

Gráficos básicos

Kathya Valdez Gómez

2022-05-18

```
library(readxl)
penguins <- read_excel("penguins.xlsx", col_types = c("skip",
  "text", "text", "numeric", "numeric",
  "numeric", "numeric", "text", "numeric"))
```

Vamos a utilizar la matriz de datos “penguins”

a) *Exportación de la matriz*

Import data/ from excel/ aceptar

b) *Exploración de la matriz*

1.- Dimensión de la matriz

```
dim(penguins)
```

```
## [1] 344 8
```

La matriz penguins tiene 8 variables y 344 individuos (observaciones)

2.- Nombre de las variables

```
colnames(penguins)
```

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

3.- Tipo de variables

```
str(penguins)
```

```
## tibble [344 x 8] (S3: tbl_df/tbl/data.frame)
## $ 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 ...
```

4.- En busca de datos perdidos

```
anyNA(penguins)
```

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

5.- Estadísticas descriptivas (para variables cuantitativas)

```
summary(penguins)
```

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

c) Configuración de la matriz

1.- Convertir las variables categóricas (cualitativas) a factores

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

2.- Se convierte a factor la variable isla

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

3.- Se convierte a factor la variable género

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

5.- Se verifica que las variables se hayan codificado de forma correcta

```
str(penguins)

## tibble [344 x 8] (S3: tbl_df/tbl/data.frame)
## $ 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          : num [1:344] 2007 2007 2007 2007 2007 ...
```

6.- Se realizan de nuevo las estadísticas descriptivas

```
summary(penguins)

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

7.- Creamos una nueva matriz de datos donde sólo se seleccionan las columnas de la 2 a la 9.

```
penguins1<-penguins[,1:8]
```

Dejamos toda la matriz importada con 8 columnas,de la 1 a la 8 porque desde el inicio ya habíamos eliminado la columna de ID

```
penguins1

## # A tibble: 344 x 8
##   especie isla      largo_pico_mm grosor_pico_mm largo_aleta_mm masa_corporal_g
##   <fct>   <fct>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 Adelie Torgersen      39.1            18.7            181            3750
## 2 Adelie Torgersen      39.5            17.4            186            3800
## 3 Adelie Torgersen      40.3            18             195            3250
## 4 Adelie Torgersen      37.8            18.1            190            3700
## 5 Adelie Torgersen      36.7            19.3            193            3450
```

```
## 6 Adelie Torgersen      39.3      20.6      190      3650
## 7 Adelie Torgersen      38.9      17.8      181      3625
## 8 Adelie Torgersen      39.2      19.6      195      4675
## 9 Adelie Torgersen      34.1      18.1      193      3475
## 10 Adelie Torgersen      42       20.2      190      4250
## # ... with 334 more rows, and 2 more variables: genero <fct>, año <dbl>
```

d) *Librerías*

Vamos a utilizar la librería ggplot2

1.- Descargamos la librería del CRAN

```
install.packages("ggplot2")
```

2.- Abrimos la librería

```
library(ggplot2)
```

e) *Boxplot*

1.- Creación de un vector de color

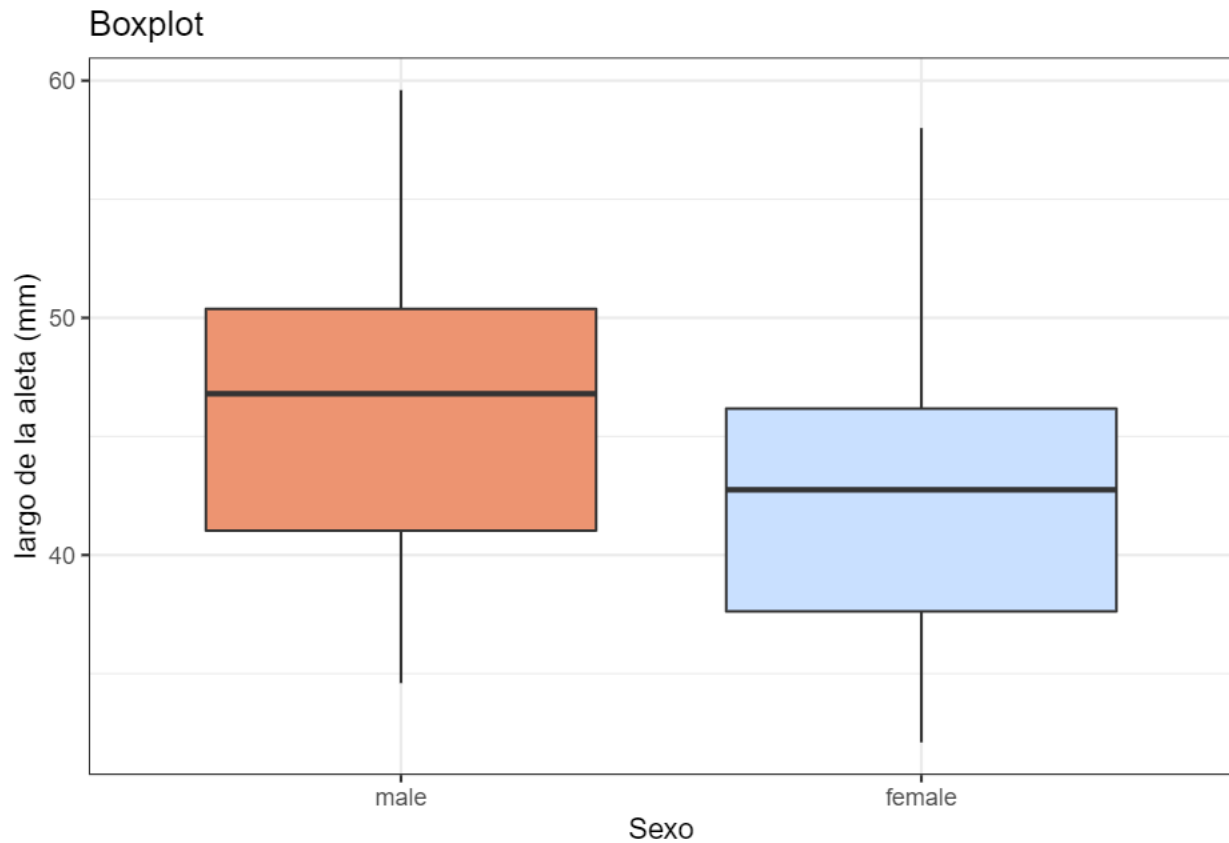
```
color=c("lightsalmon2","lightsteelblue1")
```

2.- Creación del gráfico

