUE, sep=",", row.names="Departamento")

# DESCRIBRIR DATASET ####
str (df2)
```

```
## 'data.frame':    22 obs. of  2 variables:
##  $ Rob_Hurto : num  281.9 88.7 211.9 77 234.9 ...
##  $ Homicidios: num  86.7 56.8 24.1 19 76.9 ...
```

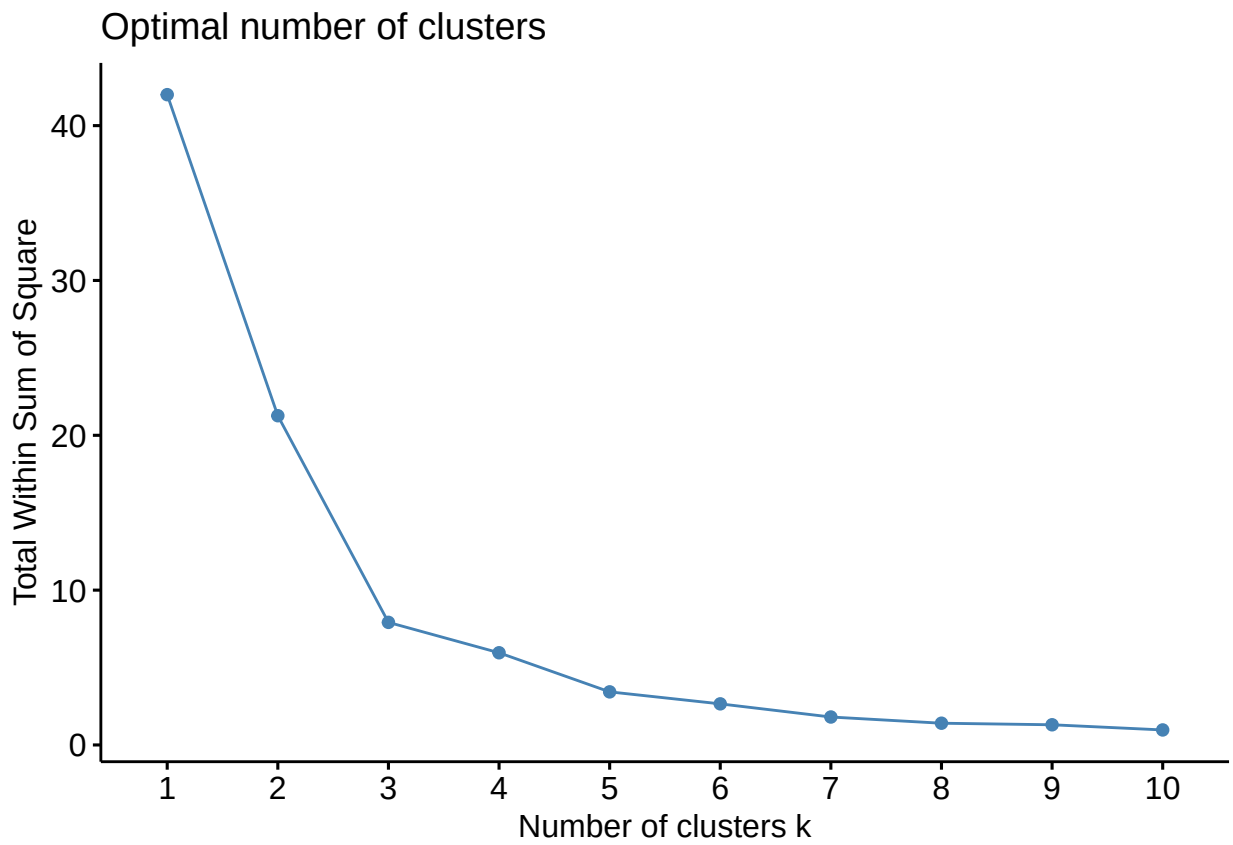
```r
describe(df2)
```

