

Trabajo Práctico N°1

Integrantes: Jeremías Laria Guaza (1629/21), Jerónimo Barragán (1472/21), Manuel Horn (321/21).

```
data <- read.table("ENNyS_menorA2.txt", header=TRUE)

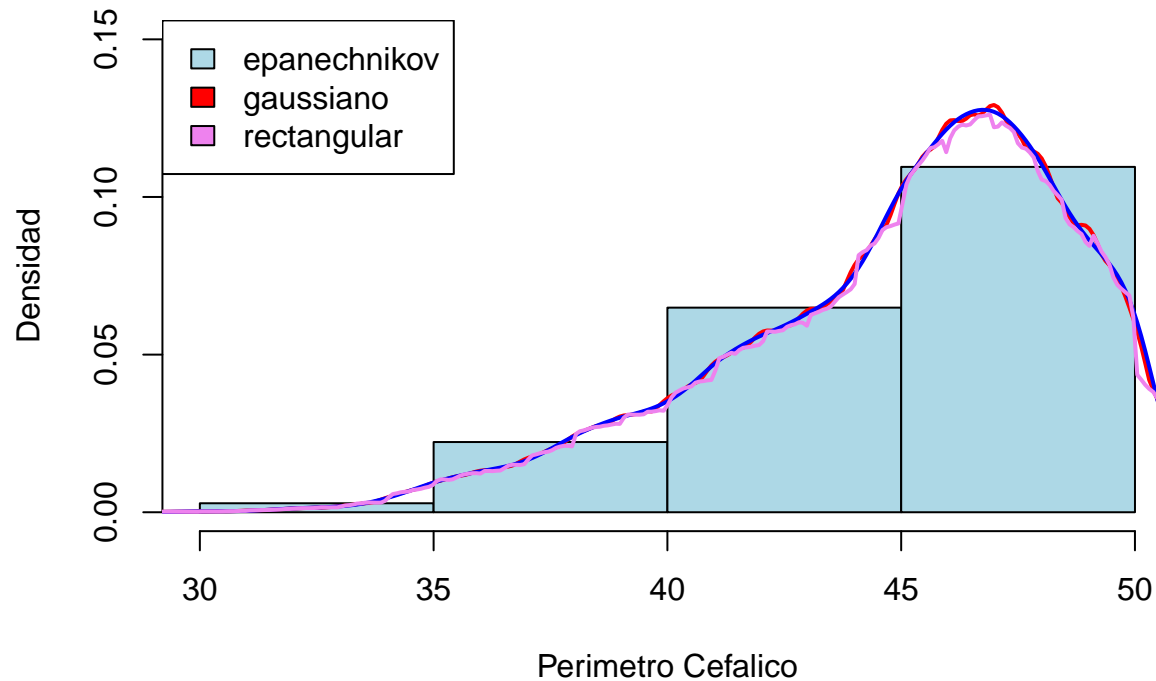
attach(data)

as.factor(Sexo)
as.factor(Tipo_embarazo)
```

Ejercicio 1

```
hist(Perim_cef, xlim = c(30, max(Perim_cef)), ylim=c(0, 0.15), freq = FALSE, col = "lightblue", main = "Comparacion de
↪  nucleos", xlab = "Perimetro Cefalico", ylab = "Densidad")
lines(density(Perim_cef, kernel="epanechnikov"), col = "red", lwd = 2)
lines(density(Perim_cef, kernel="gaussian"), col = "blue", lwd = 2)
lines(density(Perim_cef, kernel="rectangular"), col = "violet", lwd = 2)
legend("topleft",
      legend = c("epanechnikov", "gaussiano", "rectangular"),
      fill = c("lightblue", "red", "violet"),
      bg = "white")
```

