

# Penguins Project

Luis Choque

2023-08-10

```
Penguins = read.csv("C:/Users/Luis/Downloads/archive (1)/penguins.csv")
data(Penguins)
```

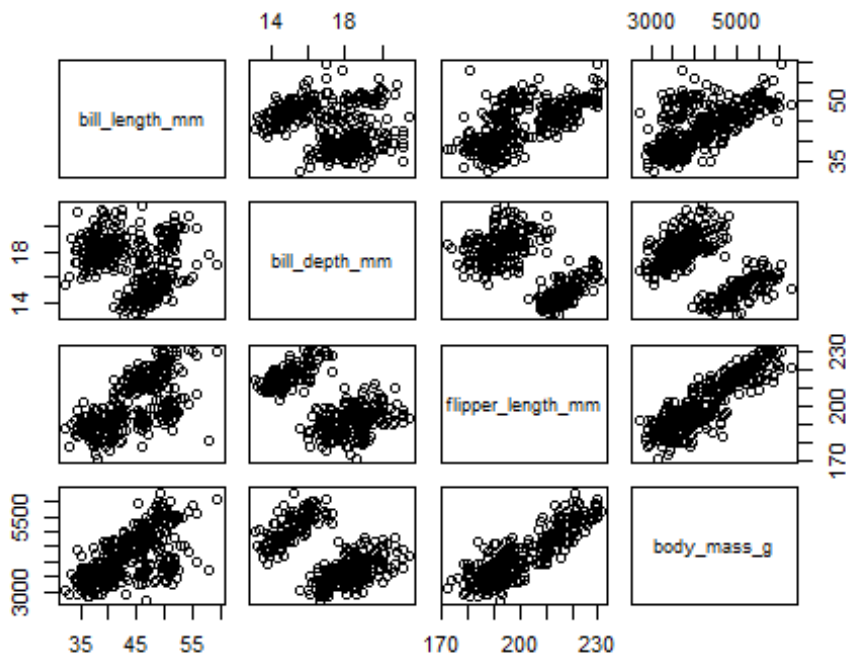
```
## Warning in data(Penguins): data set 'Penguins' not found
```

```
summary(Penguins[2:8])
```

```
##   species            island      bill_length_mm  bill_depth_mm
## Length:344      Length:344      Min.   :32.10    Min.   :13.10
## Class :character Class :character 1st Qu.:39.23    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
##                                     NA's   :2        NA's   :2
## flipper_length_mm  body_mass_g      sex
## Min.   :172.0      Min.   :2700    Length:344
## 1st Qu.:190.0      1st Qu.:3550    Class :character
## Median :197.0      Median :4050    Mode  :character
## Mean   :200.9      Mean   :4202
## 3rd Qu.:213.0      3rd Qu.:4750
## Max.   :231.0      Max.   :6300
## NA's   :2          NA's   :2
```

Realizamos la correlación entre las variables del dataset.

```
pairs(Penguins[, c("bill_length_mm", "bill_depth_mm",
"flipper_length_mm", "body_mass_g")])
```



En general, los datos se encuentran muy dispersos, parece tener una correlación positiva entre las variables. Para ello, veamos de manera numérica:

```
cor(Penguins[4:7], use = "complete.obs")
```

	bill_length_mm	bill_depth_mm	flipper_length_mm
body_mass_g			
## bill_length_mm	1.0000000	-0.2350529	0.6561813
0.5951098			
## bill_depth_mm	-0.2350529	1.0000000	-0.5838512
0.4719156			
## flipper_length_mm	0.6561813	-0.5838512	1.0000000
0.8712018			
## body_mass_g	0.5951098	-0.4719156	0.8712018
1.0000000			

En resumen, los resultados sugieren que hay relaciones significativas entre estas variables físicas de los pingüinos. La longitud y profundidad del pico, así como la longitud de la aleta, están relacionadas con la masa corporal de los pingüinos.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

library(dplyr)