```
##            vars  n  mean    sd median trimmed   mad   min    max  range skew
## Rob_Hurto     1 22 79.04 71.68  64.65   65.76 38.30 18.75 281.94 263.19 1.63
## Homicidios    2 22 43.27 27.92  34.17   42.32 32.71  4.02  90.93  86.91 0.23
##            kurtosis    se
## Rob_Hurto      1.59 15.28
## Homicidios    -1.56  5.95
```

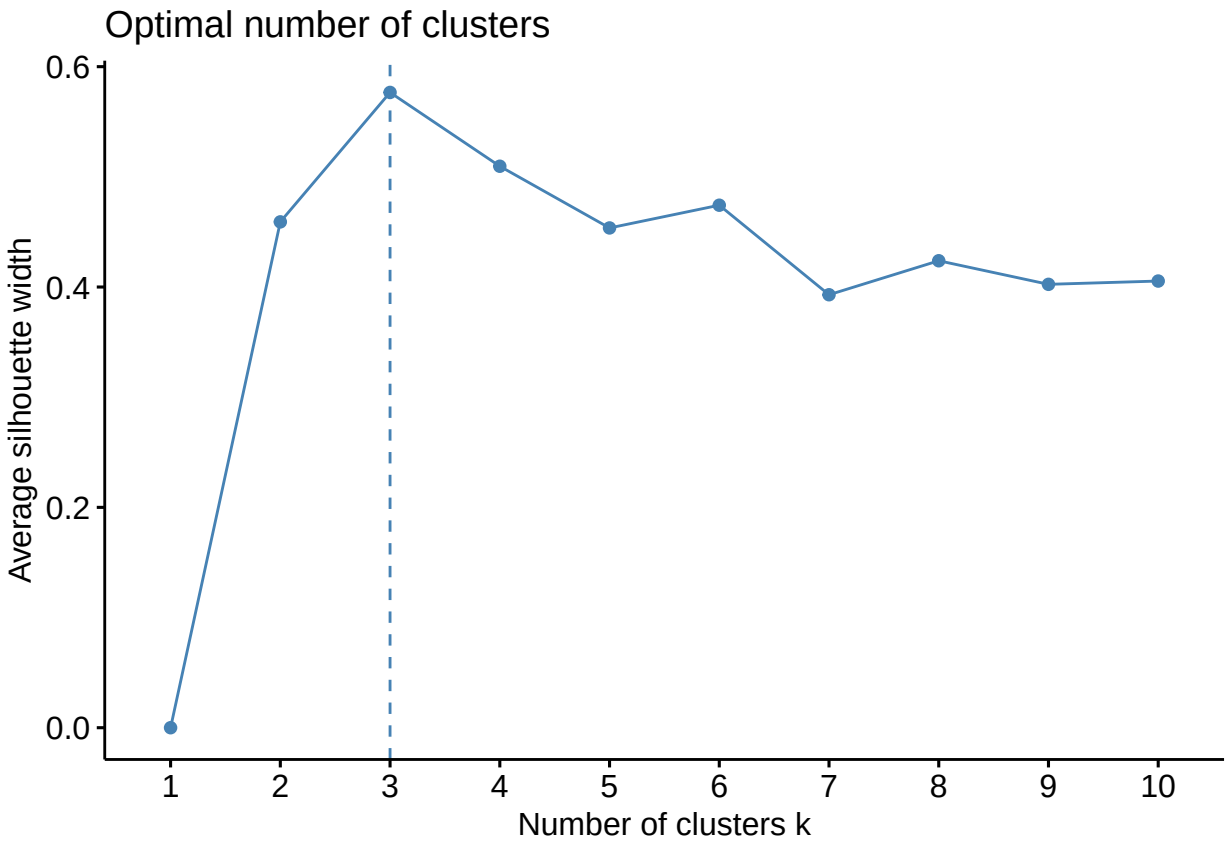
```r
# NORMALIZAR VARIABLES ####
df2 <- scale(df2) # "Scale" función para normalizar
head(df2)
```

