

Gráficos

Irais Espejo

2022-07-04

Gráficos

Matriz penguins

Importación de matriz

import data / from excel / browse / seleccionar archivo / aceptar

```
library(readxl)
penguins <- read_excel("penguins.xlsx")
```

Exploración

```
dim(penguins)
```

```
## [1] 344 9
```

Nombre de las variables

```
colnames(penguins)
```

```
## [1] "ID" "especie" "isla" "largo_pico_mm"
## [5] "grosor_pico_mm" "largo_aleta_mm" "masa_corporal_g" "genero"
## [9] "año"
```

Tipo de variables

```
str(penguins)
```

```
## tibble [344 x 9] (S3: tbl_df/tbl/data.frame)
## $ ID : chr [1:344] "i1" "i2" "i3" "i4" ...
## $ especie : chr [1:344] "Adelie" "Adelie" "Adelie" "Adelie" ...
## $ isla : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
## $ largo_pico_mm : num [1:344] 39.1 39.5 40.3 37.8 36.7 39.3 38.9 39.2 34.1 42 ...
## $ grosor_pico_mm : num [1:344] 18.7 17.4 18 18.1 19.3 20.6 17.8 19.6 18.1 20.2 ...
## $ largo_aleta_mm : num [1:344] 181 186 195 190 193 190 181 195 193 190 ...
## $ masa_corporal_g: num [1:344] 3750 3800 3250 3700 3450 ...
## $ genero : chr [1:344] "male" "female" "female" "female" ...
```

```
## $ año : num [1:344] 2007 2007 2007 2007 2007 ...
```

En busca de valores perdidos (NA's)

```
anyNA(penguins)
```

```
## [1] FALSE
```

No hay valores perdidos.

Estadística descriptiva

```
summary(penguins)
```

```
##      ID          especie          isla      largo_pico_mm
## Length:344      Length:344      Length:344      Min.       :32.10
## Class :character Class :character Class :character 1st Qu.:39.20
## Mode  :character Mode  :character Mode  :character Median :44.45
##                                         Mean  :43.92
##                                         3rd Qu.:48.50
##                                         Max.   :59.60
## grosor_pico_mm largo_aleta_mm masa_corporal_g  genero
## Min.       :13.10 Min.       :172.0 Min.       :2700 Length:344
## 1st Qu.:15.60 1st Qu.:190.0 1st Qu.:3550 Class :character
## Median :17.30 Median :197.0 Median :4050 Mode  :character
## Mean    :17.15 Mean    :200.9 Mean    :4202
## 3rd Qu.:18.70 3rd Qu.:213.2 3rd Qu.:4756
## Max.    :21.50 Max.    :231.0 Max.    :6300
## año
## Min.       :2007
## 1st Qu.:2007
## Median :2008
## Mean      :2008
## 3rd Qu.:2009
## Max.      :2009
```

Configuración de la matriz

1. Convertir las variables categóricas a factores

```
penguins$especie<-factor(penguins$especie,
                          levels=c("Adelie", "Gentoo", "Chinstrap"))
```

```
penguins$isla<-factor(penguins$isla,
                      levels=c("Torgersen", "Biscoe", "Dream"))
```

```
penguins$genero<-factor(penguins$genero,
                        levels=c("male", "female"))
```

```
penguins$año<-factor(penguins$año,
                     levels=c("2007", "2008", "2009"))
```

1.1 Verificamos que las variables cambiaron

```
str(penguins)
```

```
## tibble [344 x 9] (S3: tbl_df/tbl/data.frame)
##  $ ID                : chr [1:344] "i1" "i2" "i3" "i4" ...
##  $ especie           : Factor w/ 3 levels "Adelie","Gentoo",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ isla              : Factor w/ 3 levels "Torgersen","Biscoe",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ largo_pico_mm     : num [1:344] 39.1 39.5 40.3 37.8 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ grosor_pico_mm    : num [1:344] 18.7 17.4 18 18.1 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ largo_aleta_mm    : num [1:344] 181 186 195 190 193 190 181 195 193 190 ...
##  $ masa_corporal_g   : num [1:344] 3750 3800 3250 3700 3450 ...
##  $ genero            : Factor w/ 2 levels "male","female": 1 2 2 2 2 1 2 1 2 1 ...
##  $ año              : Factor w/ 3 levels "2007","2008",...: 1 1 1 1 1 1 1 1 1 1 ...
```