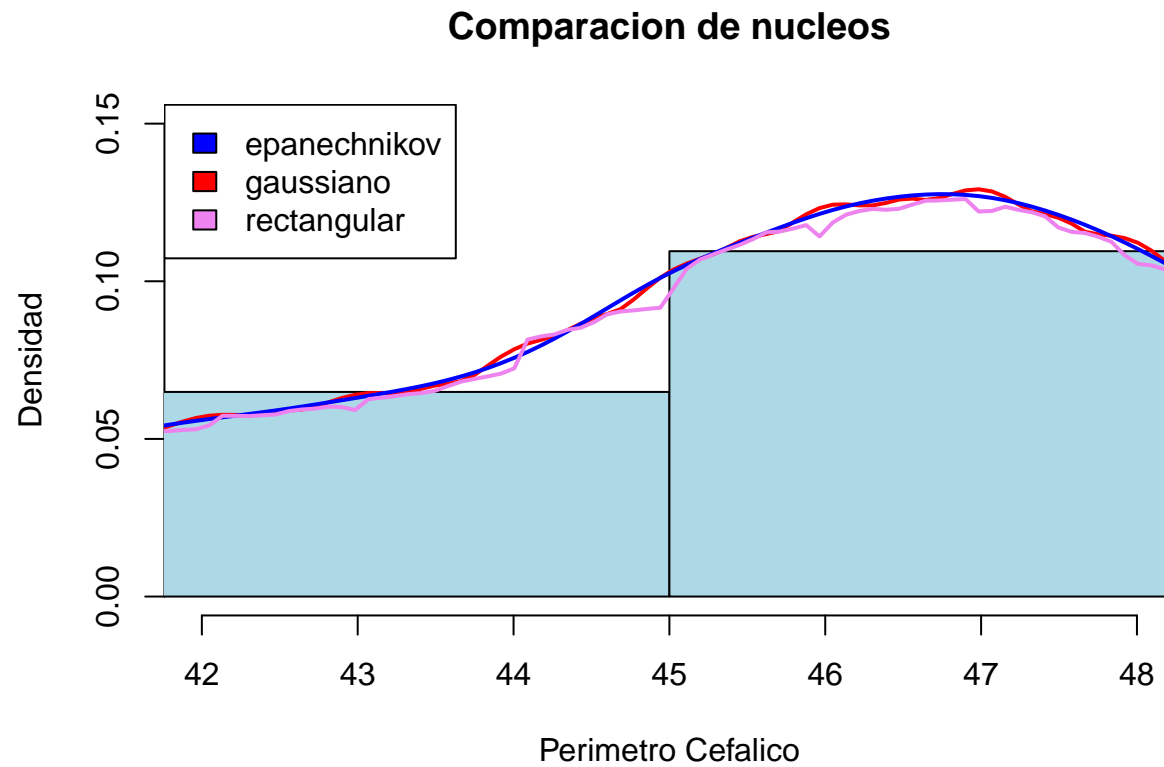
Comparacion de nucleos



En conclusión, podemos ver que las densidades estimadas por los distintos núcleos nos dan como resultado valores muy similares entre si. Podemos analizarlo en una región más acotada para ver si hay diferencias:

```
hist(Perim_cef, xlim = c(42, 48), ylim=c(0, 0.15), freq = FALSE, col = "lightblue", main = "Comparacion de nucleos", xlab =  
  ↪ "Perimetro Cefalico", ylab = "Densidad")  
lines(density(Perim_cef, kernel="epanechnikov"), col = "red", lwd = 2)  
lines(density(Perim_cef, kernel="gaussian"), col = "blue", lwd = 2)  
lines(density(Perim_cef, kernel="rectangular"), col = "violet", lwd = 2)  
legend("topleft",  
  legend = c("epanechnikov", "gaussiano", "rectangular"),  
  fill = c("blue", "red", "violet"),
```

```
bg = "white")
```