```
##                 Rob_Hurto Homicidios
## Guatemala      2.83054070  1.5556254
## El Progreso    0.13525478  0.4835076
## Sacatepéquez   1.85331934 -0.6880351
## Chimaltenango -0.02803739 -0.8691538
## Escuintla      2.17396865  1.2064038
## Santa Rosa    -0.28162379  1.0077499
```

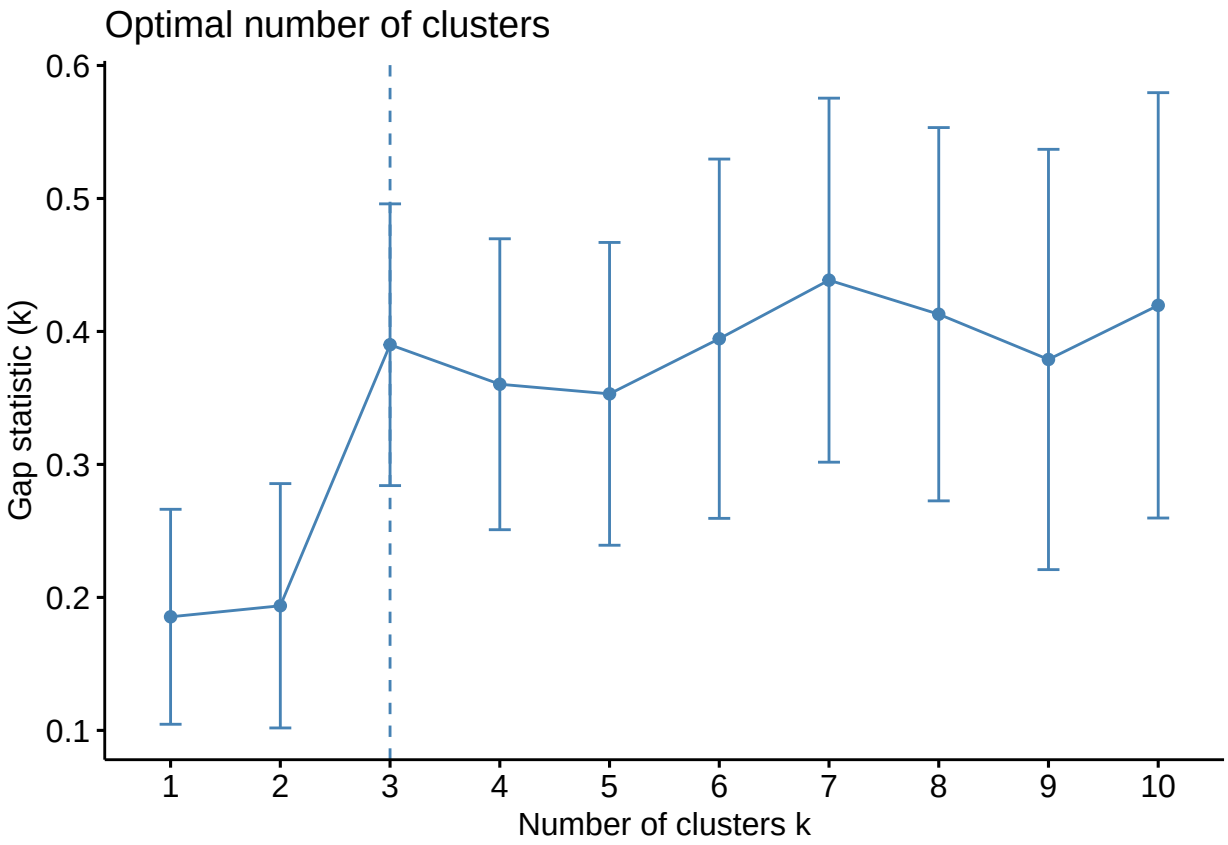
```r
# ESTIMAR EL NÚMERO DE CLUSTERS ####
# Elbow, silhouette o gap_stat  method
fviz_nbclust(df2, kmeans, method = "wss") # Hay que encontar la rodilla
```



