

# Laboratorio 8, Tópicos en análisis datos 1

Joshua Cervantes Artavia - Moisés Monge Cordonero

2023-11-02

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
```

```
tryCatch(  
  {  
    # Directorio donde se ubica el qmd  
    directory <- dirname(rstudioapi::getSourceEditorContext())$path  
    setwd(directory) # Establecer el directorio del archivo como la raiz  
  },  
  error = function(e) {  
    message("")  
    print("")  
  }  
)
```

```
[1] ""
```

```
source("cod/set_up.R")
```

```
-- 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    3.4.4      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
Loading required package: tictoc
```

# 1 Aplicaciones del método de k-means

## 1.1 Notas escolares

```
# We read the excel with the data
df_notas_escolares <- read.xlsx("./data/Ejercicios-Cap3.xlsx", "9.NotasFrancesas")

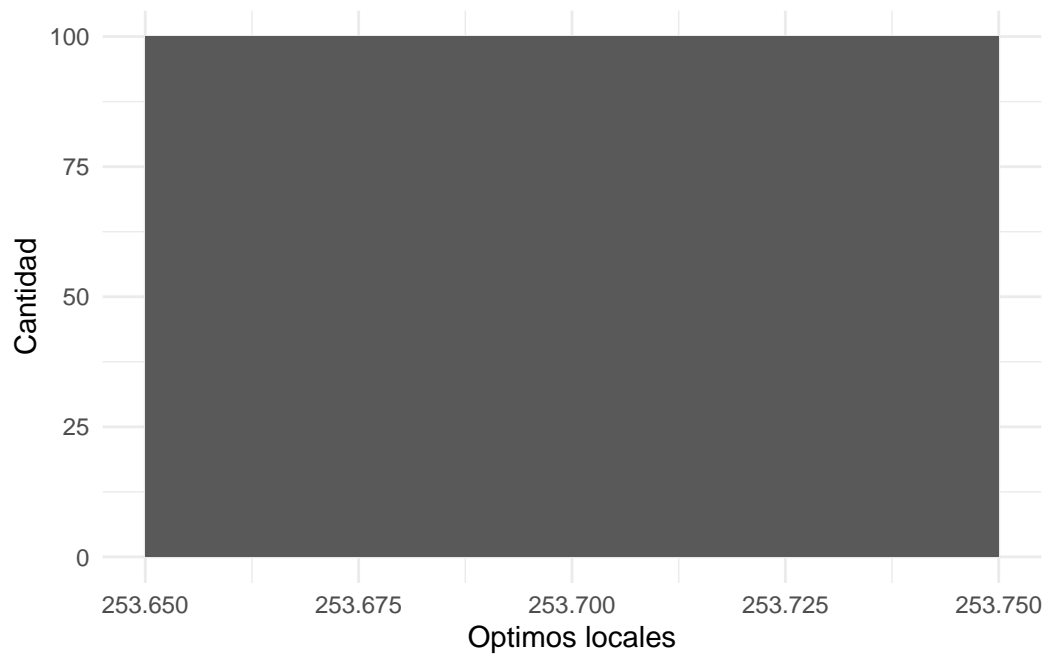
# We make the name of rows the name of the studentes
rownames(df_notas_escolares) <- df_notas_escolares[, 1]

# We delete the first column
df_notas_escolares <- df_notas_escolares[, -1]

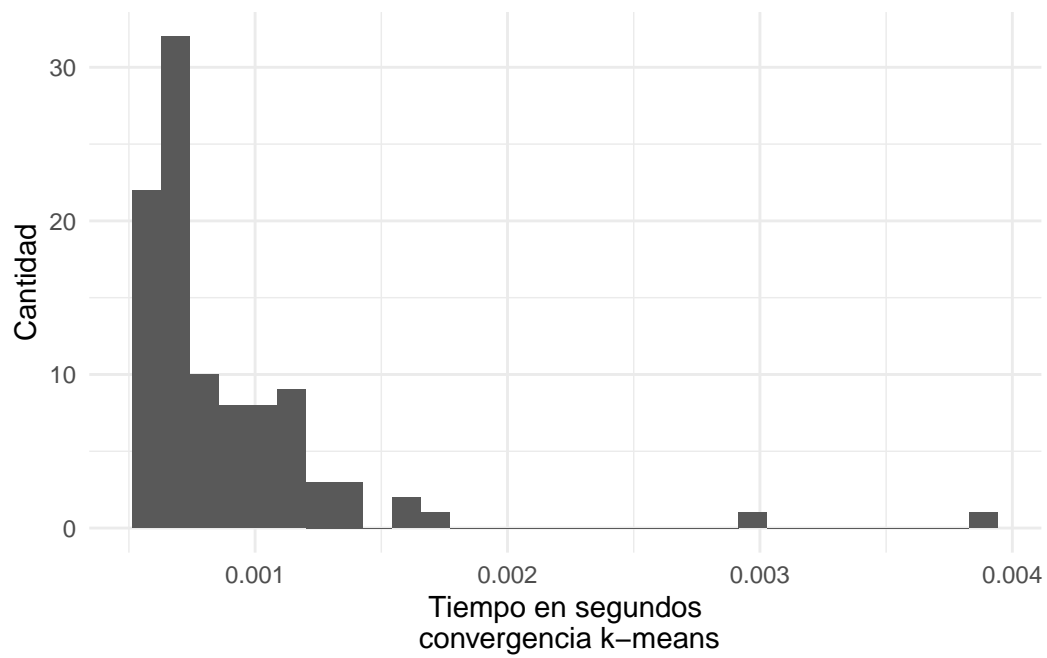
# We estimate some of the point asked
notas_k_2 <- fn_punto_1(df = df_notas_escolares, k = 2)
notas_k_3 <- fn_punto_1(df = df_notas_escolares, k = 3)
notas_k_4 <- fn_punto_1(df = df_notas_escolares, k = 4)

# We print the summary asked for the point
notas_k_2$resumen
```

\$plot\_optimos



`$plot_tiempo`



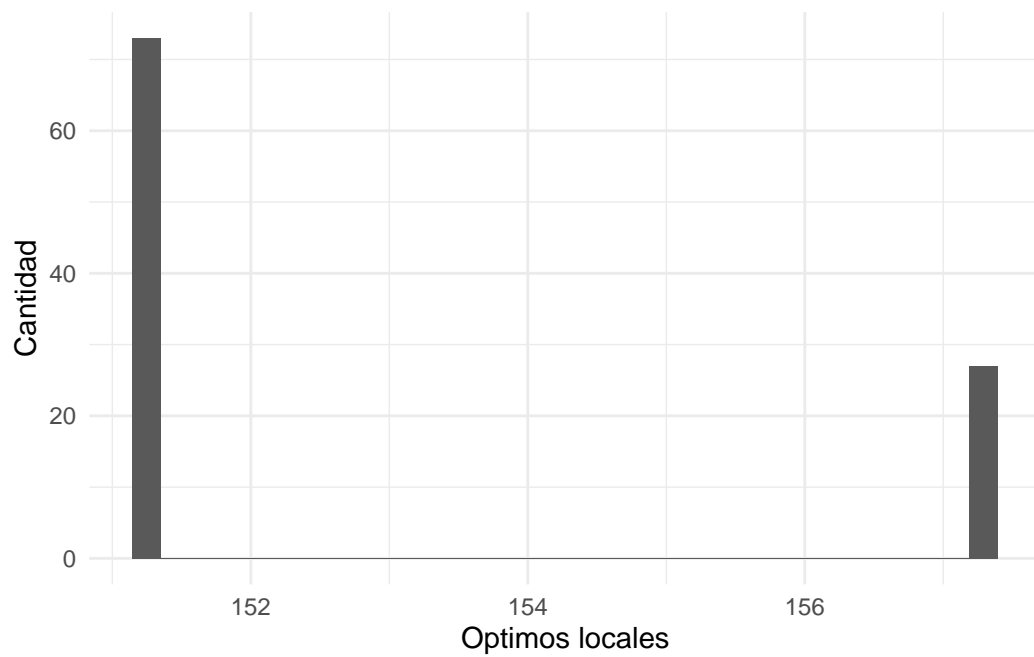
```
$optimo_promedio  
[1] 253.7125
```