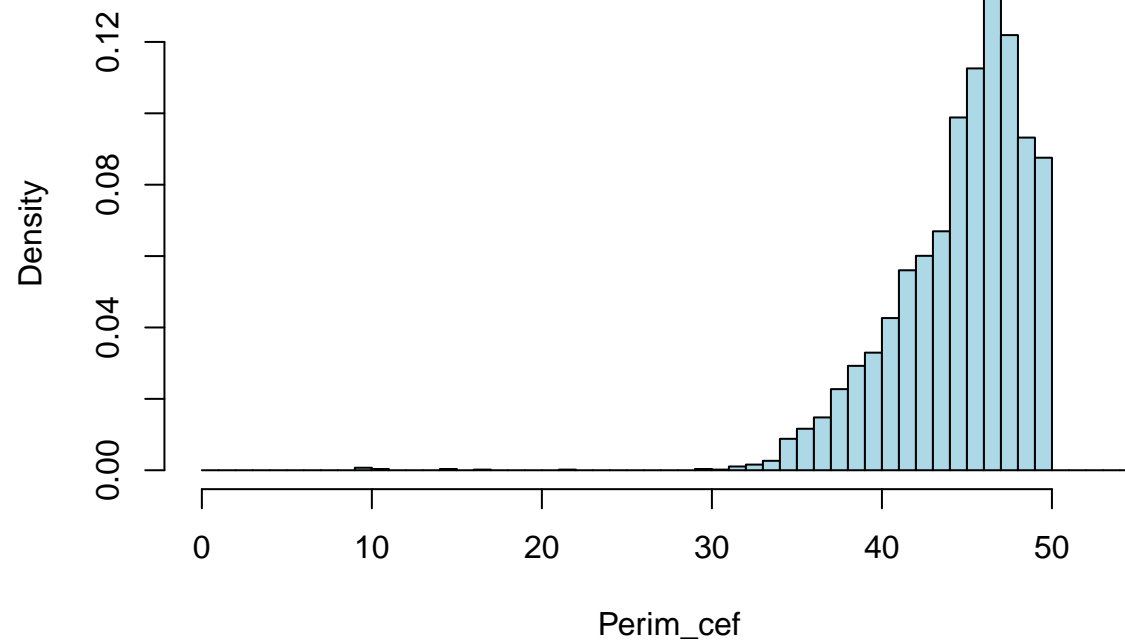


Acá podemos ver cómo tanto el núcleo epanechnikov como el núcleo gaussiano nos dan una gráfica mas “suave” que el núcleo rectangular.

Ejercicio 2

```
histograma <- hist(Perim_cef, breaks = seq(from=0, to=55, by=1), freq = FALSE, col = "lightblue")
```

Histogram of Perim_cef



```
densidad_bines_filtradas <- histograma$density[42:48]
probabilidad_histograma <- sum(densidad_bines_filtradas)

densidad_epanechnikov <- density(Perim_cef, kernel="epanechnikov")
probabilidad_densidad <- integrate(function(x) {approx(densidad_epanechnikov$x, densidad_epanechnikov$y, xout=x)$y},
  ↪ lower=42, upper=48)
```

Para calcular la probabilidad estimada usando el histograma, recordamos que es igual a $\#\{42 \leq \text{Perim_cef} \leq 48\} / n$ donde n es el total de realizaciones. Como el alto de cada bin C_j es $\#\{C_j\} / (n \cdot |C_j|)$ y $|C_j| = 1$ para todo j en nuestro gráfico, tenemos directamente que la probabilidad es la suma de las alturas de los bins entre 42 y 48.

En cuanto a la densidad, para calcular la probabilidad, integramos la función de densidad estimada con el núcleo de epanechnikov entre 42 y 48.

Ejercicio 3

El valor de la ventana R lo ajusta dependiendo del conjunto de datos y del método utilizado. Para Perim_cef, el valor utilizado es:

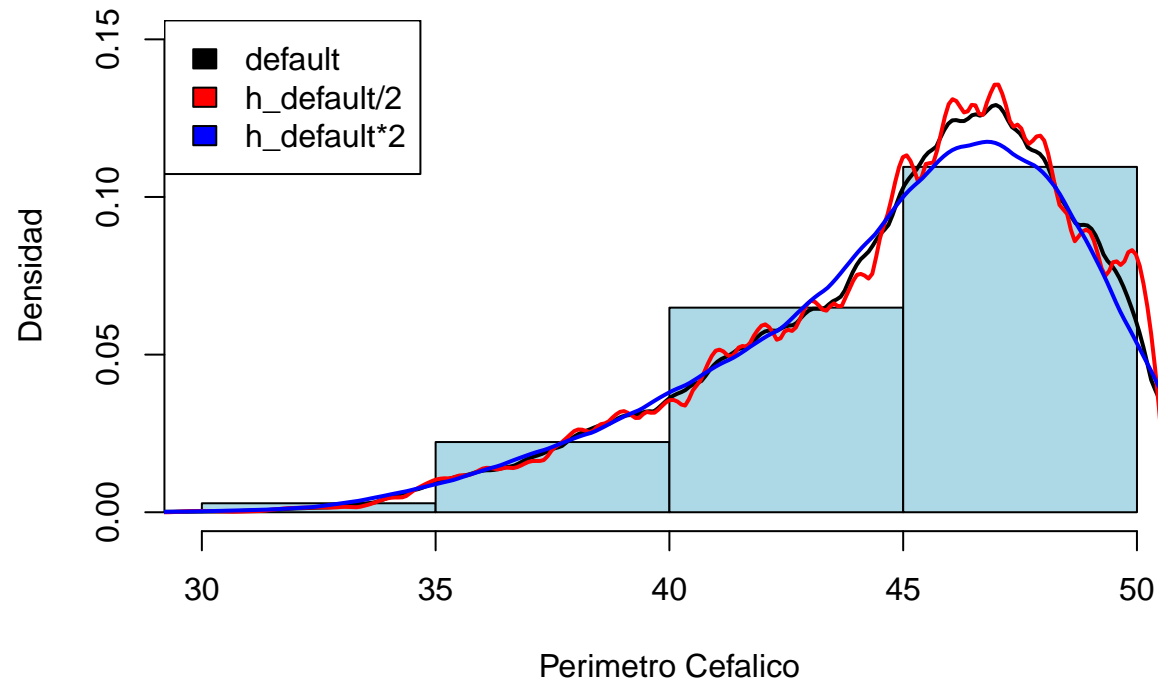
```
h_default <- bw.nrd0(Perim_cef)
h_default
```

```
## [1] 0.5841513
```