```
fviz_nbclust(df2, kmeans, method = "silhouette") # Sugiere el número de clusters
```




```
fviz_nbclust(df2, kmeans, method = "gap_stat") # Sugiere el número de clusters
```



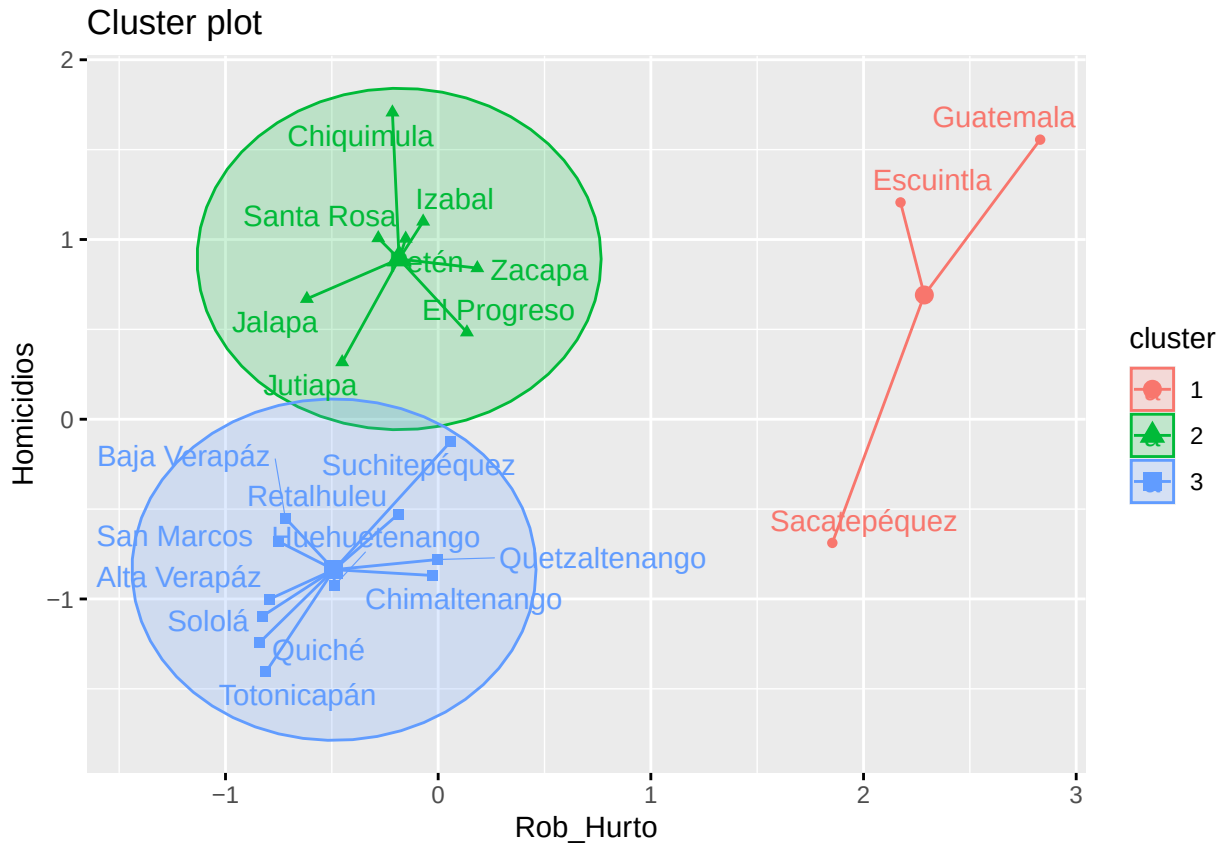