```
$mejor_optimo  
[1] 253.7125
```

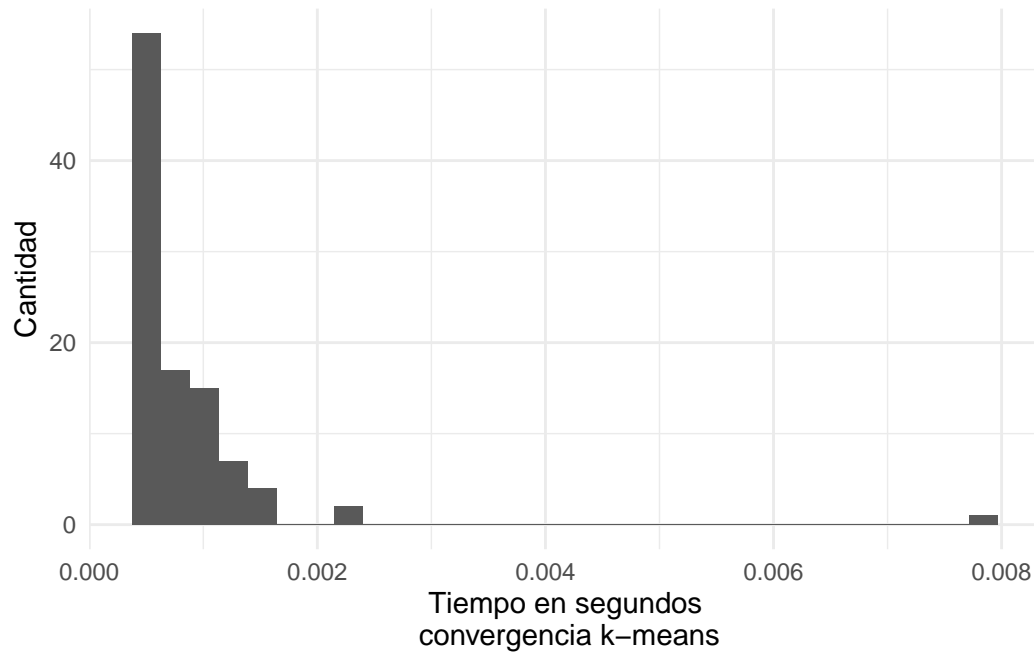
```
$atraccion_mejor_optimo  
[1] 100
```

```
notas_k_3$resumen
```

```
$plot_optimos
```



```
$plot_tiempo
```



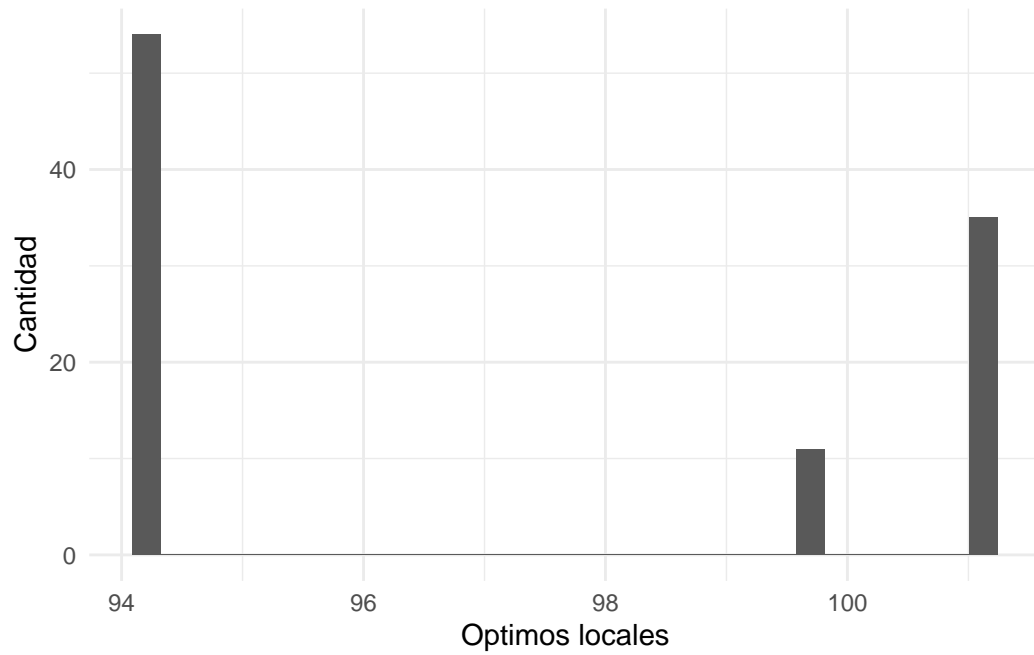
```
$optimo_promedio  
[1] 152.9646
```

```
$mejor_optimo  
[1] 151.3333
```

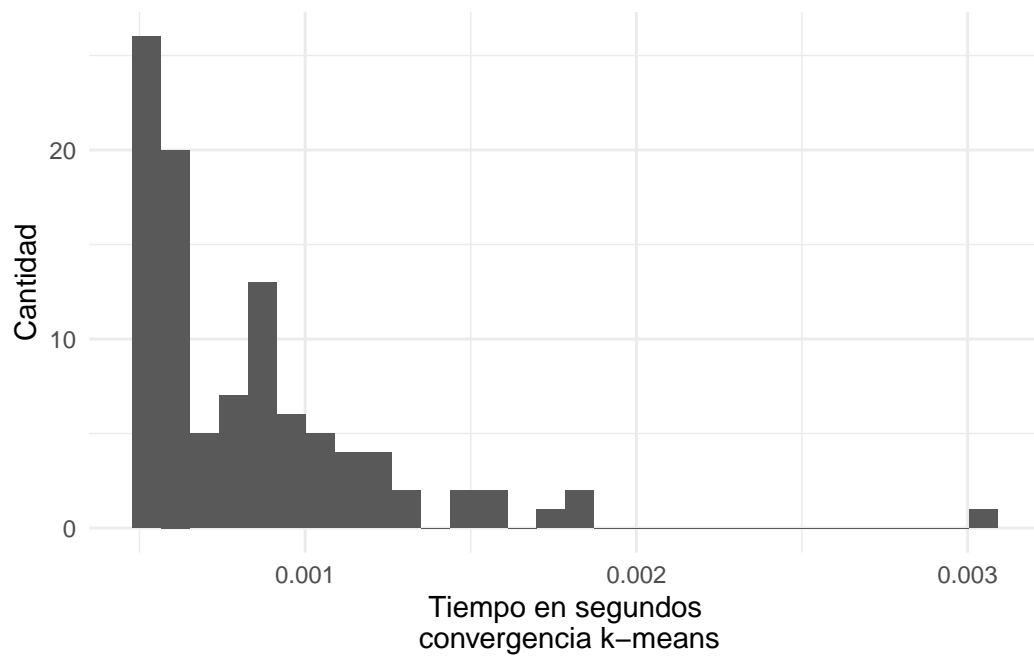
```
$atraccion_mejor_optimo  
[1] 73
```

