

Laboratorio 8, Tópicos en análisis datos 1

Joshua Isaac Cervantes Artavia

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
```

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")

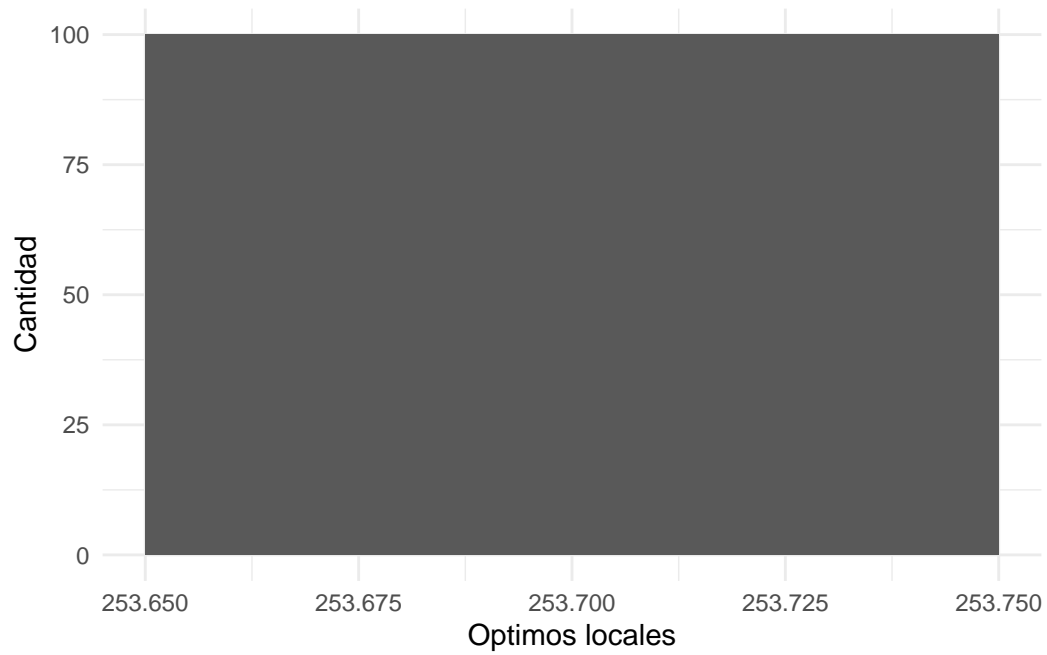
# We make the name of rows the name of the studentes
rownames(df_notas_escolares) <- df_notas_escolares[, 1]

# 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



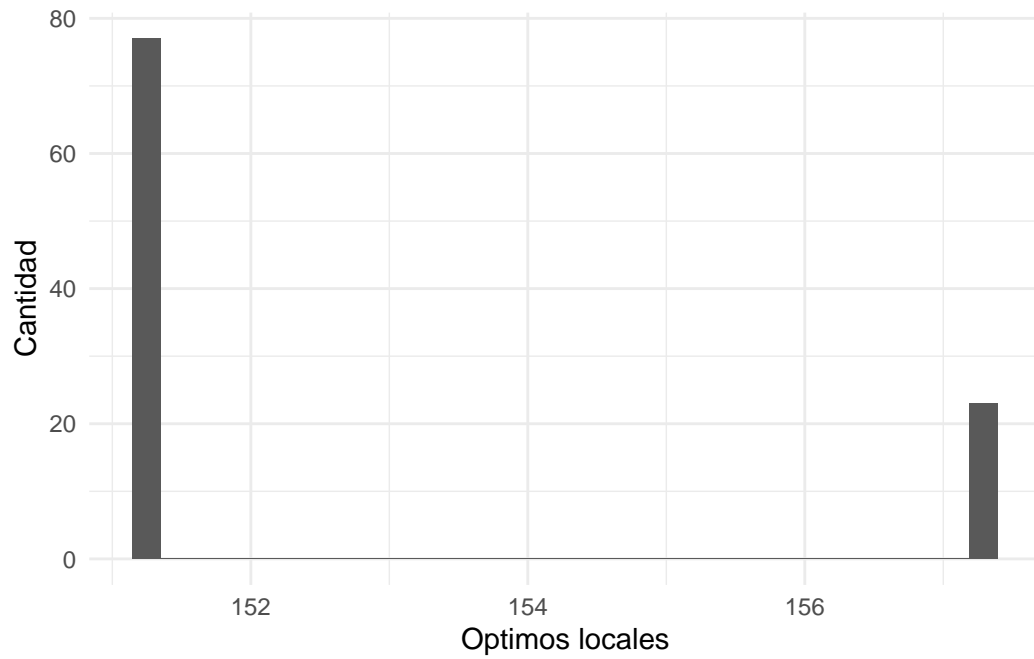
```
$optimo_promedio  
[1] 253.7125
```

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

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