```
# CALCULAR LOS CLUSTER SUGERIDOS POR LOS MÉTODOS ####
# Son dos los clusters sugeridos
k3_2009 <- kmeans(df2, centers = 3, nstart = 25)
k3_2009 # me dice en que cluster ha quedado cada uno de los estados
```

```
## K-means clustering with 3 clusters of sizes 3, 8, 11
##
## Cluster means:
##    Rob_Hurto Homicidios
## 1  2.2859429  0.6913314
## 2 -0.1836762  0.8914722
## 3 -0.4898563 -0.8368884
##
## Clustering vector:
##      Guatemala    El Progreso   Sacatepéquez  Chimaltenango      Escuintla
##              1              2              1              3              1
##     Santa Rosa         Sololá   Totonicapán Quetzaltenango  Suchitepéquez
##              2              3              3              3              3
##     Retalhuleu     San Marcos  Huehuetenango         Quiché   Baja Verapáz
##              3              3              3              3              3
##   Alta Verapáz          Petén         Izabal         Zacapa     Chiquimula
##              3              2              2              2              2
##         Jalapa        Jutiapa
##              2              2
##
```

```
## Within cluster sum of squares by cluster:
## [1] 3.411244 1.803463 2.703439
##  (between_SS / total_SS =  81.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
# PLOTEANDO CLUSTERS ####
fviz_cluster(k3_2009, data = df2, ellipse.type = "euclid",repel =TRUE,
             star.plot = TRUE) #ellipse.type= "t", "norm", "euclid"
```

```
## Too few points to calculate an ellipse
```




```
# 2020
# IMPORTAR DATASET ####
df <- read.table("/Users/erickfernandochaconflores/Downloads/GTP.csv",
                 header=TRUE, sep=",", row.names="Departamento")

# DESCRIBRIR DATASET ####
str (df)
```

```
## 'data.frame':    22 obs. of  2 variables:
##  $ Robo_Hurto: num  110.1 38.3 36.9 28.9 120.6 ...
##  $ Homicidios: num  23.52 19.16 5.39 7.91 38.78 ...
```

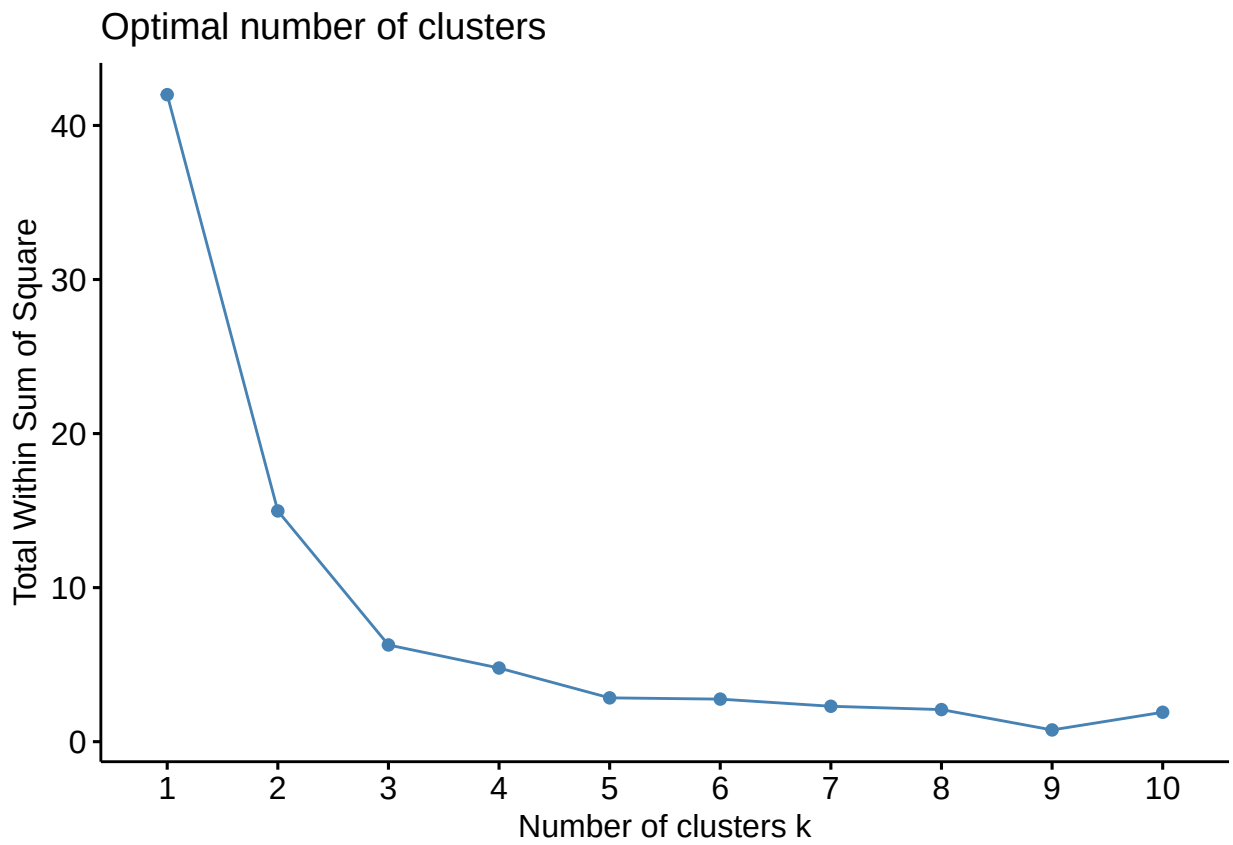
```r
describe(df)
```