1.2 Realizamos el conteo final y visualizamos la matriz de datos

```
summary(penguins)
```

```
##      ID                especie            isla      largo_pico_mm
## Length:344          Adelie       :152   Torgersen: 52   Min.       :32.10
## Class :character    Gentoo       :124   Biscoe    :168   1st Qu.:39.20
## Mode  :character    Chinstrap: 68    Dream     :124   Median  :44.45
##                                     Mean    :43.92
##                                     3rd Qu.:48.50
##                                     Max.    :59.60
## grosor_pico_mm  largo_aleta_mm  masa_corporal_g  genero      año
## Min.       :13.10  Min.       :172.0  Min.       :2700  male   :170  2007:110
## 1st Qu.:15.60  1st Qu.:190.0  1st Qu.:3550  female:174  2008:114
## Median  :17.30  Median :197.0  Median  :4050                2009:120
## Mean    :17.15  Mean    :200.9  Mean    :4202
## 3rd Qu.:18.70  3rd Qu.:213.2  3rd Qu.:4756
## Max.    :21.50  Max.    :231.0  Max.    :6300
```

Este paso no es necesario 2. Creamos una nueva matriz de datos donde se seleccionan las columnas de la 2 a la 9.

```
penguins1<-penguins[,2:9]
```

Librerías para gráficos

1. Descargar el paquete ggplot2
2. Abrir la librería

```
library(ggplot2)
```

Boxplot

1. Creación de un vector de color

```
color=c("paleturquoise1","pink")
```

2. Creación del gráfico

```
BX<-ggplot(penguins, aes(x=genero, y=largo_pico_mm))+  
  geom_boxplot(fill=color)+  
  ggtitle("Boxplot")+  
  xlab("Género")+  
  ylab("largo de la aleta (mm)")+  
  theme_bw()
```