Veamos la comparación de los graficos si utilizamos la mitad de la ventana y el doble de esta.

```
hist(Perim_cef, xlim = c(30, max(Perim_cef)), ylim=c(0, 0.15), freq = FALSE, col = "lightblue", main = "Comparacion de
↪ ventanas", xlab = "Perimetro Cefalico", ylab = "Densidad")
lines(density(Perim_cef, kernel="epanechnikov"), col = "black", lwd = 2)
lines(density(Perim_cef, kernel="epanechnikov", bw = h_default/2), xlim=c(30,50), col = "red", lwd = 2)
lines(density(Perim_cef, kernel="epanechnikov", bw = h_default*2), xlim=c(30,50), col = "blue", lwd = 2)
legend("topleft",
      legend = c("default", "h_default/2", "h_default*2"),
      fill = c("black", "red", "blue"),
      bg = "white")
```

Comparacion de ventanas



Podemos ver que utilizando la mitad del valor de la ventana que utiliza por defecto R, el gráfico va a tener más picos, pero se encuentra mas pegado a la densidad estimada con la ventana calculada por defecto. Por otro lado, la densidad estimada con el doble del valor calculado por defecto es mas “suave” pero en comparación a la anterior se aleja más de la densidad calculada por defecto.

Ejercicio 4

```
require (tidyverse)
```

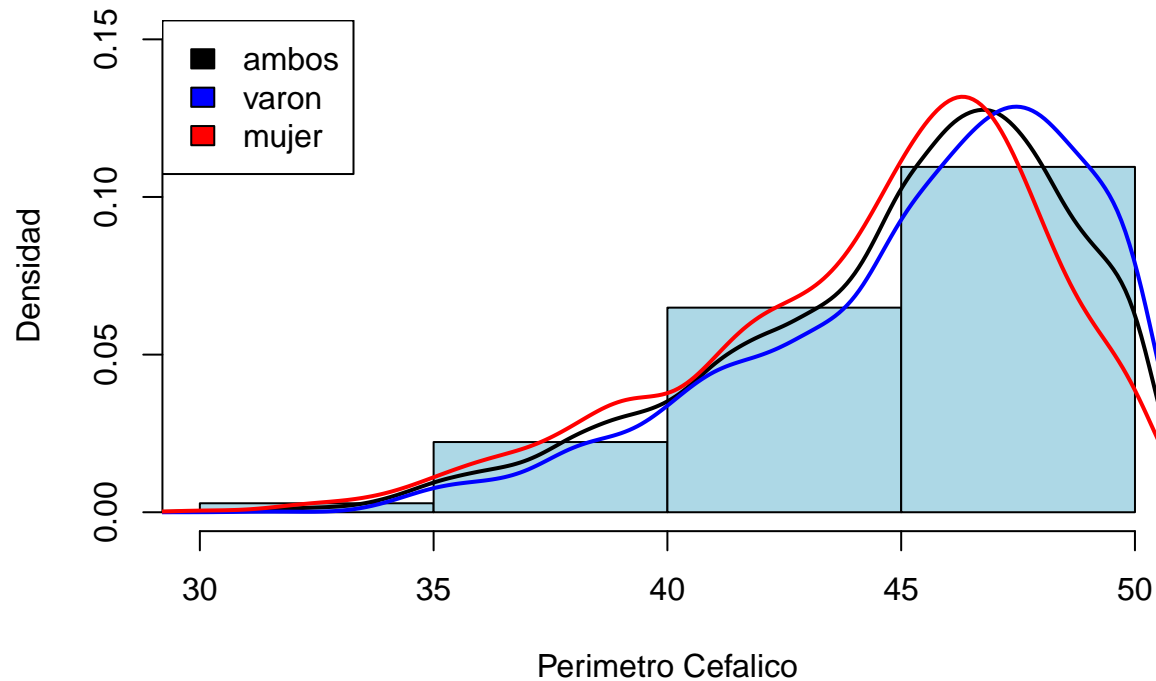
```
## Loading required package: tidyverse
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
Perim_cef_varon <-data%>%filter(Sexo=="Varon")%>%pull(Perim_cef)
Perim_cef_mujer <-data%>%filter(Sexo=="Mujer")%>%pull(Perim_cef)
```

```
hist(Perim_cef, xlim = c(30, max(Perim_cef)), ylim=c(0, 0.15), freq = FALSE, col = "lightblue", main = "Comparacion de
↳ Densidades de Perimetro por Sexo", xlab = "Perimetro Cefalico", ylab = "Densidad")
lines(density(Perim_cef, kernel="gaussian"), col = "black", lwd = 2)
lines(density(Perim_cef_varon, kernel="gaussian"), col = "blue", lwd = 2)
lines(density(Perim_cef_mujer, kernel="gaussian"), col = "red", lwd = 2)
legend("topleft",
      legend = c("ambos", "varon", "mujer"),
      fill = c("black", "blue", "red"),
      bg = "white")
```