```
notas_k_4$resumen
```

```
$plot_optimos
```



`$plot_tiempo`



```
$optimo_promedio  
[1] 97.225
```

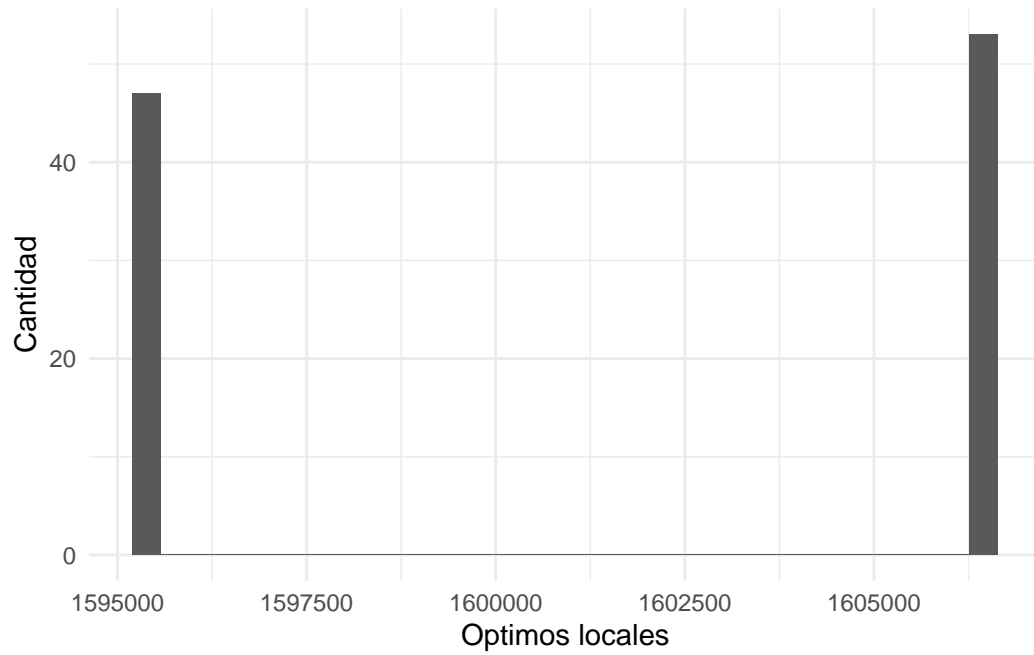
```
$mejor_optimo  
[1] 94.20833
```

```
$atraccion_mejor_optimo  
[1] 54
```

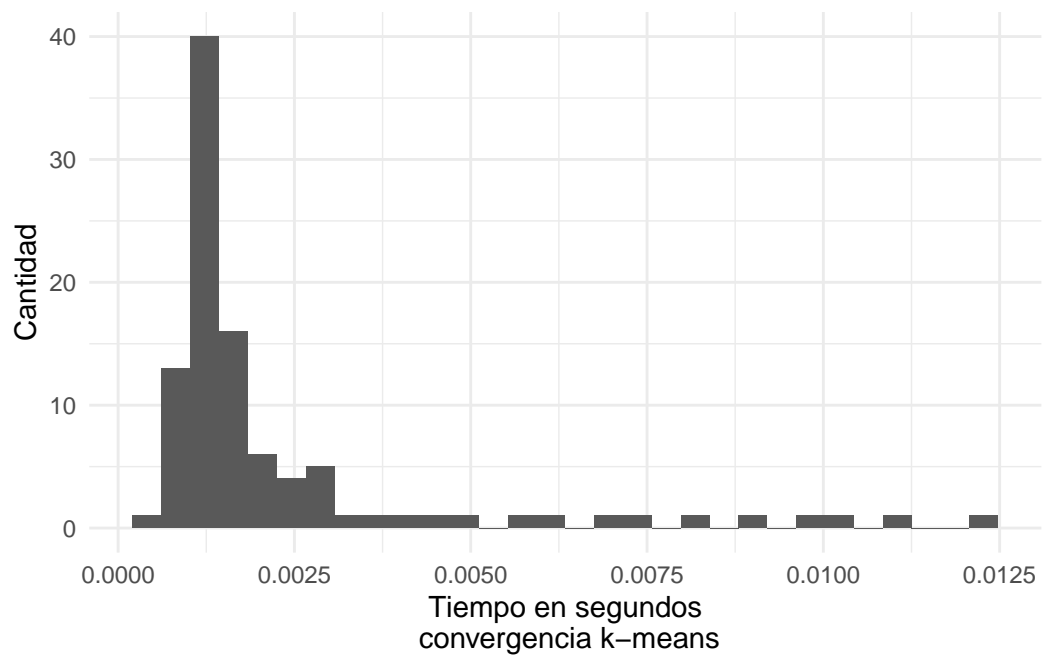
## 1.2 Notas Amiard

```
# We read the excel with the data  
df_amiard <- read.xlsx("./data/Ejercicios-Cap3.xlsx", "10.Amiard")  
  
# We make the name of rows the name of the studentes  
rownames(df_amiard) <- df_amiard[, 1]  
  
# We delete the first column  
df_amiard <- df_amiard[, -1]  
  
# We estimate some of the point asked  
notas_k_2 <- fn_punto_1(df = df_amiard, k = 2)  
notas_k_3 <- fn_punto_1(df = df_amiard, k = 3)  
notas_k_4 <- fn_punto_1(df = df_amiard, k = 4)  
  
# We print the summary asked for the point  
notas_k_2$resumen
```

```
$plot_optimos
```



`$plot_tiempo`





```
$optimo_promedio
```

```
[1] 1601332
```

```
$mejor_optimo
```

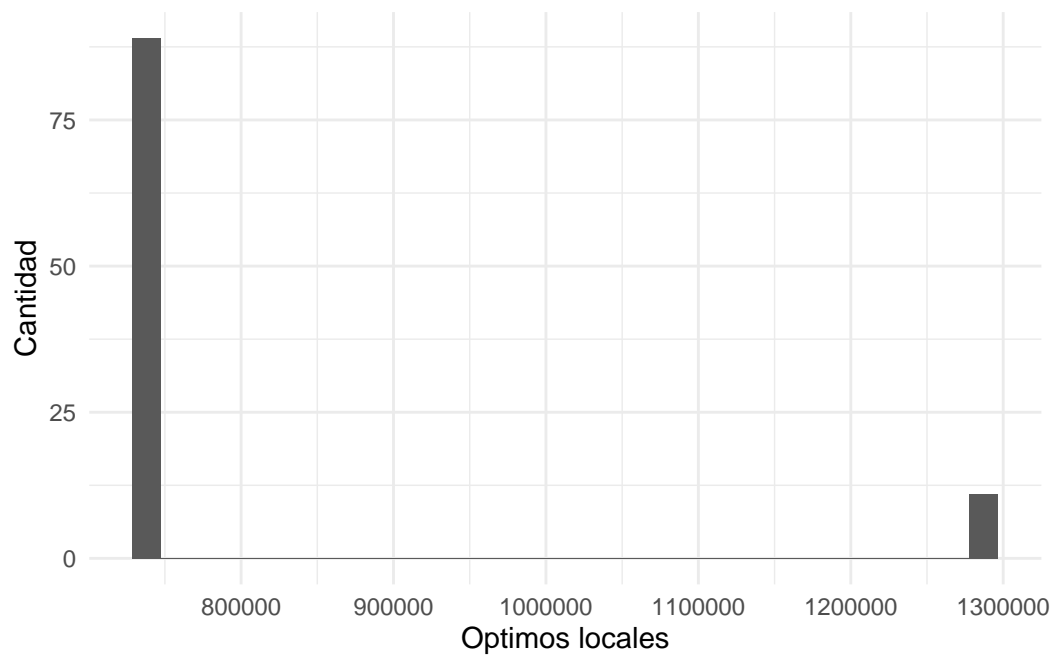
```
[1] 1595470
```