3. Visualización del boxplot

BX

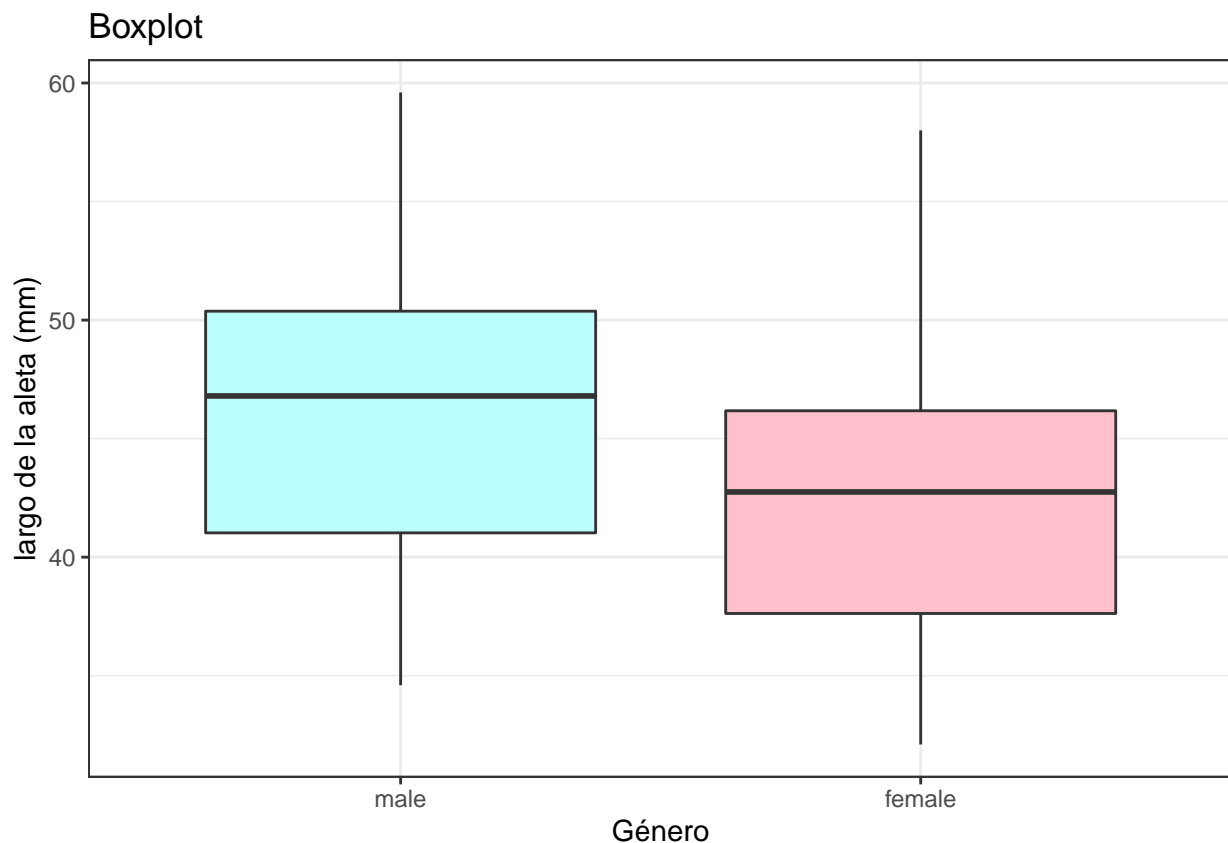


Gráfico de barras

1. Creación de un vector de color

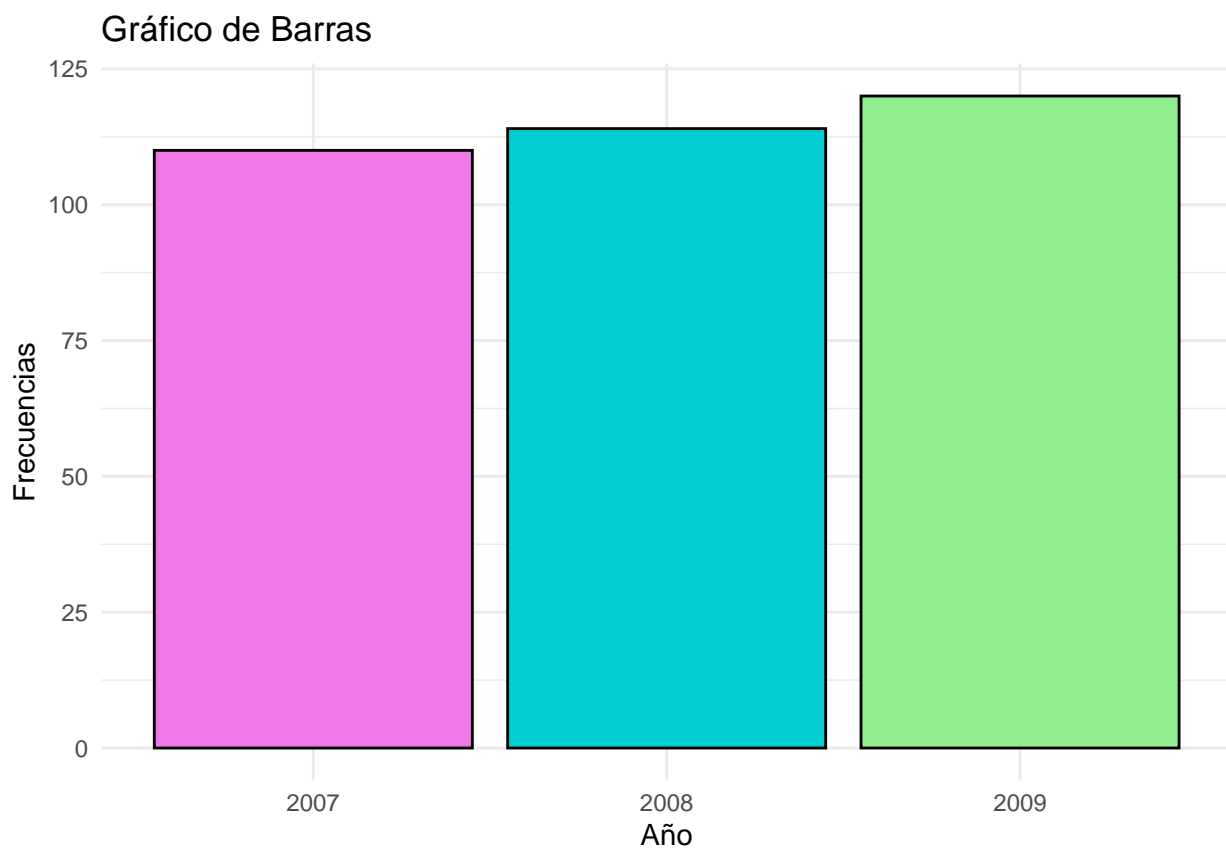
```
color=c("orchid2", "darkturquoise", "lightgreen")
```