Comparacion de Densidades de Perimetro por Sexo



En el gráfico vemos que los varones tienden a tener un perímetro cefálico mas grande respecto a las mujeres. Esto se llega a ver dado que en perímetros cefálicos más bajos hay una mayor densidad de mujeres, y en perímetros cefálicos más grandes hay una mayor densidad de varones. La gráfica de varones está más desplazada hacia la derecha respecto a la de mujeres.

Ejercicio 5

```
Perim_cef_simple <-data%>%filter(Tipo_embarazo=="Simple")%>%pull(Perim_cef)
Perim_cef_multiple<-data%>%filter(Tipo_embarazo=="Multiple")%>%pull(Perim_cef)
```

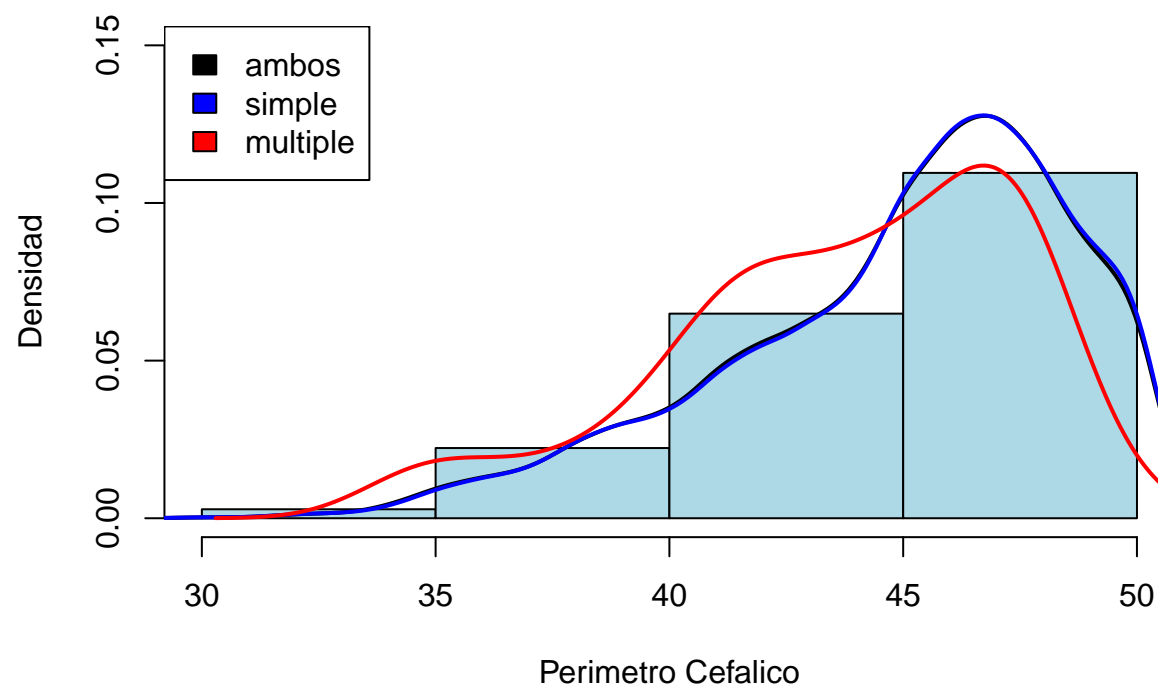


```

hist(Perim_cef, xlim = c(30, max(Perim_cef)), ylim=c(0, 0.15), freq = FALSE, col = "lightblue", main = "Comparacion de
  ↳ Densidades de Perimetro por Tipo de Embarazo", xlab = "Perimetro Cefalico", ylab = "Densidad")
lines(density(Perim_cef, kernel="gaussian"), col = "black", lwd = 2)
lines(density(Perim_cef_simple, kernel="gaussian"), col = "blue", lwd = 2)
lines(density(Perim_cef_multiple, kernel="gaussian"), col = "red", lwd = 2)
legend("topleft",
      legend = c("ambos", "simple", "multiple"),
      fill = c("black", "blue", "red"),
      bg = "white")