```
notas_k_3$resumen
```

```
$plot_optimos
```



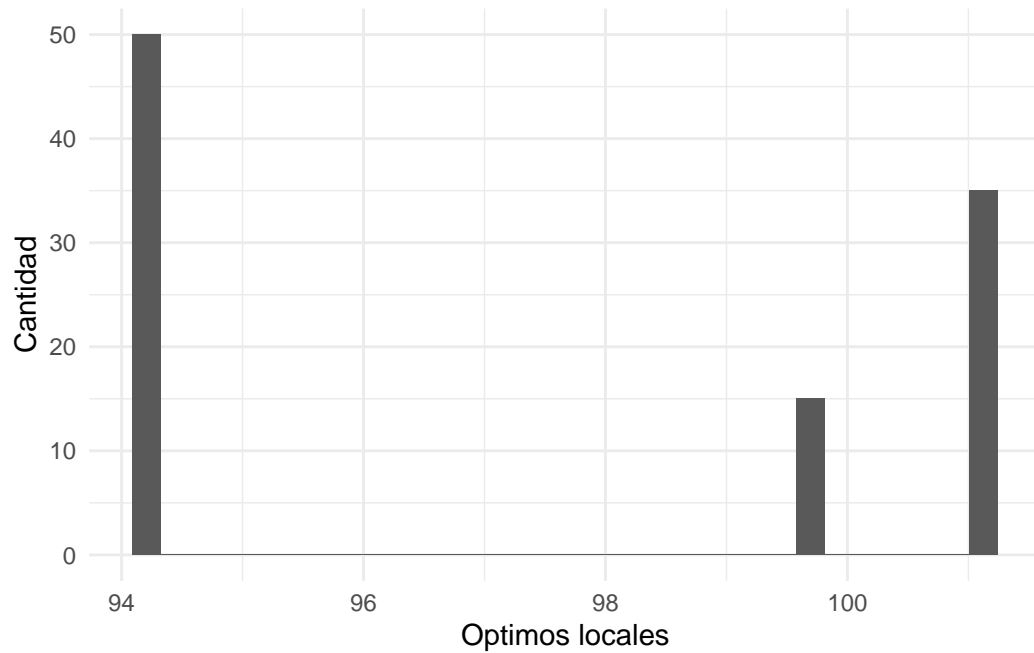
```
$optimo_promedio  
[1] 152.7229
```

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

```
$atraccion_mejor_optimo  
[1] 77
```

```
notas_k_4$resumen
```

```
$plot_optimos
```



```
$optimo_promedio
```

```
[1] 97.44167
```

```
$mejor_optimo
```

```
[1] 94.20833
```

```
$atraccion_mejor_optimo
```

```
[1] 50
```

1.2 Notas Amiard

```
# We read the excel with the data
df_amiard <- read.xlsx("./data/Ejercicios-Cap3.xlsx", "10.Amiard")

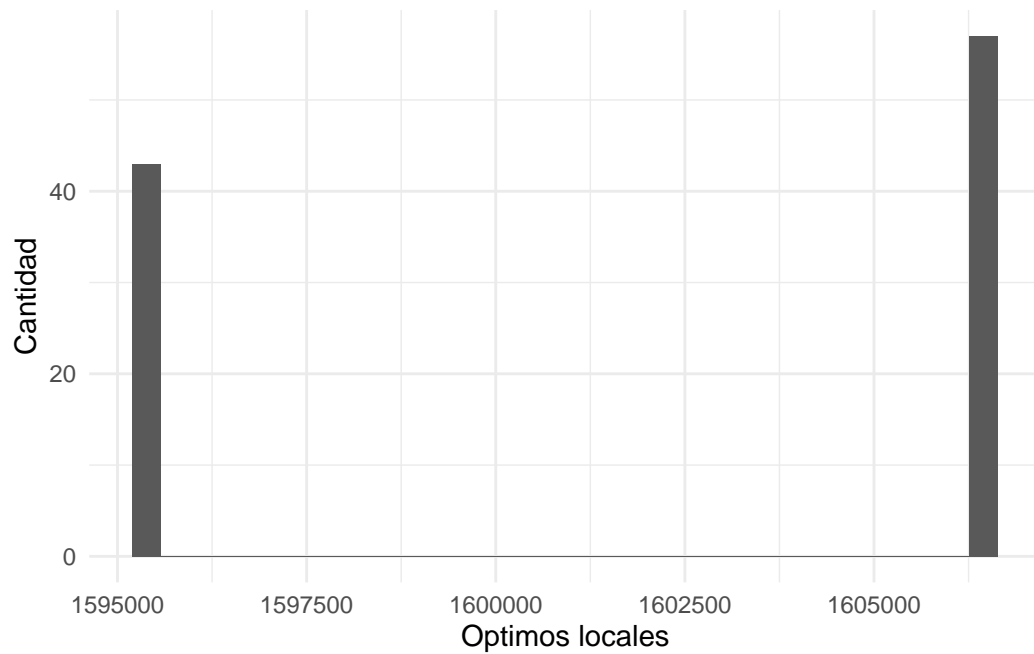
# We make the name of rows the name of the studentes
rownames(df_amiard) <- df_amiard[, 1]