2. Creación del gráfico

```
GB1<-ggplot(penguins, aes(x=año))+  
  geom_bar(colour= "black", fill=color)+  
  ggtitle("Gráfico de Barras")+  
  xlab("Año")+  
  ylab("Frecuencias")+  
  theme_minimal()
```

3. Visualización del gráfico

GB1



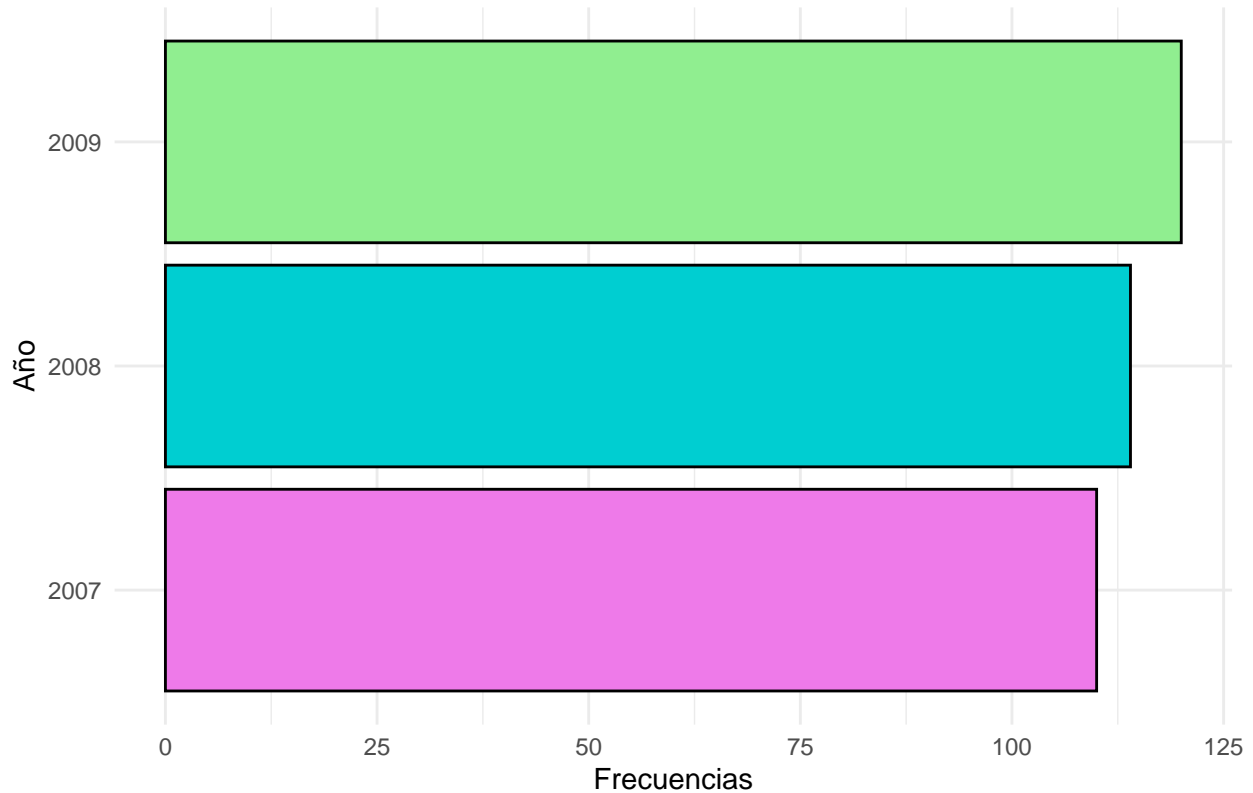
4. Barras horizontales

```
GB2<-ggplot(penguins, aes(x=año))+  
  geom_bar(colour= "black", fill=color)+  
  ggtitle("Gráfico de Barras")+  
  xlab("Año")+  
  ylab("Frecuencias")+  
  coord_flip()+  
  theme_minimal()
```