```

Comparacion de Densidades de Perimetro por Tipo de Embarazo



Podemos ver que la densidad de los tipos de embarazo simple es casi idéntica a la densidad total, esto se debe a que la cantidad de tipos de embarazo múltiples es muy baja respecto a la cantidad total de datos.

```
table(Tipo_embarazo)
```

```
## Tipo_embarazo  
## Multiple    Simple  
##      190      5486
```

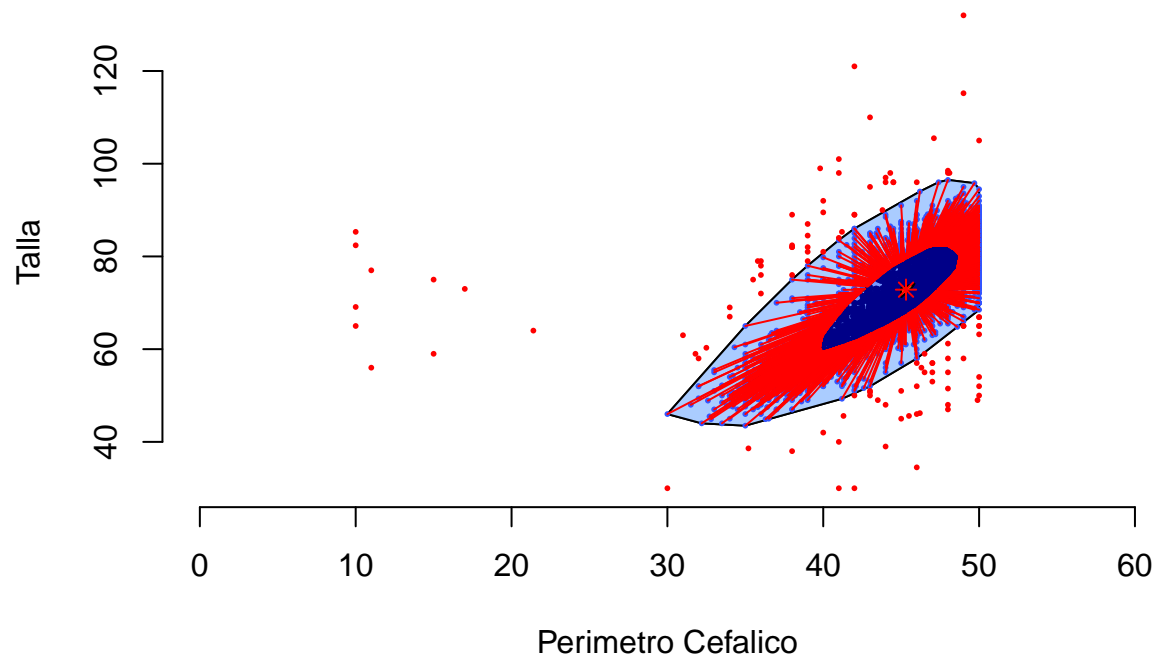
Por otro lado, el perímetro cefálico de los bebés que nacieron en un embarazo múltiple suele ser más chico respecto a los bebés que nacen de un tipo de embarazo simple. Esto se debe a que en perímetros cefálicos más bajos hay una mayor cantidad de tipos de embarazo múltiple y en perímetros cefálicos más altos hay una mayor cantidad de tipos de embarazo simples.

Ejercicio 6

```
library(aplpack)
```

```
## Warning: package 'aplpack' was built under R version 4.2.2
```

```
bagplot(Perim_cef, Talla, xlim = c(0, 60), xlab = "Perimetro Cefalico", ylab = "Talla")
```