```
BX<-ggplot(penguins1, aes(x=genero, y=largo_pico_mm))+
  geom_boxplot(fill=color)+
  ggtitle("Boxplot")+
  xlab("Sexo")+
  ylab("largo de la aleta (mm)")+
  theme_bw()
```

3.- Visualización del boxplot

```
BX
```



f) *Gráfico de barras*

1.- Creación de un vector de color

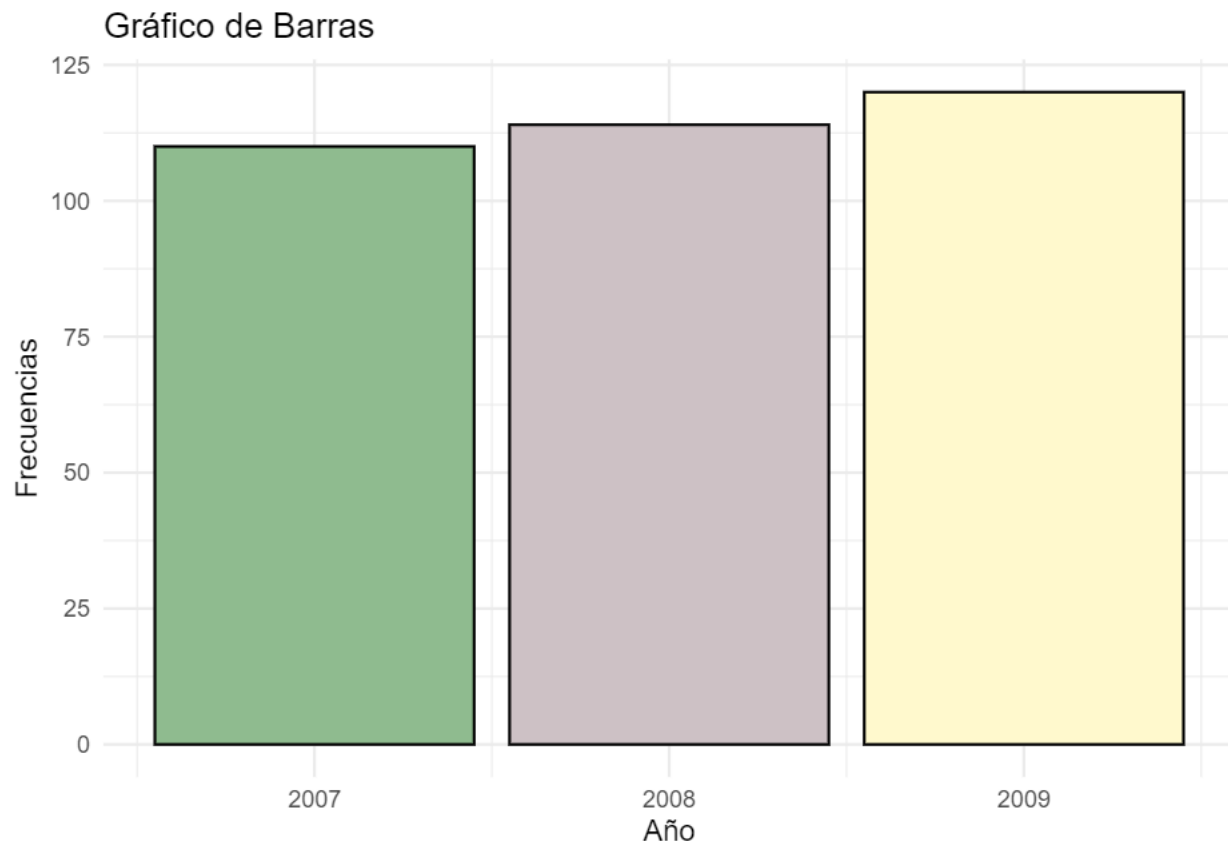
```
color=c("darkseagreen", "lavenderblush3", "lemonchiffon1")
```

2.- Creación del gráfico