5. Visualización del objeto

GB2

Gráfico de Barras



Histograma

1. Construcción del gráfico

```
HG<-ggplot(penguins, aes(x=largo_aleta_mm))+  
  geom_histogram(col="black", fill="seagreen2")+  
  ggtitle("Histograma")+  
  xlab("Largo de la aleta (mm)")+  
  ylab("Frecuencias")+  
  theme_classic()
```

2. Visualización del gráfico

HG

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

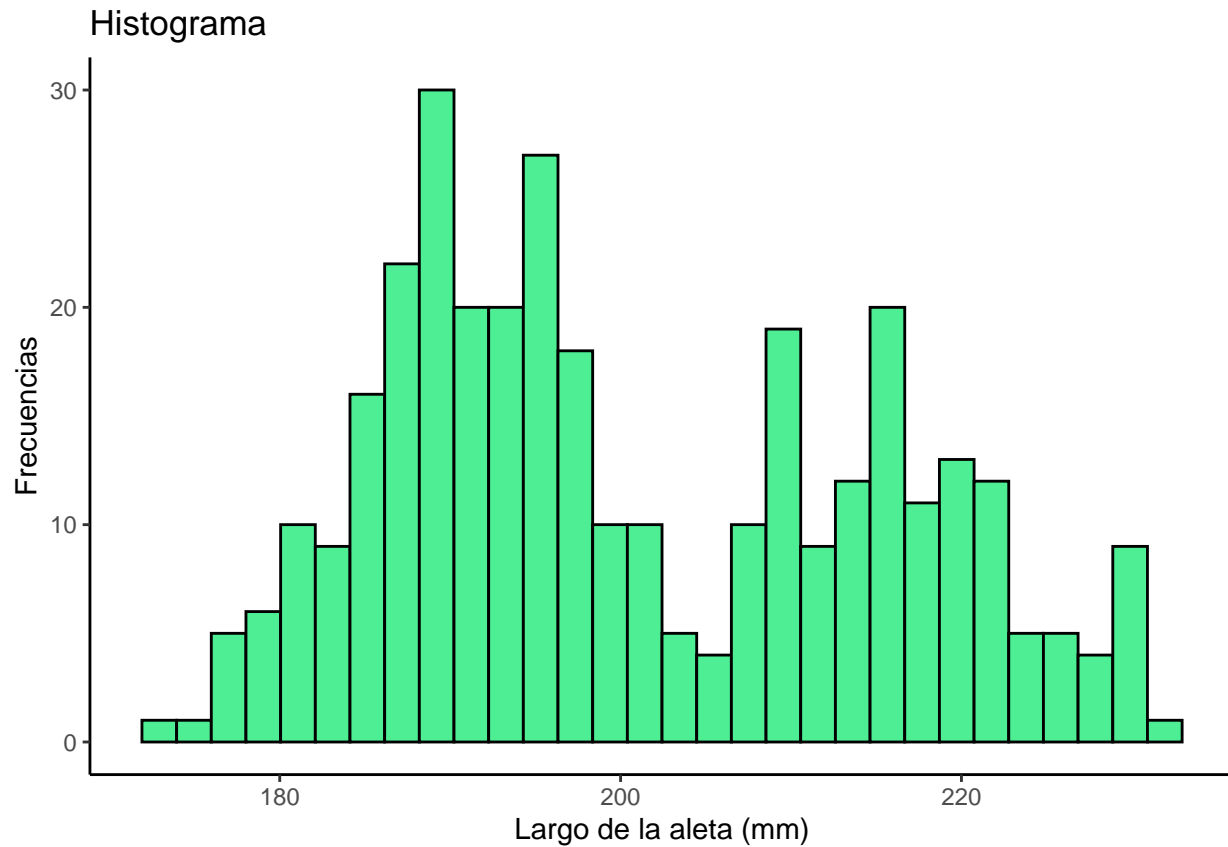


Gráfico de dispersión

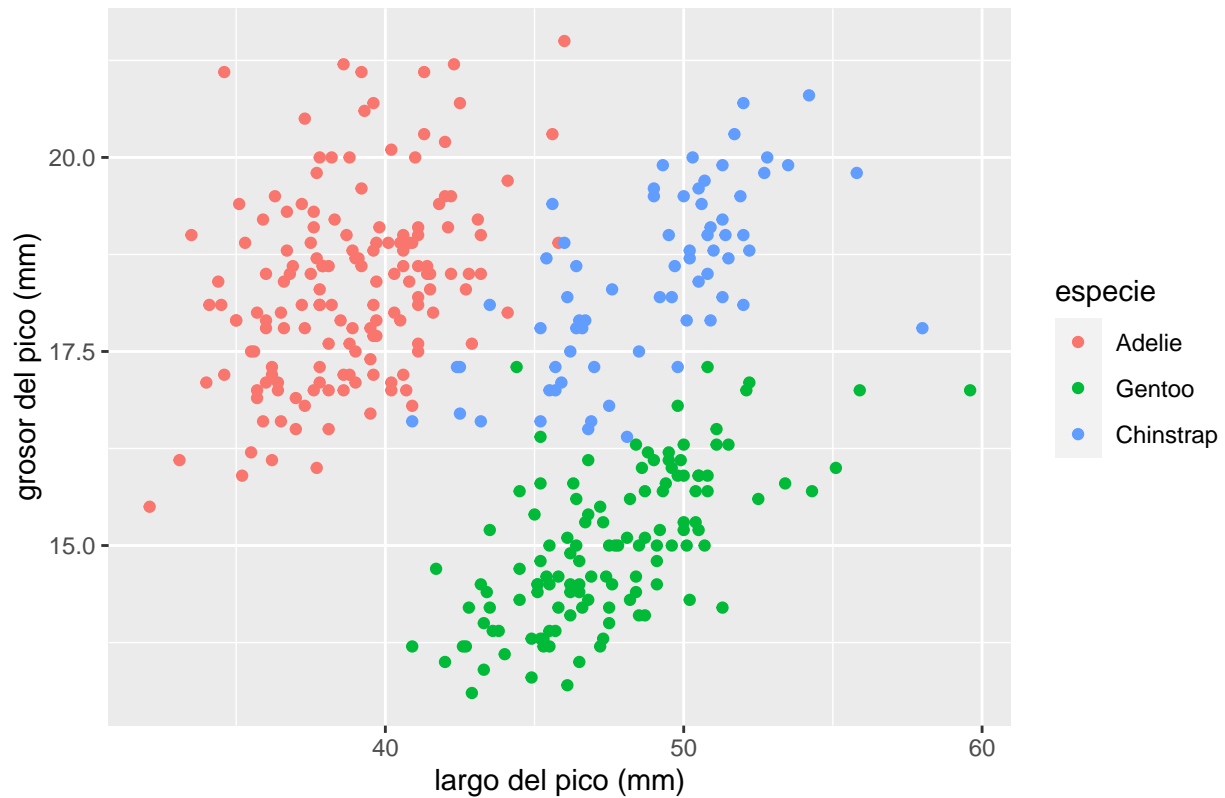
1. Construcción del gráfico

```
GD<-ggplot(penguins, aes(x=largo_pico_mm, y=grosor_pico_mm))+  
  geom_point(aes(color=especie))+  
  ggtitle("Gráfico de dispersión")+  
  xlab("largo del pico (mm)")+  
  ylab("grosor del pico (mm)")+  
  theme_gray()
```