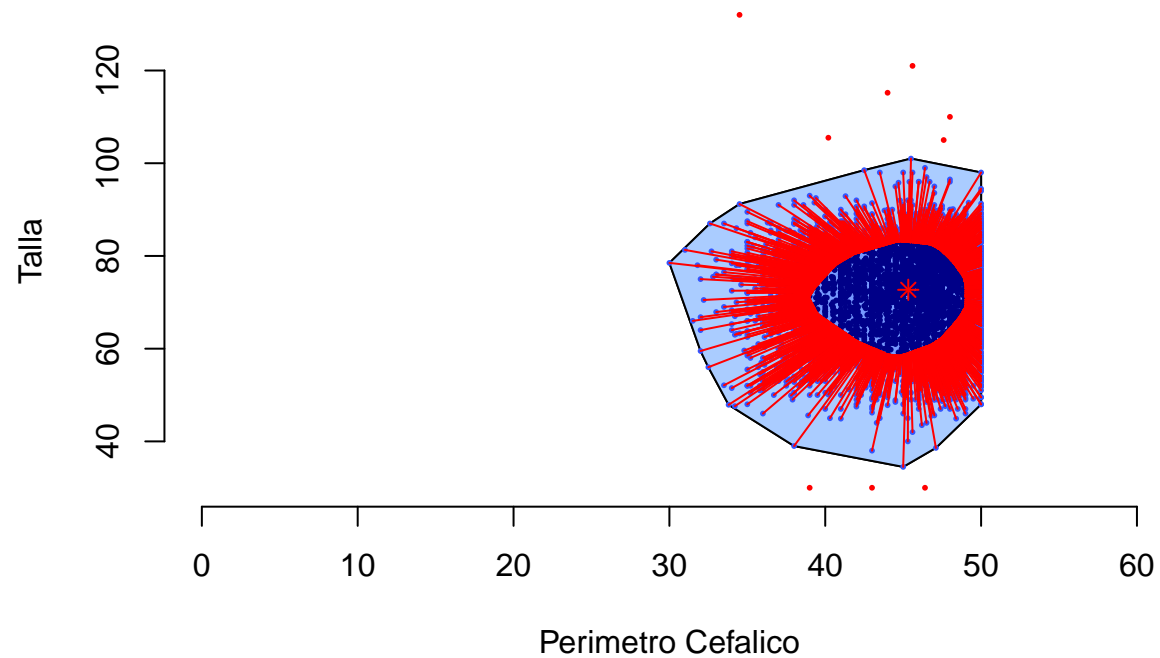
Se registraron algunos valores atípicos, como por ejemplo casos con perímetro cercano a 10 y talla en aproximadamente 90. De la misma manera, hay casos que tienen un perímetro de aproximadamente 45 y de talla cercana a 40.

Ejercicio 7

```
filtro_datos_atipicos <- data%>%filter(Perim_cef > 25)%>%filter(Talla>35)%>%filter(Talla<110)%>%pull(Perim_cef)

bagplot(filtro_datos_atipicos, Talla, xlim = c(0, 60), xlab = "Perimetro Cefalico", ylab = "Talla")
```

```
## Warning in cbind(x, y): number of rows of result is not a multiple of vector
## length (arg 1)
```



Podemos ver que al quitar la mayor parte de datos atípicos del gráfico anterior, el nuevo bagplot sigue conteniendo datos atípicos. Así mismo, hay datos que antes eran atípicos pero ya no lo son en el nuevo gráfico, puesto que ahora se encuentran dentro del polígono azul.

Ejercicio 8

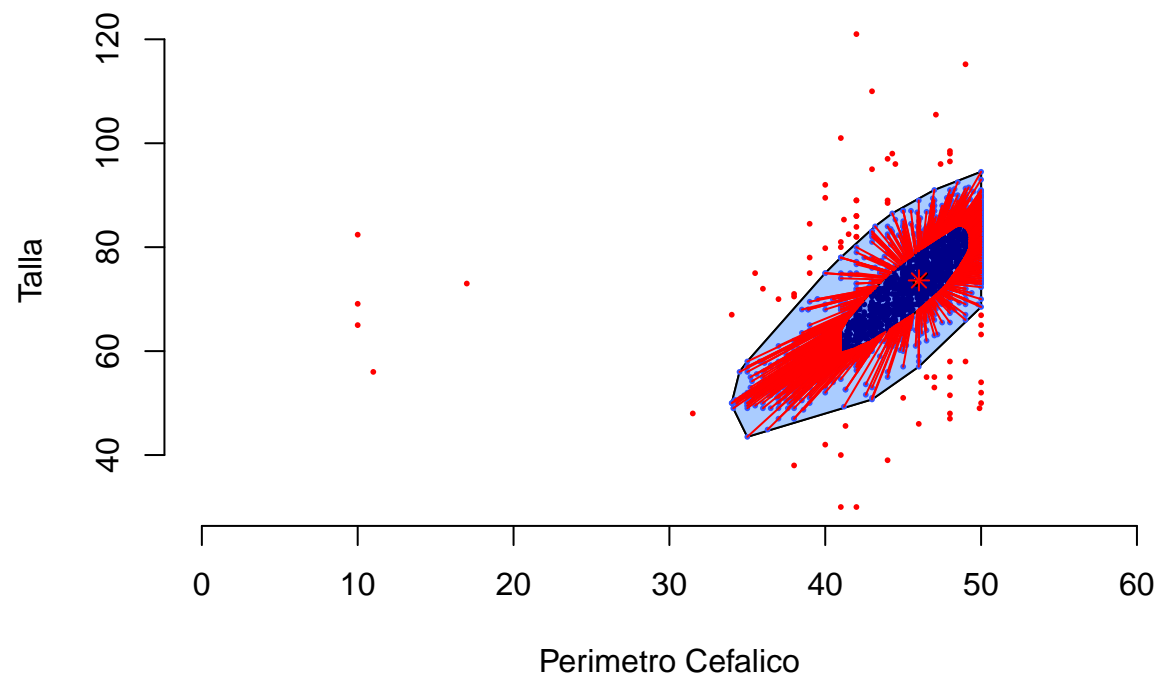
```

Talla_varon <- data%>%filter(Sexo=="Varon")%>%pull(Talla)
Talla_mujer <- data%>%filter(Sexo=="Mujer")%>%pull(Talla)

bagplot(Perim_cef_varon, Talla_varon, xlim = c(0, 60), xlab = "Perimetro Cefalico", ylab = "Talla")
title(main = "Perimetro Cefalico contra Talla en Varones")