```
GB1<-ggplot(penguins1, aes(x=año))+
  geom_bar(colour= "black", fill=color)+
  ggtitle("Gráfico de Barras")+
  xlab("Año")+
  ylab("Frecuencias")+
  theme_minimal()
```

3.- Visualización del gráfico

```
GB1
```



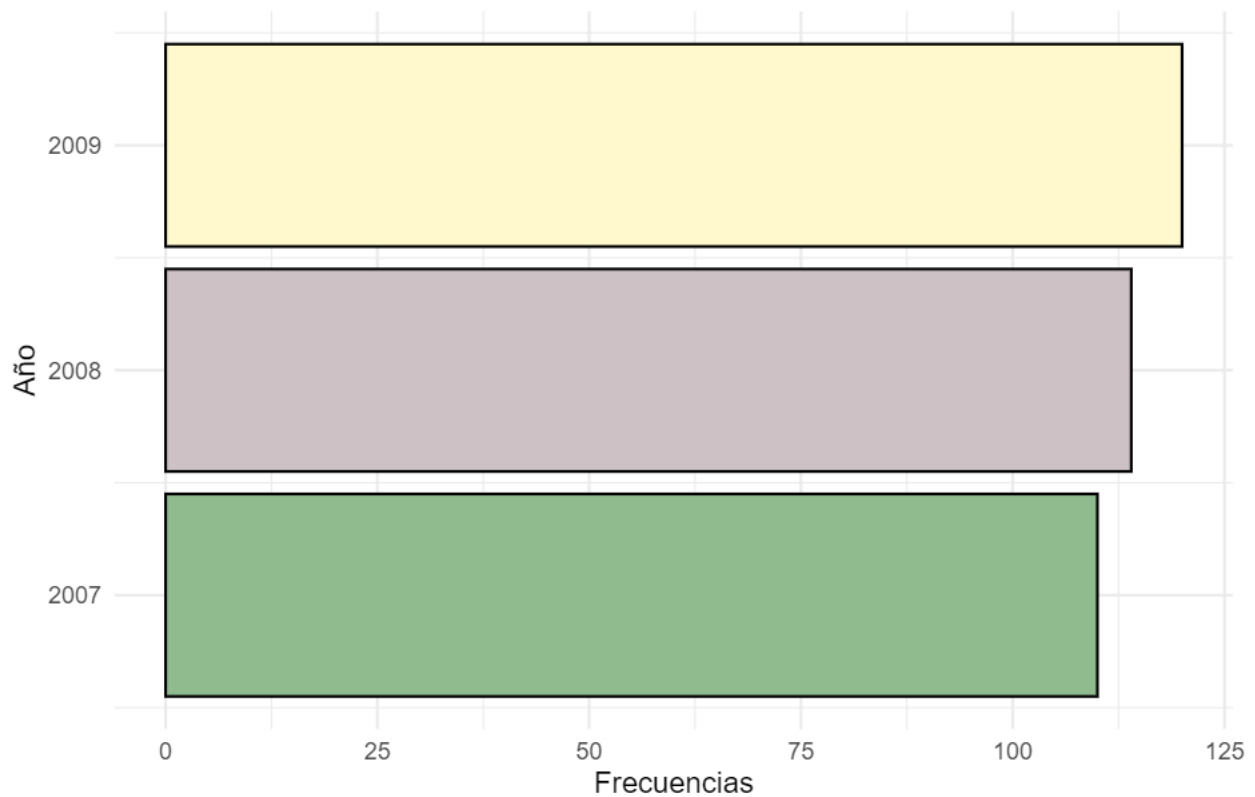
4.- Se agrega la capa de coord_flip para visualizar las barras horizontales