2. Visualización del objeto

```
GD
```

Gráfico de dispersión



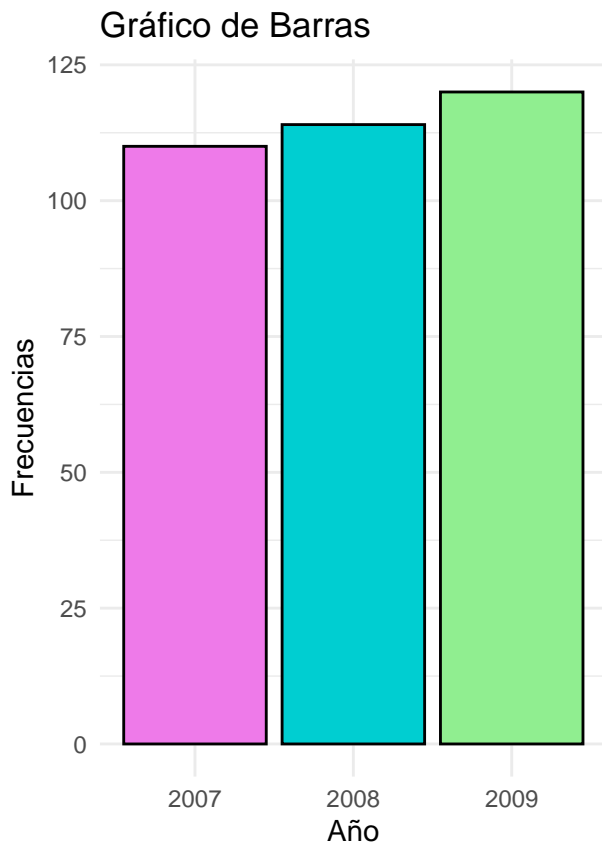
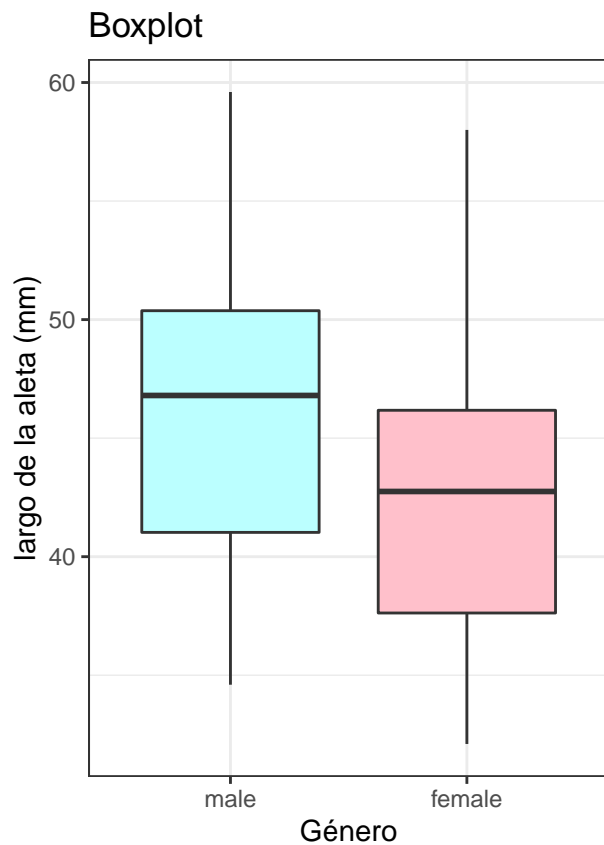
Organización de los gráficos