```
$atraccion_mejor_optimo
```

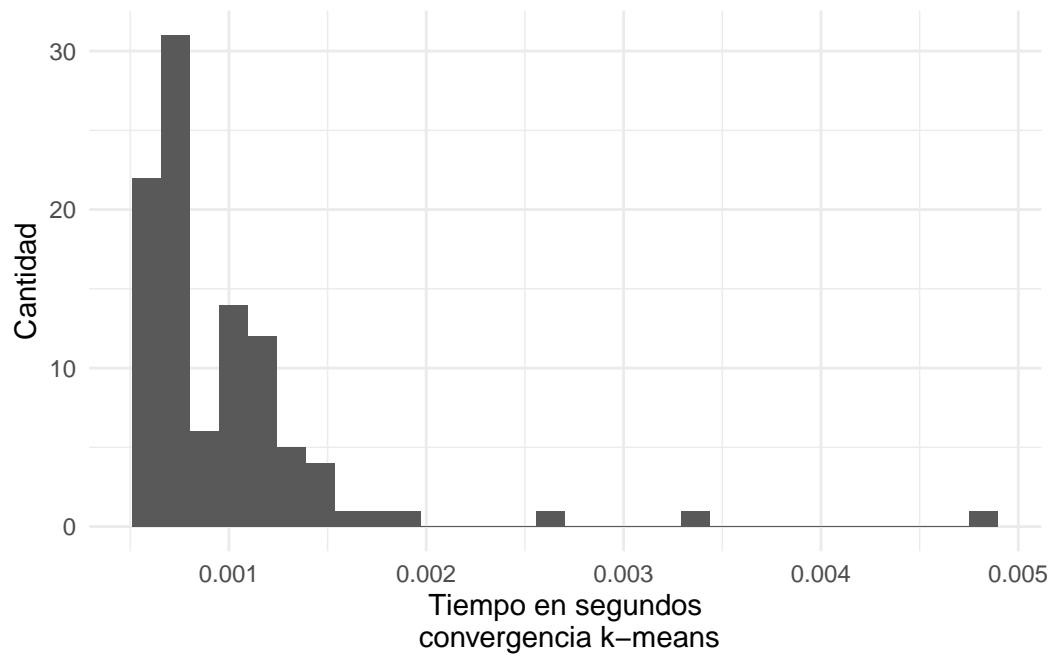
```
[1] 47
```

```
notas_k_3$resumen
```

```
$plot_optimos
```



```
$plot_tiempo
```



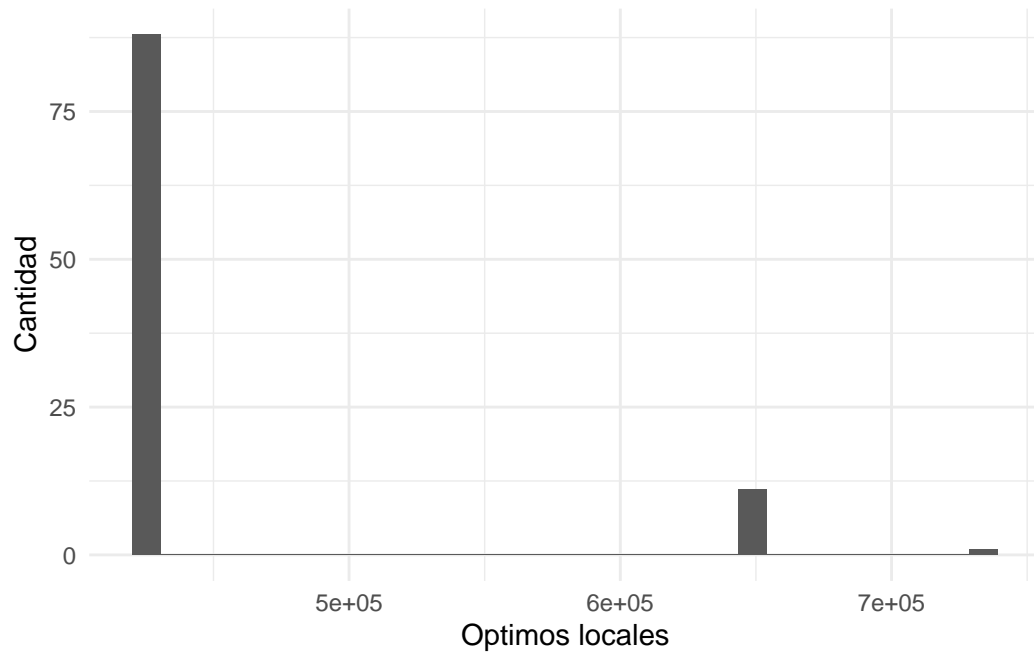
```
$optimo_promedio  
[1] 801292.2
```

```
$mejor_optimo  
[1] 740907.8
```

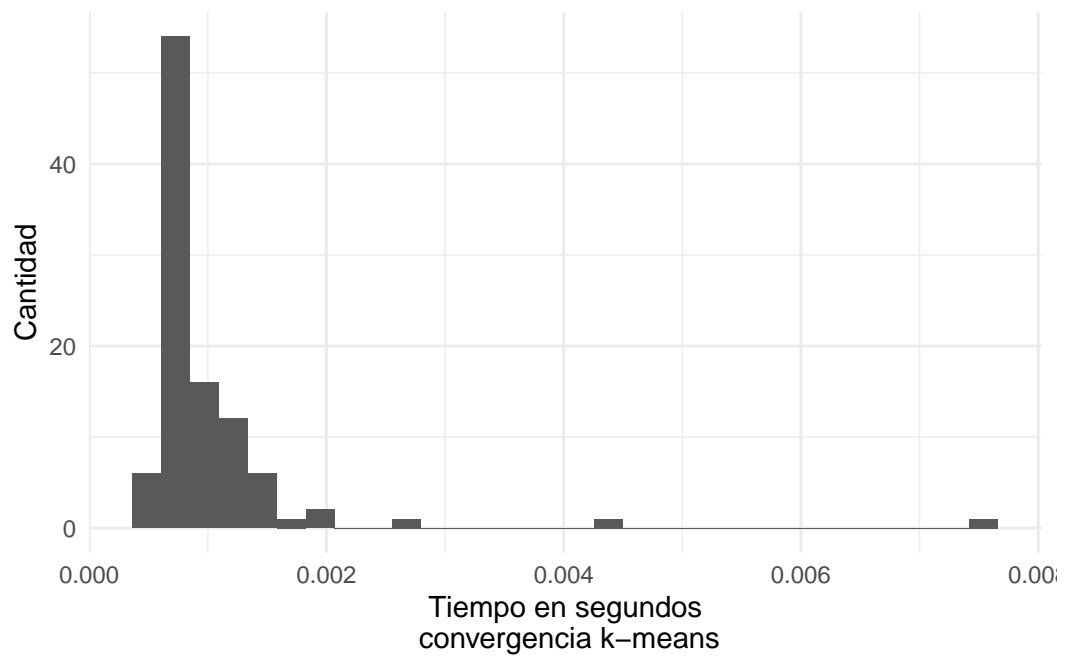
```
$atraccion_mejor_optimo  
[1] 89
```