```
##            vars  n  mean    sd median trimmed   mad  min    max  range skew
## Robo_Hurto    1 22 32.19 33.54  24.31   26.08 20.89 2.05 120.64 118.58 1.56
## Homicidios    2 22 16.36 13.54  12.64   15.20 13.40 1.60  44.27  42.67 0.62
##            kurtosis   se
## Robo_Hurto     1.22 7.15
## Homicidios    -1.02 2.89
```

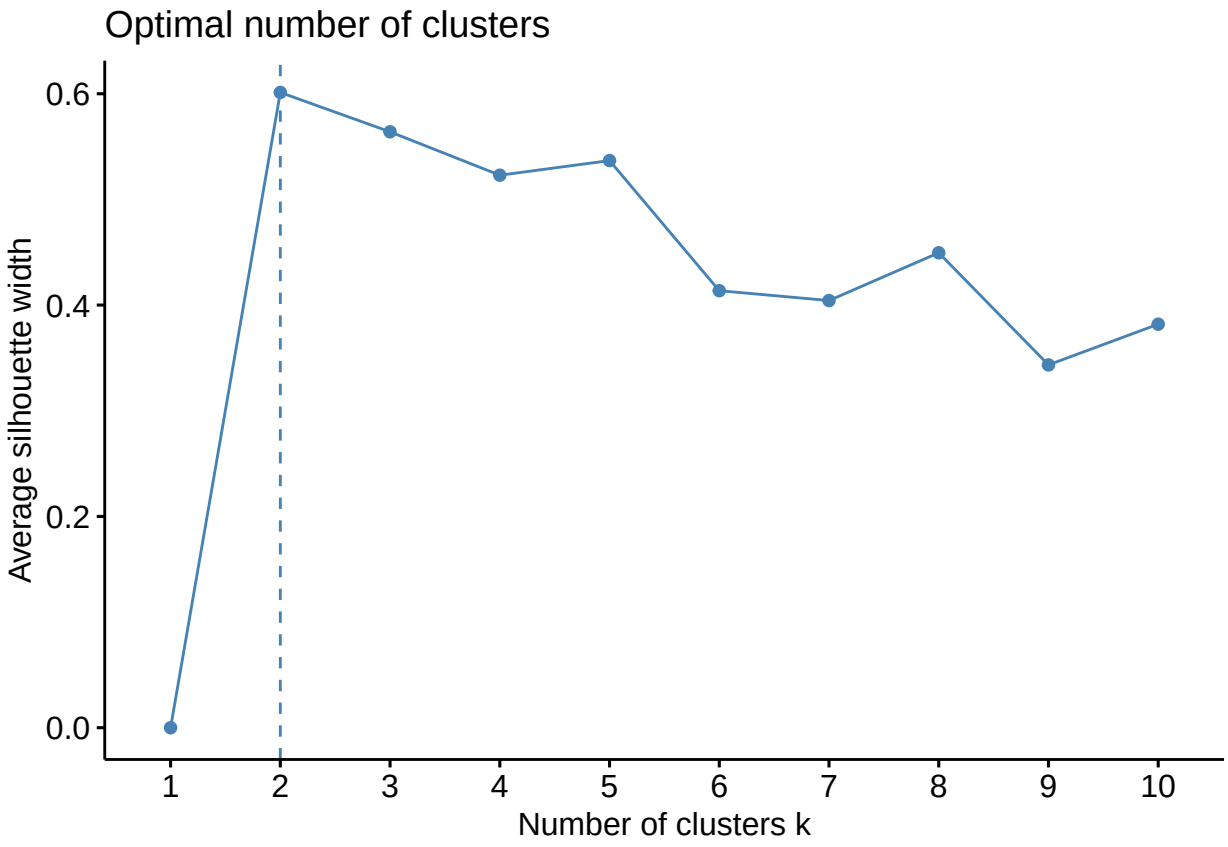
```r
# NORMALIZAR VARIABLES ####
df <- scale(df) # "Scale" función para normalizar
head(df)
```