```
GB2<-ggplot(penguins1, aes(x=año))+  
  geom_bar(colour= "black", fill=color)+  
  ggtitle("Gráfico de Barras")+  
  xlab("Año")+  
  ylab("Frecuencias")+  
  coord_flip()+  
  theme_minimal()
```

5.- Visualización del gráfico

GB2

Gráfico de Barras



g) *Histograma*

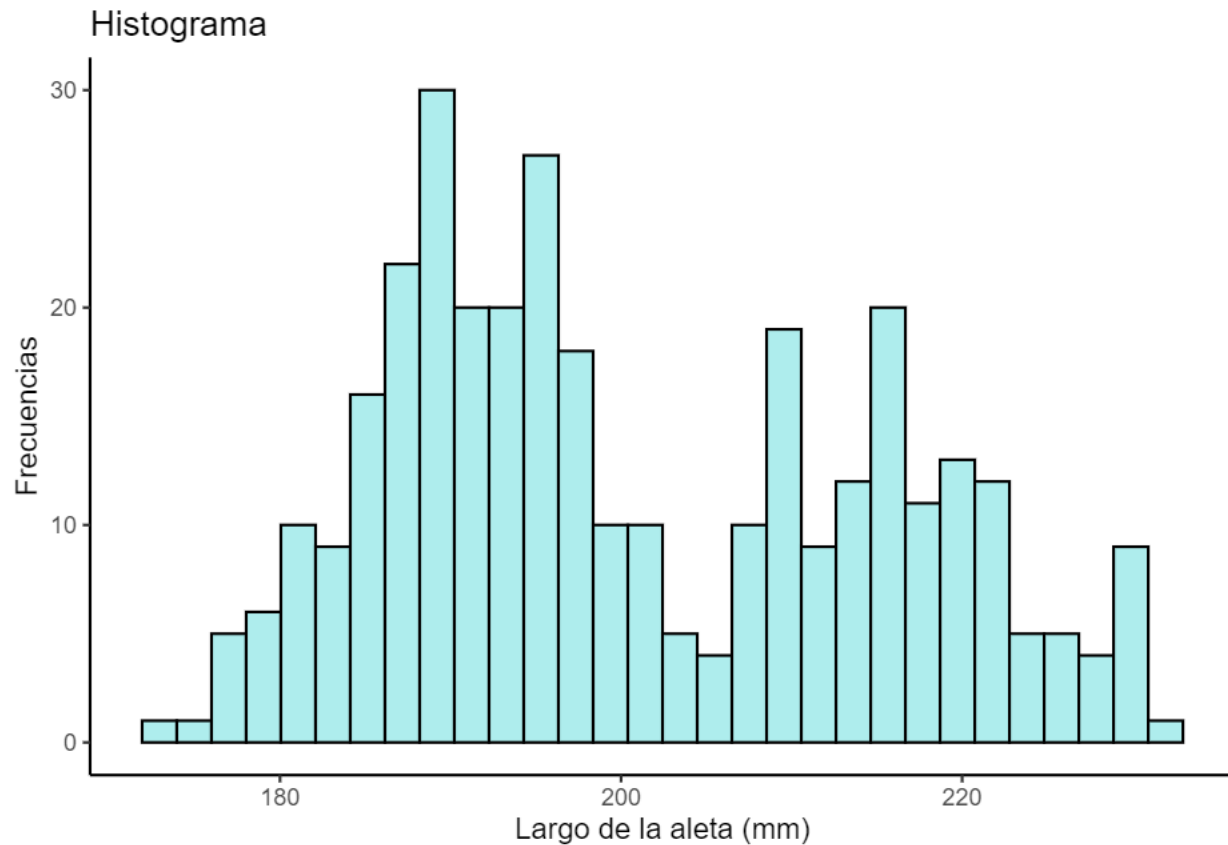
1.- Construcción del gráfico