```
notas_k_4$resumen
```

```
$plot_optimos
```



`$plot_tiempo`



```
$optimo_promedio
```

```
[1] 452036
```

```
$mejor_optimo
```

```
[1] 420471.9
```

```
$atraccion_mejor_optimo
```

```
[1] 22
```

### 1.3 Notas proteínas

```
# We read the excel with the data
df_proteinas <- read.xlsx("./data/Ejercicios-Cap3.xlsx", "12.Proteinas")

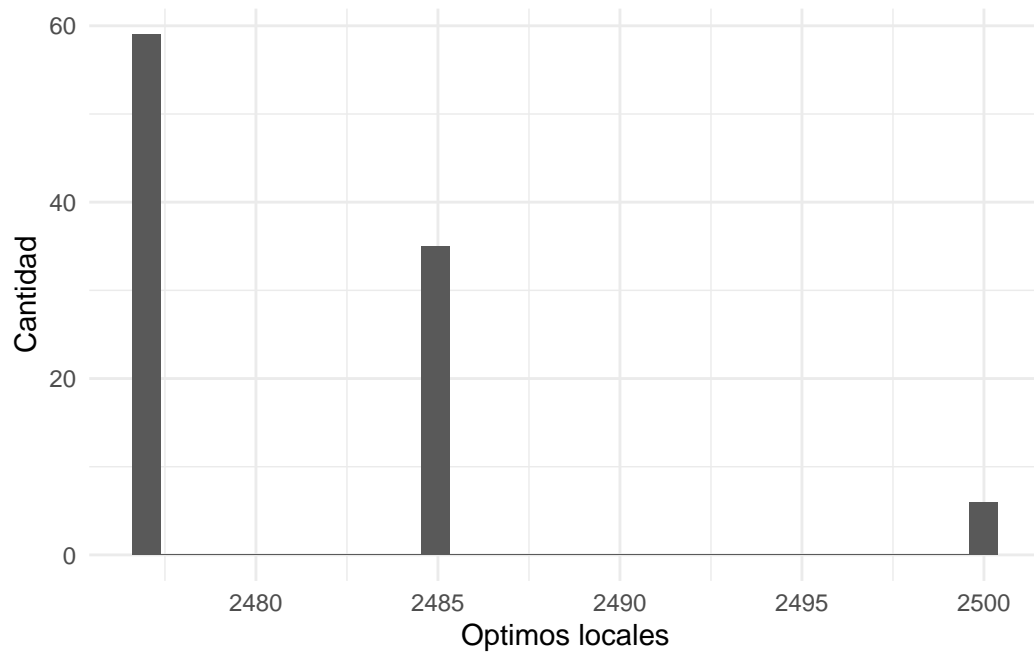
# We make the name of rows the name of the studentes
rownames(df_proteinas) <- df_proteinas[, 1]

# We delete the first column
df_proteinas <- df_proteinas[, -1]

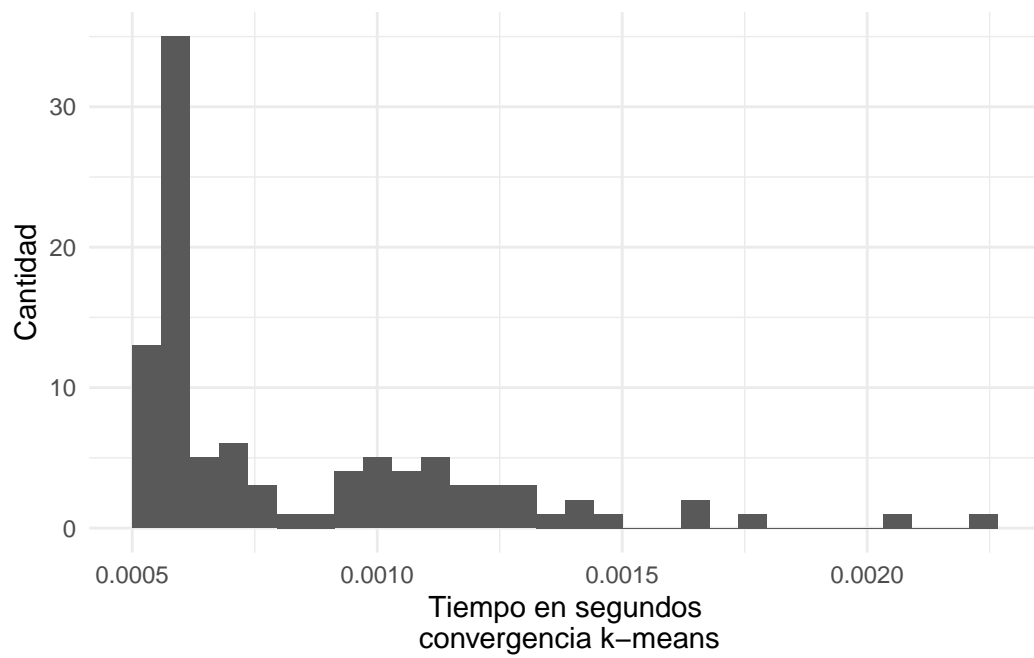
# We estimate some of the point asked
notas_k_2 <- fn_punto_1(df = df_proteinas, k = 2)
notas_k_3 <- fn_punto_1(df = df_proteinas, k = 3)
notas_k_4 <- fn_punto_1(df = df_proteinas, k = 4)

# We print the summary asked for the point
notas_k_2$resumen
```

```
$plot_optimos
```



`$plot_tiempo`



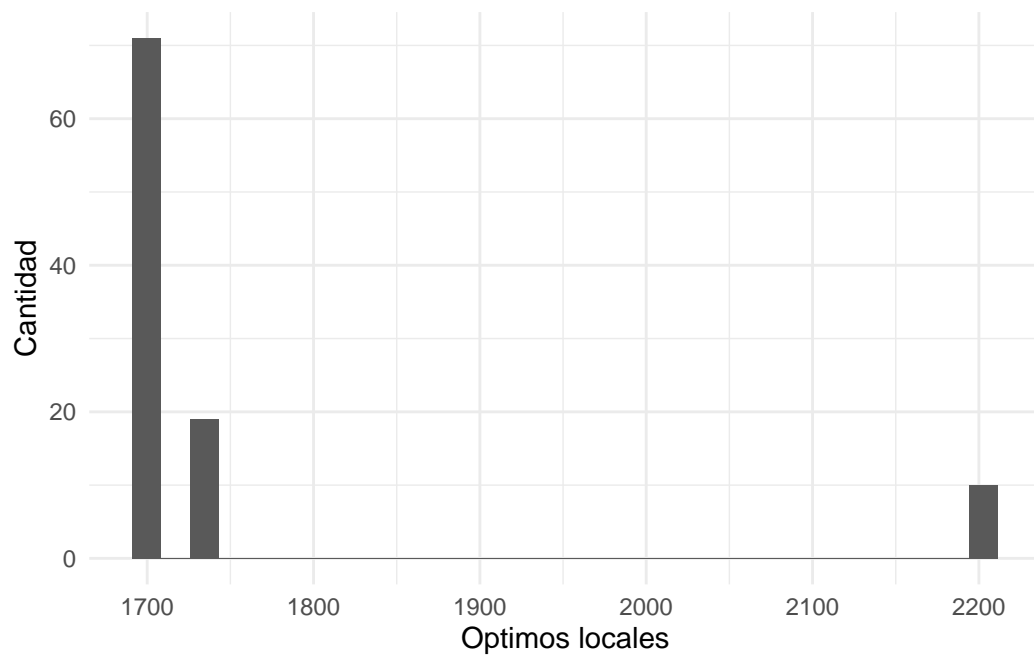
```
$optimo_promedio  
[1] 2480.999
```