```
##                Robo_Hurto Homicidios
## Guatemala       2.32182601  0.5295274
## El Progreso     0.18288375  0.2075201
## Sacatepéquez    0.14124312 -0.8104430
## Chimaltenango -0.09792875 -0.6240250
## Escuintla       2.63665826  1.6567957
## Santa Rosa     -0.52715060  0.4165858
```

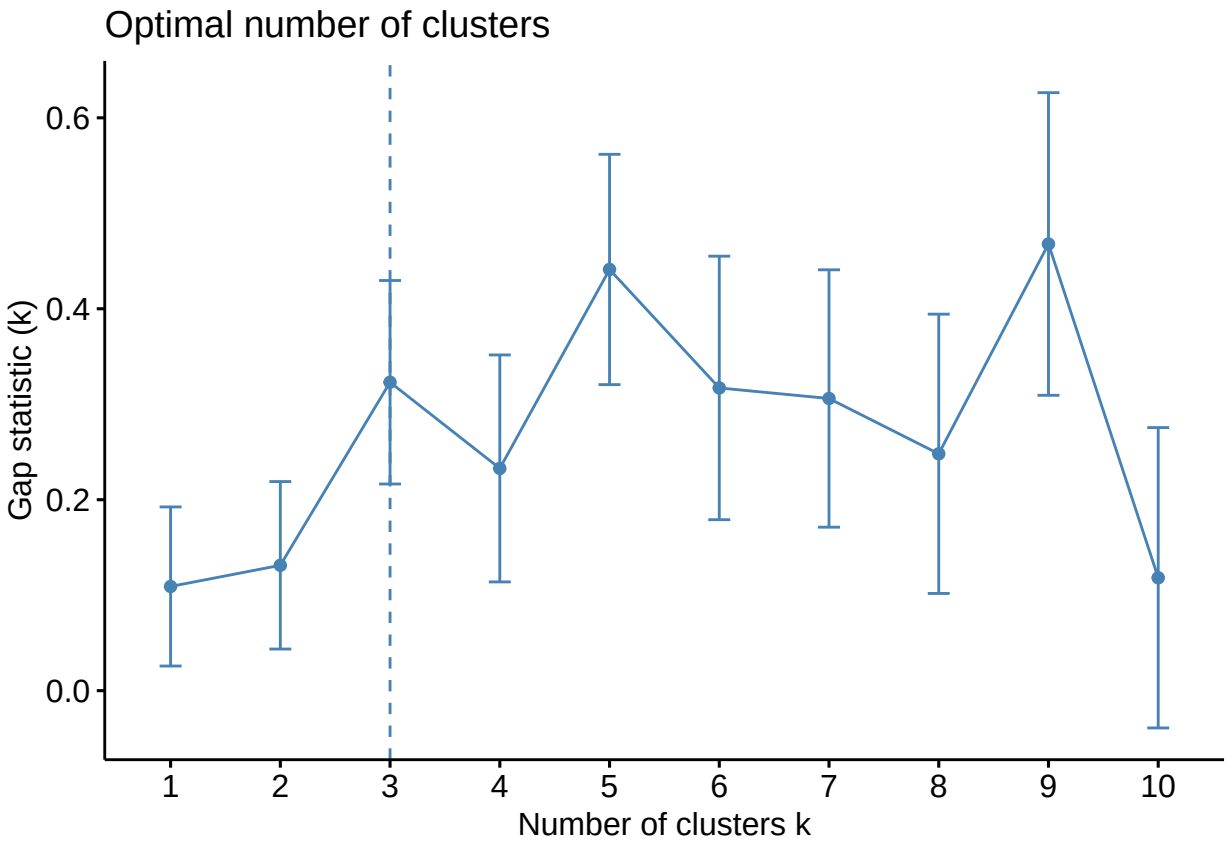
```r
# ESTIMAR EL NÚMERO DE CLUSTERS ####
# Elbow, silhouette o gap_stat  method
fviz_nbclust(df, kmeans, method = "wss") # Hay que encontar la rodilla
```