```
HG<-ggplot(penguins1, aes(x=largo_aleta_mm))+  
  geom_histogram(col="black", fill="paleturquoise")+  
  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`.
```

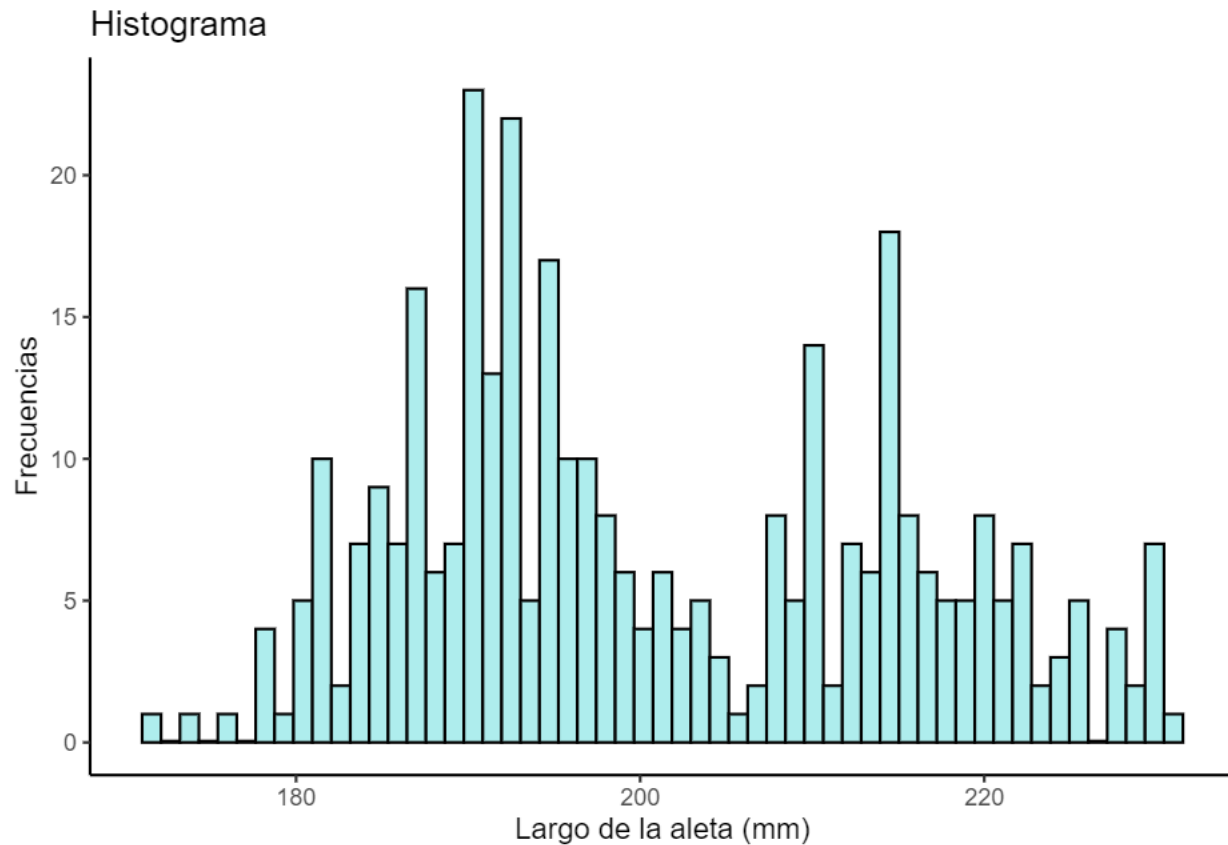


3.- Mejorar el gráfico usando el parámetro binwidth

```
HG1<-ggplot(penguins1, aes(x=largo_aleta_mm))+  
  geom_histogram(col="black", fill="paleturquoise", binwidth = 1.1)+  
  ggtitle("Histograma")+  
  xlab("Largo de la aleta (mm)")+  
  ylab("Frecuencias")+  
  theme_classic()
```

4.- Visualización del gráfico

HG1



h) *Gráficos de dispersión*

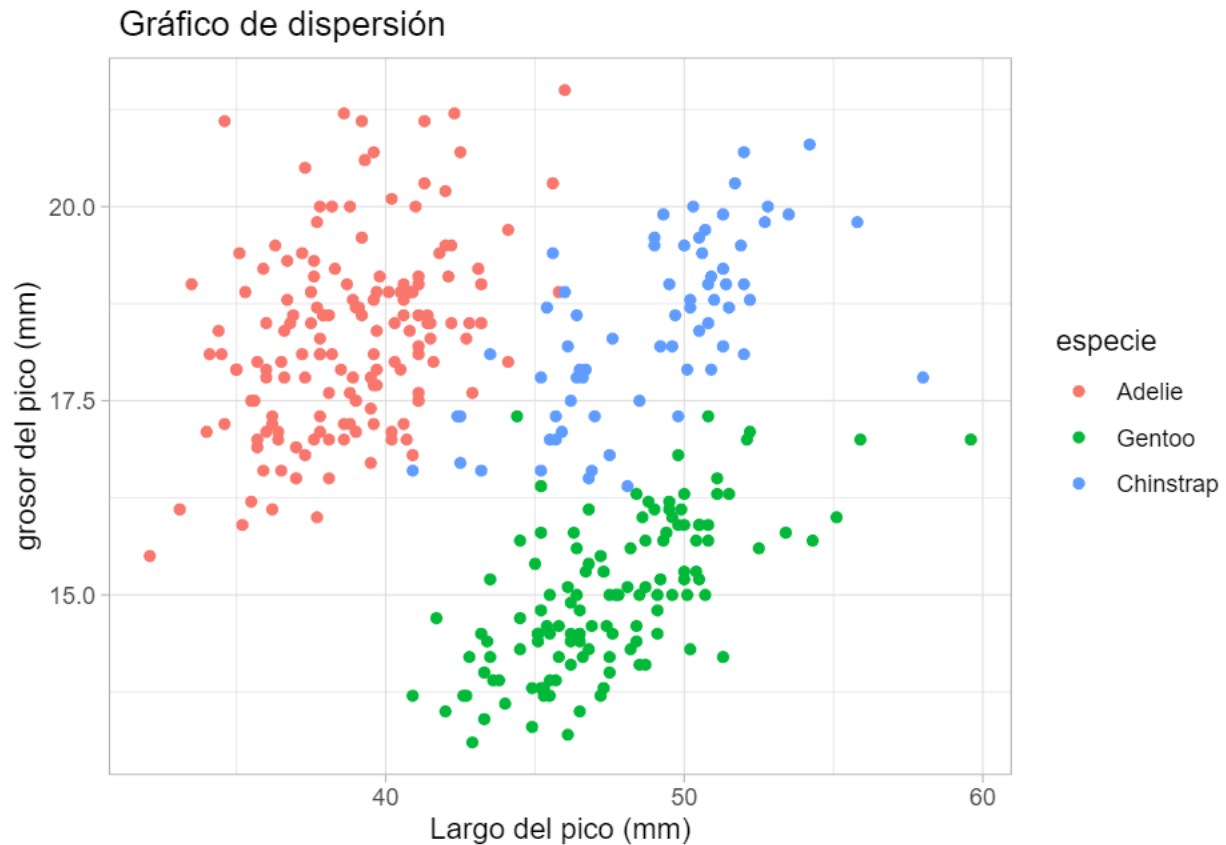
Para dos variables cualitativas y posiblemente dependientes

#1.- Construcción del gráfico