```
$mejor_optimo  
[1] 2476.749
```

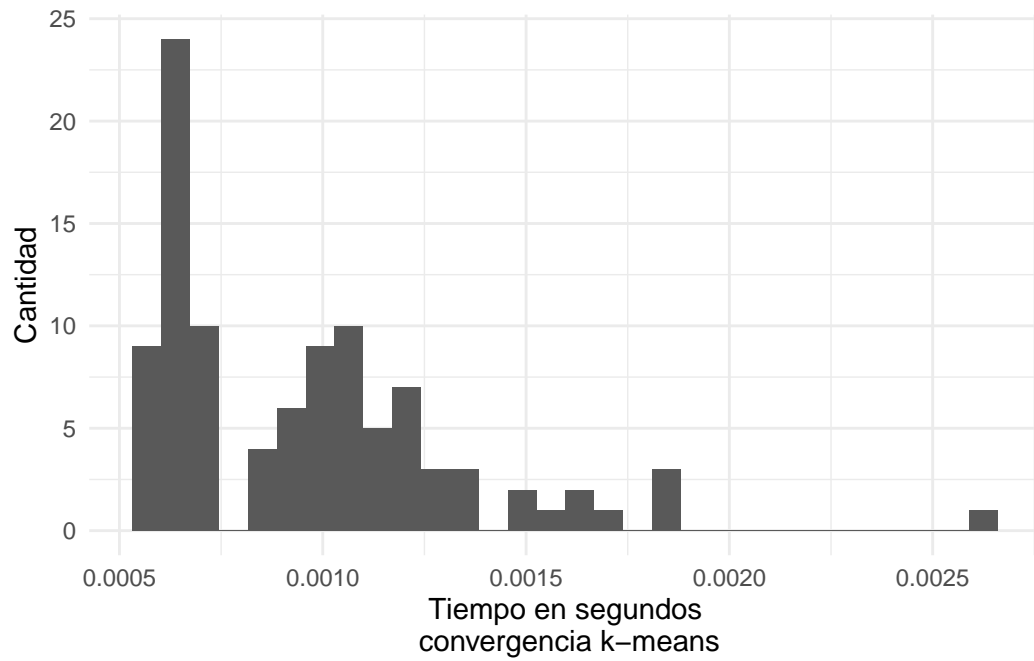
```
$atraccion_mejor_optimo  
[1] 59
```

```
notas_k_3$resumen
```

```
$plot_optimos
```



```
$plot_tiempo
```



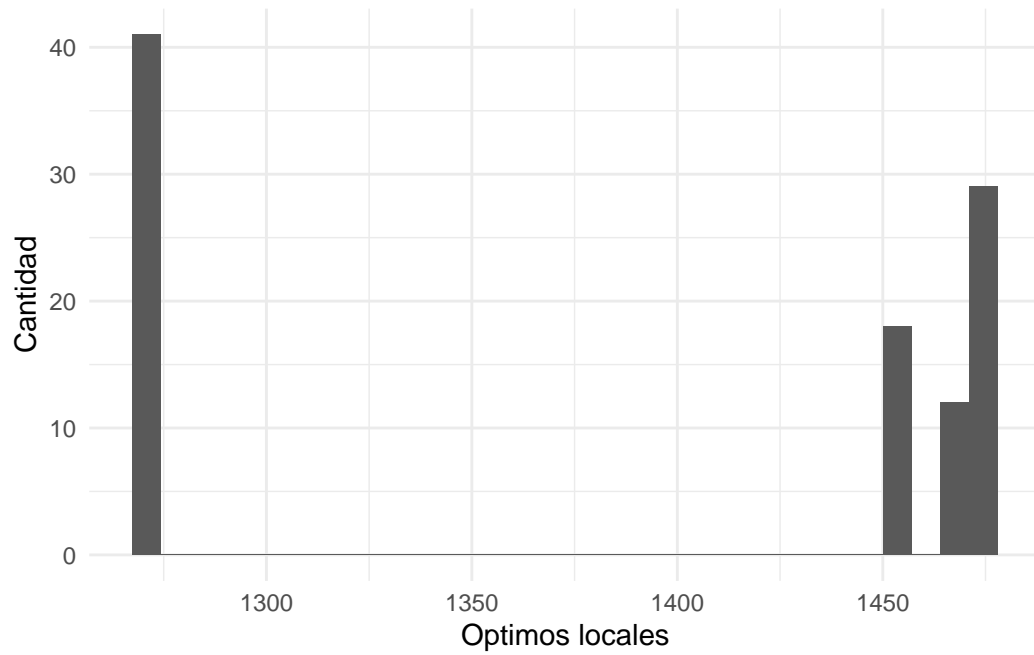
```
$optimo_promedio  
[1] 1762.476
```

```
$mejor_optimo  
[1] 1707.05
```

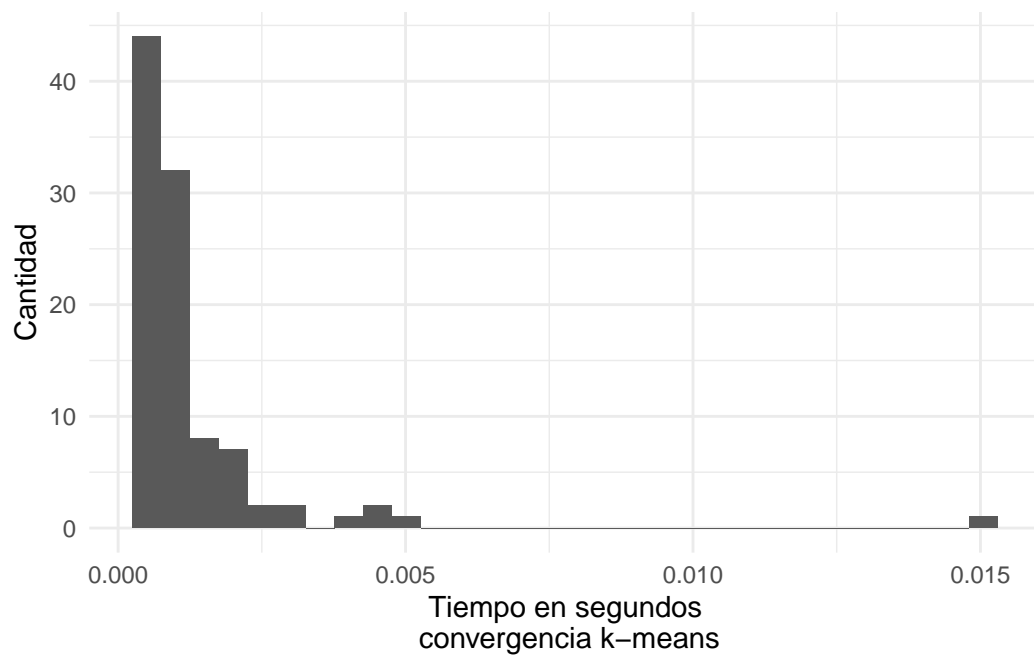
```
$atraccion_mejor_optimo  
[1] 71
```