```r
fviz_nbclust(df, kmeans, method = "silhouette") # Sugiere el número de clusters
```




```r
fviz_nbclust(df, kmeans, method = "gap_stat") # Sugiere el número de clusters
```




```
# CALCULAR LOS CLUSTER SUGERIDOS POR LOS MÉTODOS ####
# Son dos los clusters sugeridos
k3_2020 <- kmeans(df, centers = 3, nstart = 25)
k3_2020 # me dice en que cluster ha quedado cada uno de los estados
```

```
## K-means clustering with 3 clusters of sizes 3, 7, 12
##
## Cluster means:
##   Robo_Hurto Homicidios
## 1  2.2849234  1.4161958
## 2 -0.1333748  0.7284216
## 3 -0.4934289 -0.7789615
##
## Clustering vector:
##      Guatemala    El Progreso   Sacatepéquez  Chimaltenango      Escuintla
##              1              2              3              3              1
##     Santa Rosa         Sololá   Totonicapán Quetzaltenango  Suchitepéquez
##              2              3              3              3              3
##     Retalhuleu     San Marcos  Huehuetenango         Quiché   Baja Verapáz
##              3              3              3              3              3
##   Alta Verapáz          Petén         Izabal         Zacapa     Chiquimula
##              3              2              1              2              2
##         Jalapa        Jutiapa
##              2              2
##
```

```
## Within cluster sum of squares by cluster:
## [1] 1.537592 2.693185 2.048021
##  (between_SS / total_SS =  85.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
# PLOTEANDO CLUSTERS ####
fviz_cluster(k3_2020, data = df, ellipse.type = "euclid",repel = TRUE,
             star.plot = TRUE) #ellipse.type= "t", "norm", "euclid"
```

```
## Too few points to calculate an ellipse
```