```
GD<-ggplot(penguins1, 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_light()
```

2.- Visualización de la variable

GD



i) Organización de gráficos

1.- Descargar el paquete del CRAN

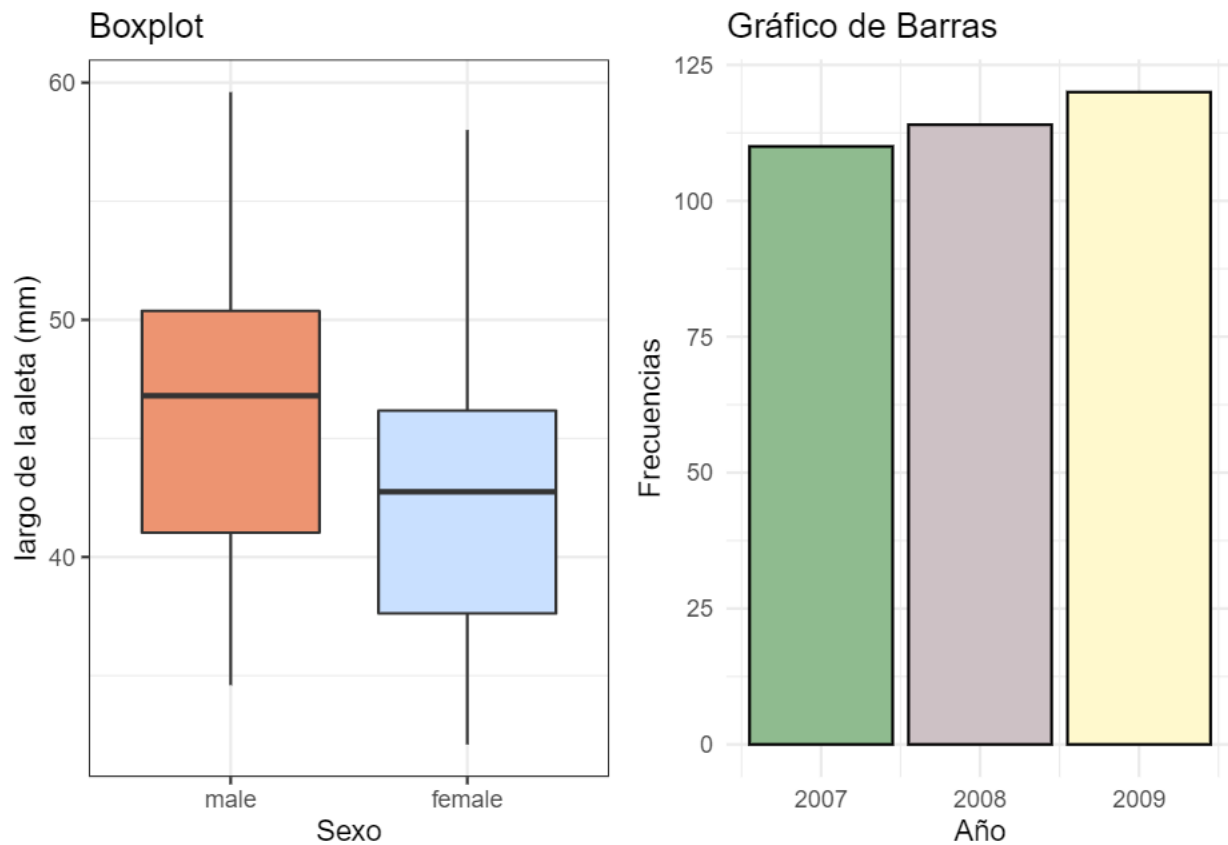
```
install.packages("gridExtra")
```

#2.- Abrir la librería