```
notas_k_4$resumen
```

```
$plot_optimos
```



`$plot_tiempo`





```
$optimo_promedio  
[1] 1384.582
```

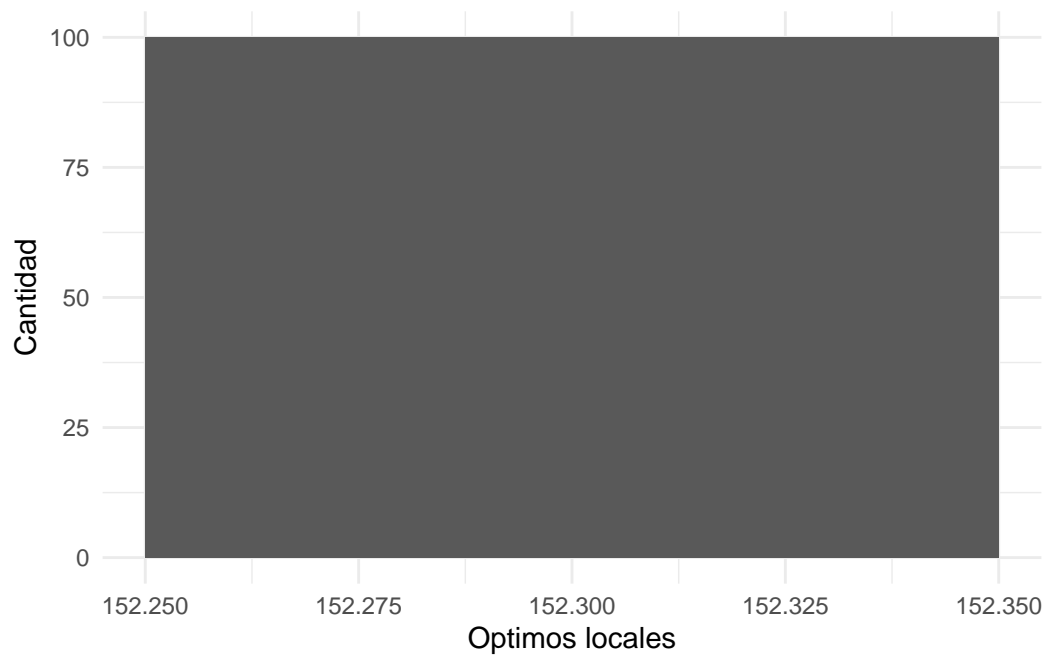
```
$mejor_optimo  
[1] 1269.05
```

```
$atraccion_mejor_optimo  
[1] 41
```

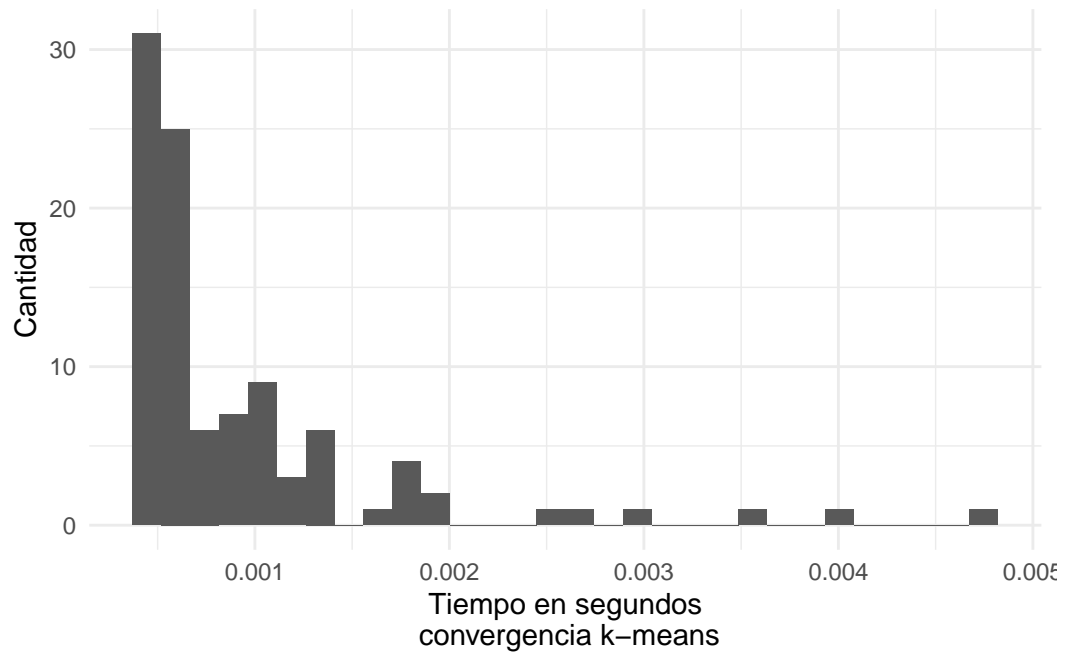
## 1.4 Iris

```
# We estimate some of the point asked  
notas_k_2 <- fn_punto_1(df = iris[, -5], k = 2)  
notas_k_3 <- fn_punto_1(df = iris[, -5], k = 3)  
notas_k_4 <- fn_punto_1(df = iris[, -5], k = 4)  
  
# We print the summary asked for the point  
notas_k_2$resumen
```

```
$plot_optimos
```



```
$plot_tiempo
```



```
$optimo_promedio
```

```
[1] 152.348
```

```
$mejor_optimo
```

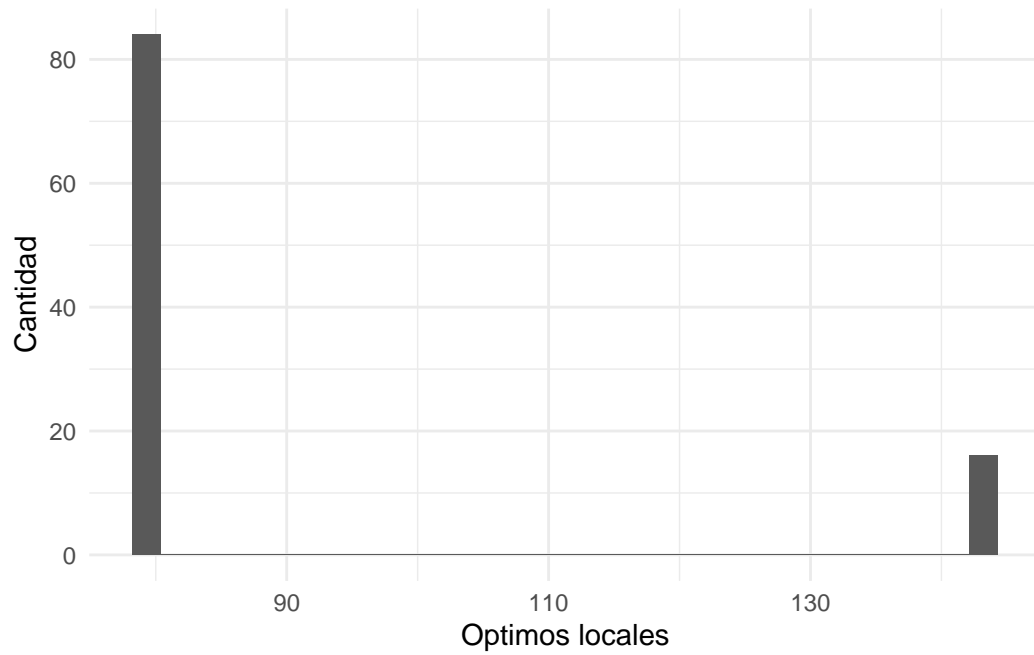
```
[1] 152.348
```