# 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



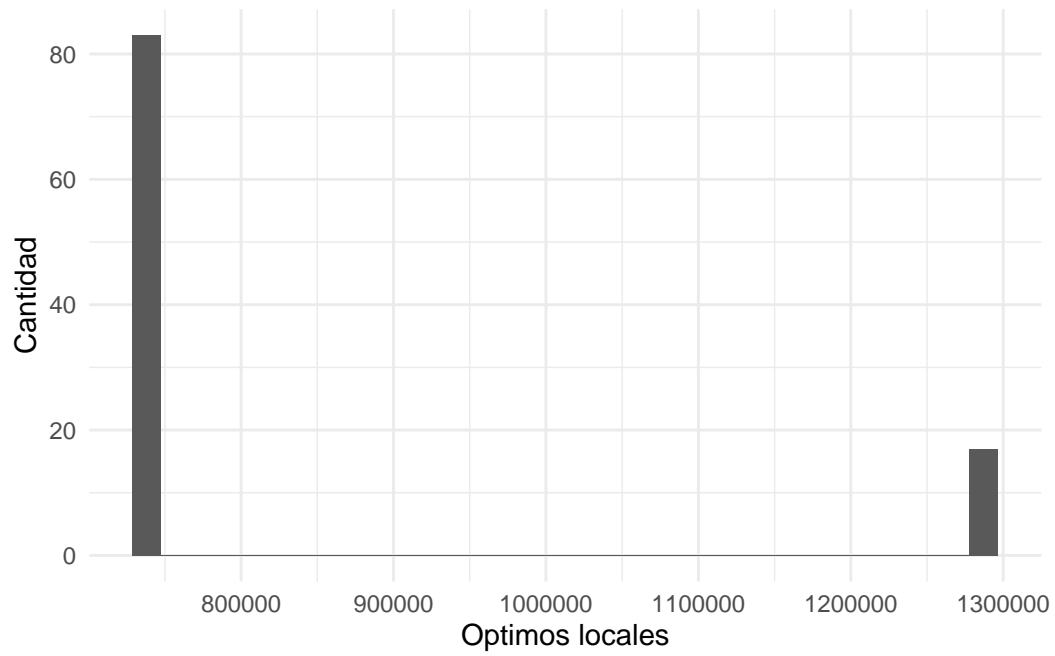
```
$optimo_promedio
[1] 1601775
```

```
$mejor_optimo
[1] 1595470
```

```
$atraccion_mejor_optimo
[1] 43
```

```
notas_k_3$resumen
```

```
$plot_optimos
```



```
$optimo_promedio
```

```
[1] 834229.1
```

```
$mejor_optimo
```