```
library(gridExtra)
```

3.- Organización de 2 gráficos en dos columnas y una fila

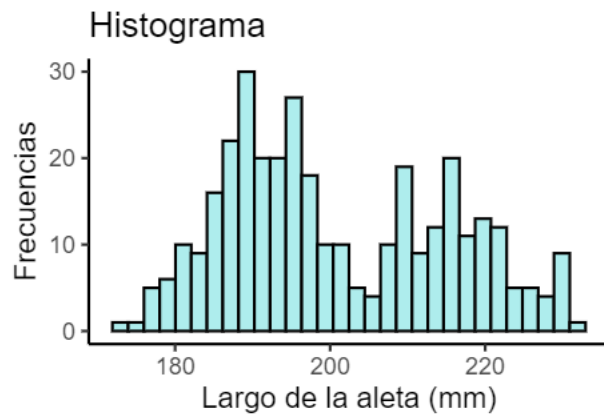
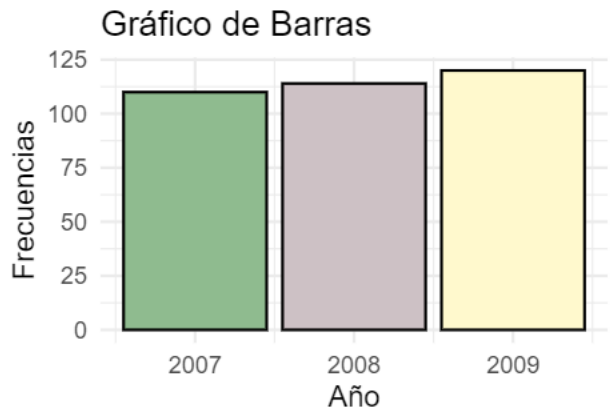
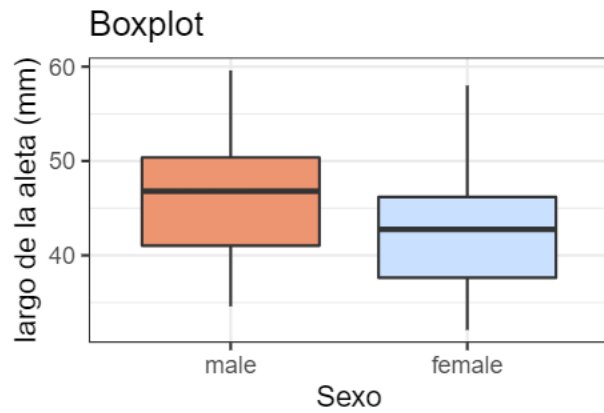
```
grid.arrange(BX, GB1, nrow=1, ncol=2)
```



4.- Organización de 3 gráficos en dos columnas y dos filas

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

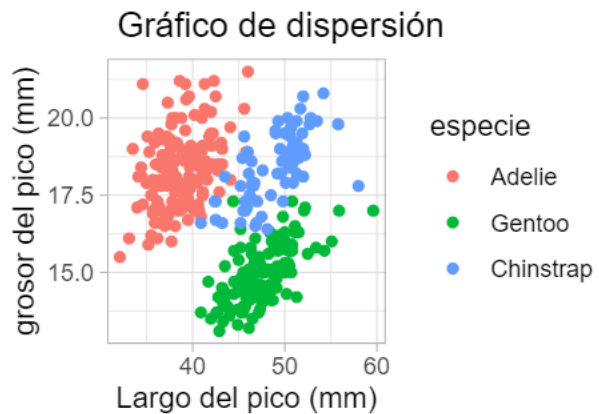
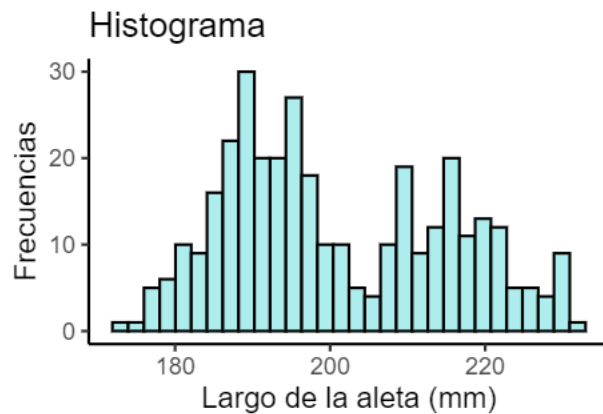
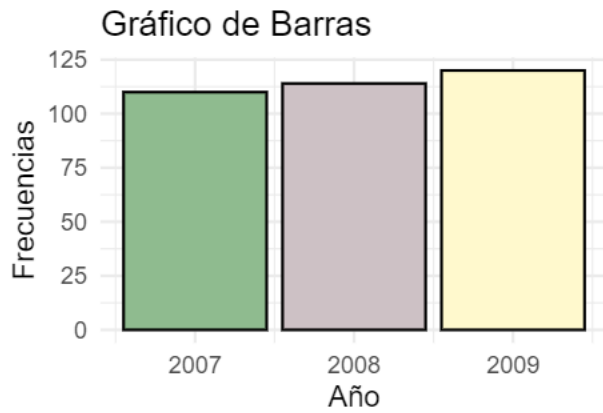
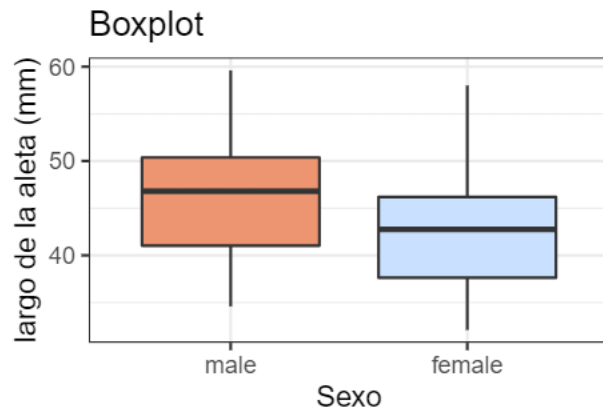
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



5.- Organización de 4 gráficos en dos columnas y dos filas