```

Perimetro Cefalico contra Talla en Varones

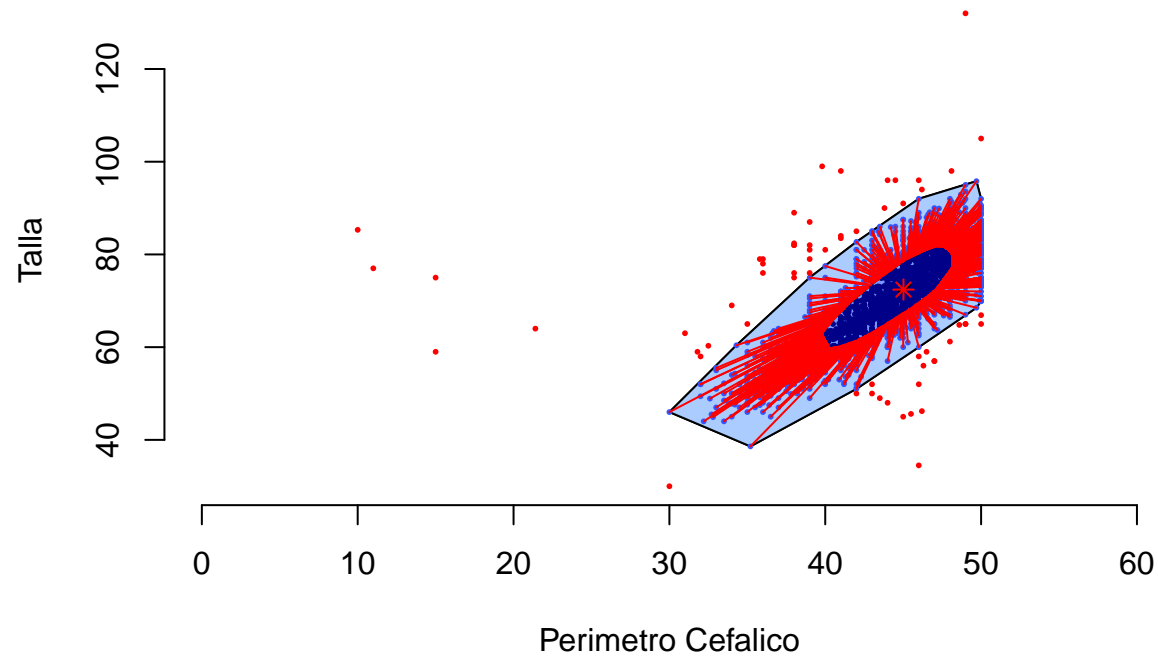


```

bagplot(Perim_cef_mujer, Talla_mujer, xlim = c(0, 60), xlab = "Perimetro Cefalico", ylab = "Talla")
title(main = "Perimetro Cefalico contra Talla en Mujeres")

```

Perimetro Cefalico contra Talla en Mujeres



Se puede ver como el bagplot del perímetro cefálico en varones está levemente desplazado hacia la derecha respecto al de mujeres, lo cual tiene sentido ya que anteriormente graficamos la distribución de estos y pudimos ver que los varones tienden a tener un perímetro cefálico mas grande respecto a las mujeres. Los valores atípicos seguimos viéndolos al igual que en el ítem 6.