```
$atraccion_mejor_optimo
```

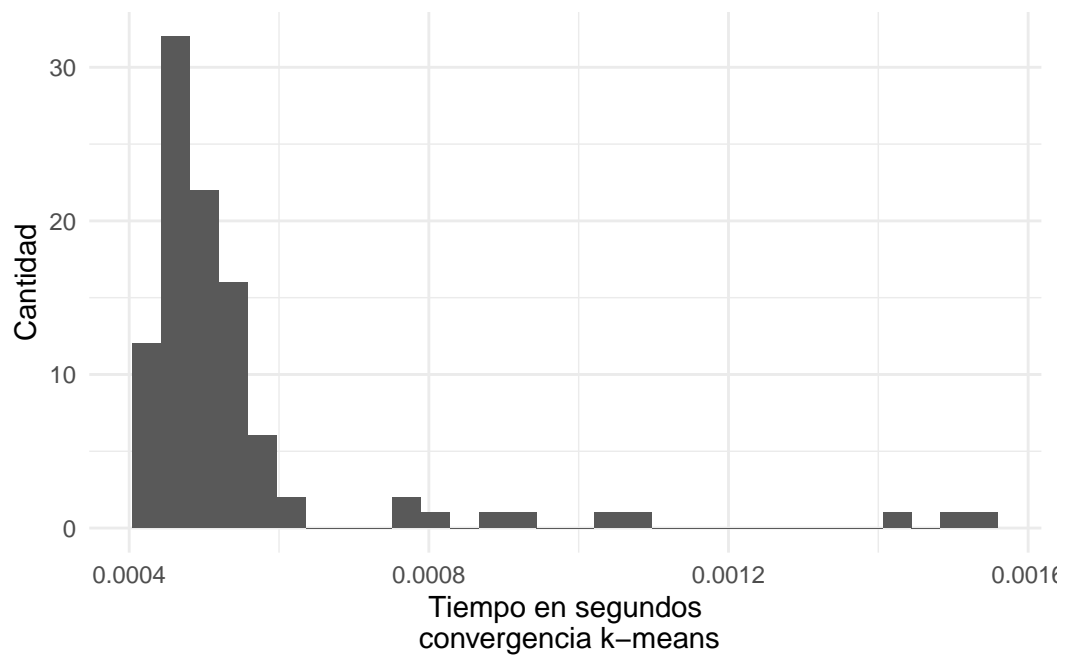
```
[1] 100
```

```
notas_k_3$resumen
```

```
$plot_optimos
```



`$plot_tiempo`



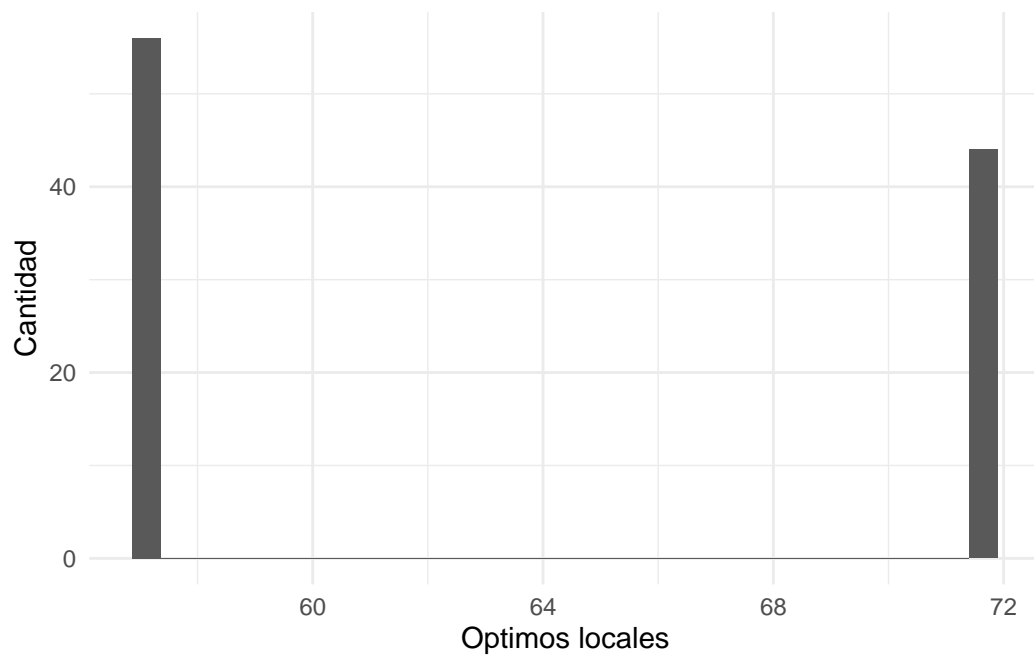
```
$optimo_promedio  
[1] 89.07577
```

```
$mejor_optimo  
[1] 78.85144
```

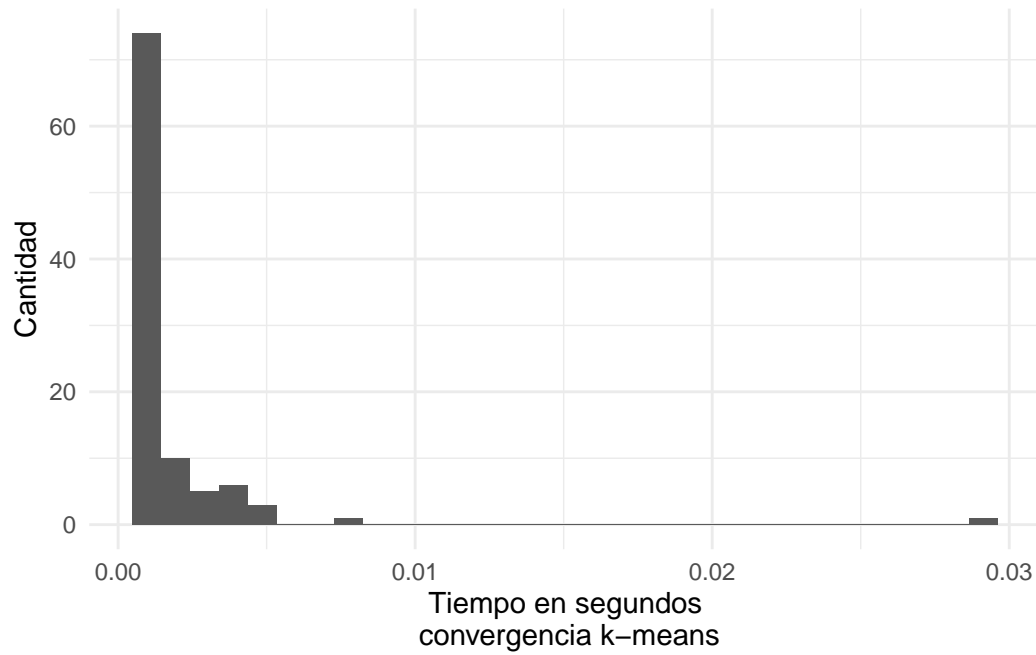
```
$atraccion_mejor_optimo  
[1] 84
```

```
notas_k_4$resumen
```

```
$plot_optimos
```



```
$plot_tiempo
```



```
$optimo_promedio  
[1] 63.53149
```

```
$mejor_optimo  
[1] 57.22847
```

```
$atraccion_mejor_optimo  
[1] 29
```