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

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

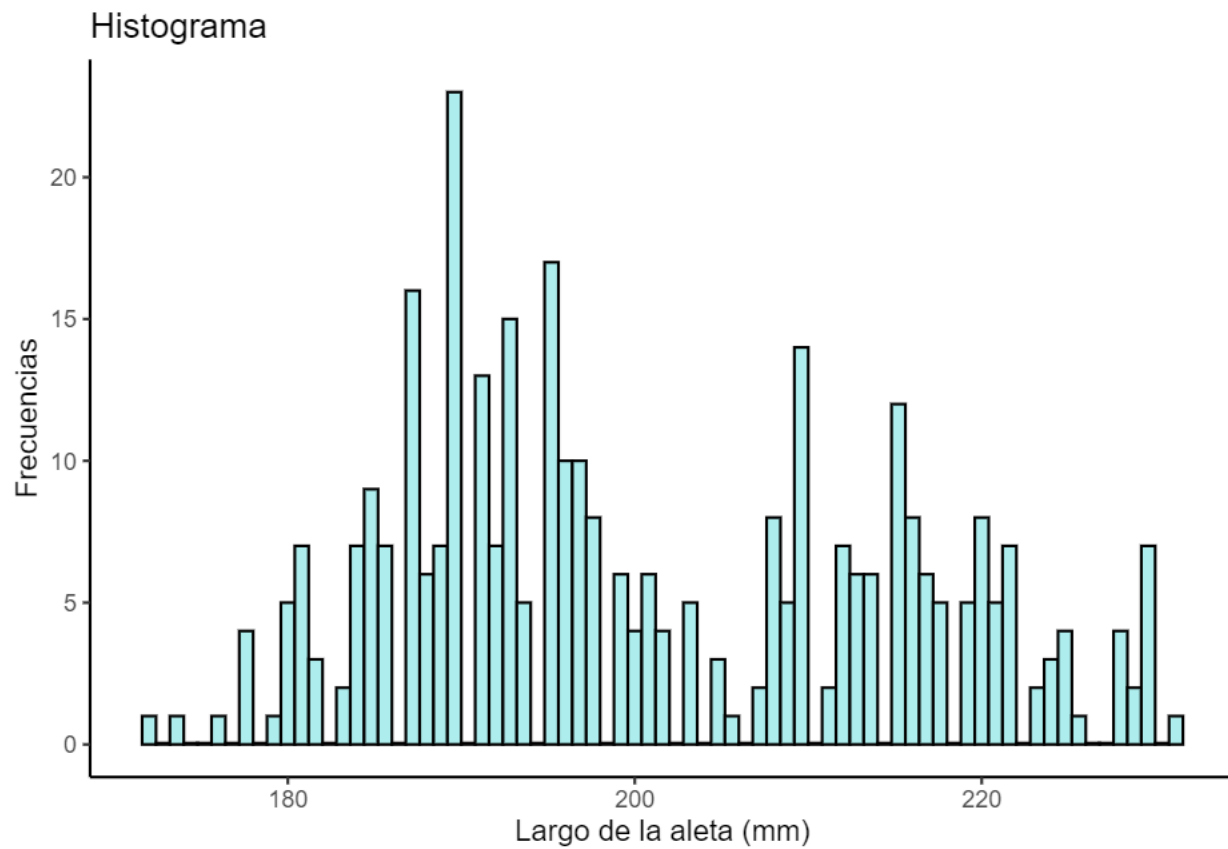


5.1 Mejorar el histograma para la organización de los gráficos

```
HG2<-ggplot(penguins1, aes(x=largo_aleta_mm))+
  geom_histogram(col="black", fill="paleturquoise", binwidth = 0.8)+
  ggtitle("Histograma")+
  xlab("Largo de la aleta (mm)")+
  ylab("Frecuencias")+
  theme_classic()
```

6.- Visualización del gráfico

HG2



7.- copiar y pegar el paso 5, agregando el histograma mejorado

```
grid.arrange(BX, GB1, HG2, GD, nrow=2, ncol=2)
```