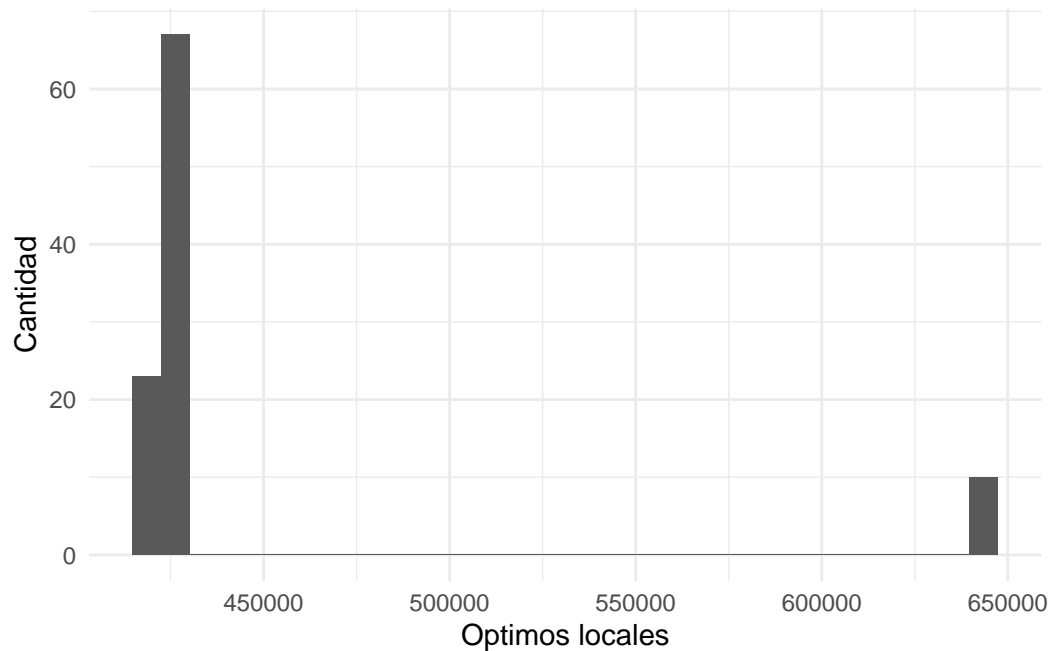
```
[1] 740907.8
```

```
$atraccion_mejor_optimo
```

```
[1] 83
```

```
notas_k_4$resumen
```

```
$plot_optimos
```



```
$optimo_promedio
```

```
[1] 446759.5
```

```
$mejor_optimo
```

```
[1] 420471.9
```

```
$atraccion_mejor_optimo
```

```
[1] 23
```

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]

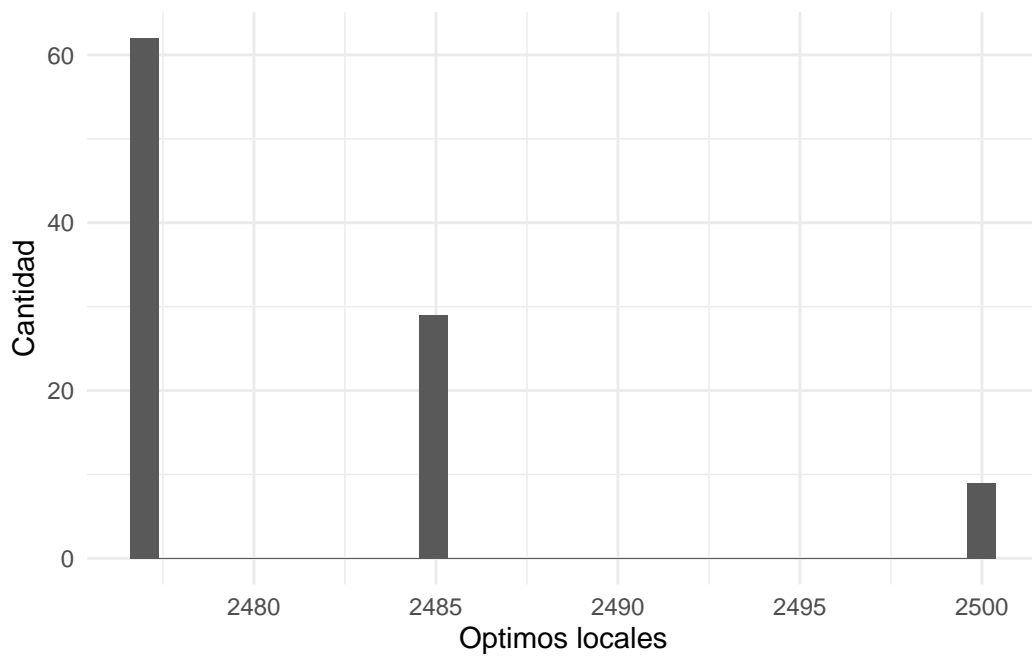
# We delete the first column
df_proteinas <- df_proteinas[, -1]
```