1. Descargar el paquete gridExtra
2. Abrir la librería

```
library(gridExtra)
```

Organización 2 gráficos en una fila y dos columnas

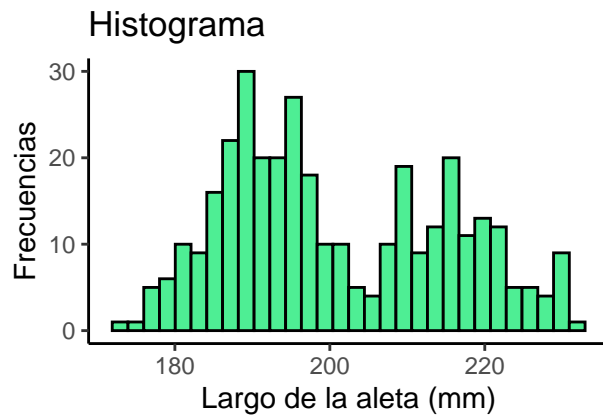
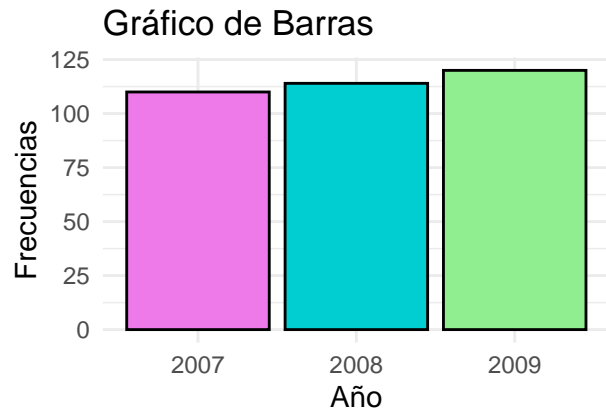
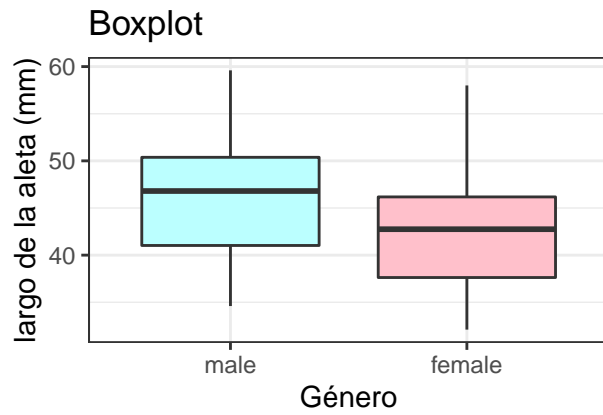
```
grid.arrange(BX,GB1, nrow=1, ncol=2)
```

Organización 3 gráficos en dos filas y dos columnas

```
grid.arrange(BX,GB1,HG, nrow=2, ncol=2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Organización 4 gráficos en dos filas y dos columnas

```
grid.arrange(BX,GB1,HG,GD, nrow=2, ncol=2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