```
# 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
```



```
$optimo_promedio
```

```
[1] 2481.196
```

```
$mejor_optimo
```

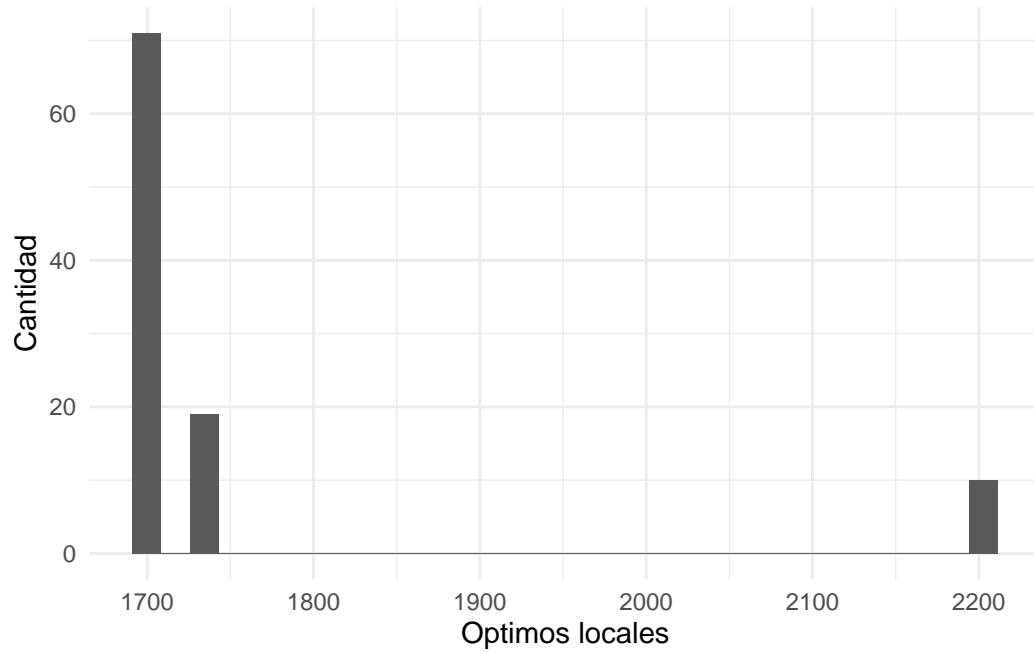
```
[1] 2476.749
```

```
$atraccion_mejor_optimo
```

```
[1] 62
```

```
notas_k_3$resumen
```

```
$plot_optimos
```



```
$optimo_promedio
```

```
[1] 1762.476
```

```
$mejor_optimo
```

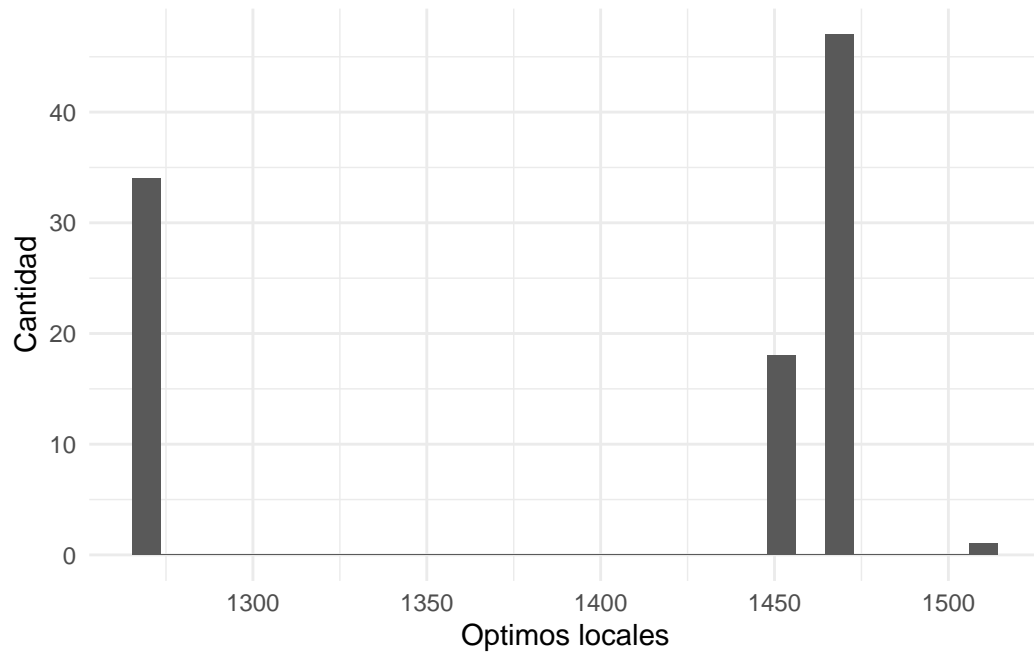
```
[1] 1707.05
```

```
$atraccion_mejor_optimo
```

```
[1] 71
```

```
notas_k_4$resumen
```

```
$plot_optimos
```



```
$optimo_promedio
```

```
[1] 1399.42
```

```
$mejor_optimo
```

```
[1] 1269.05
```

```
$atraccion_mejor_optimo
```

```
[1] 34
```