##
## Attaching package: 'dplyr'
```

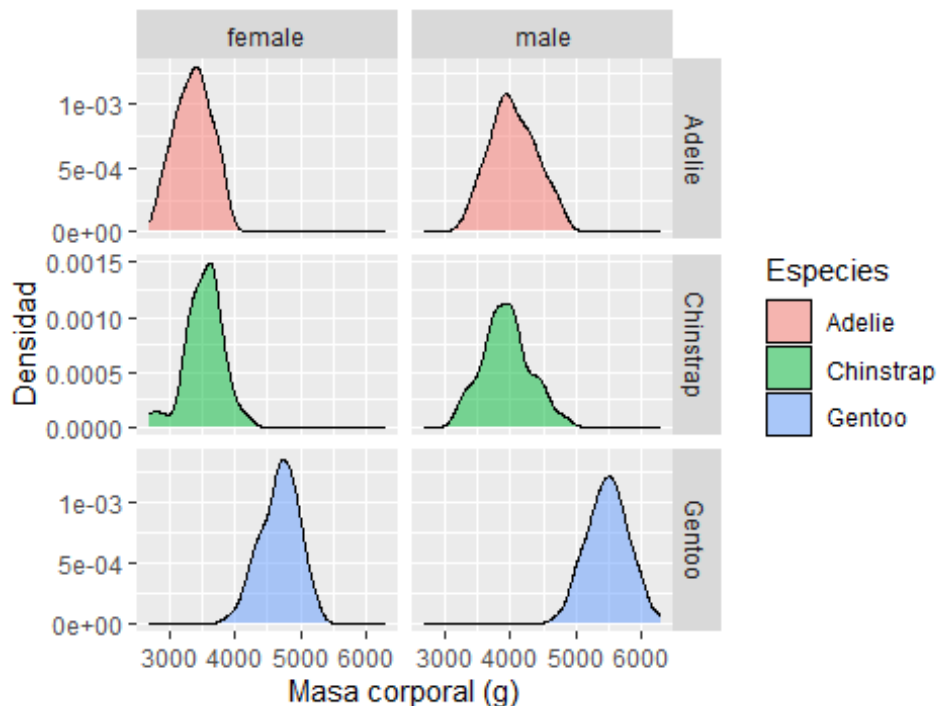
```
## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

filtered_penguins <- Penguins %>% filter(!is.na(sex))

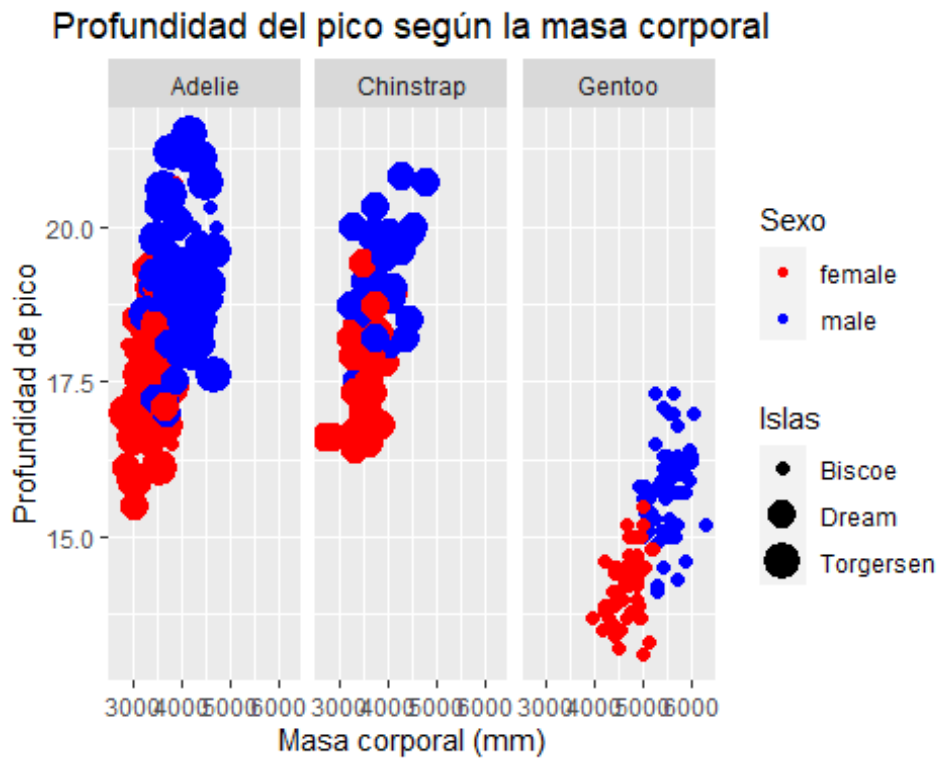
ggplot(filtered_penguins, aes(x = body_mass_g, fill = species)) +
  geom_density(alpha = 0.5) +
  facet_grid(cols = vars(sex), rows = vars(species), scales = "free_y") +
  labs(x = "Masa corporal (g)", y = "Densidad",
       title = "Densidad de Masa Corporal por Especie y Sexo", fill =
"Especies") +
  theme(plot.title = element_text(hjust = 0.5))
```

### Densidad de Masa Corporal por Especie y Sexo

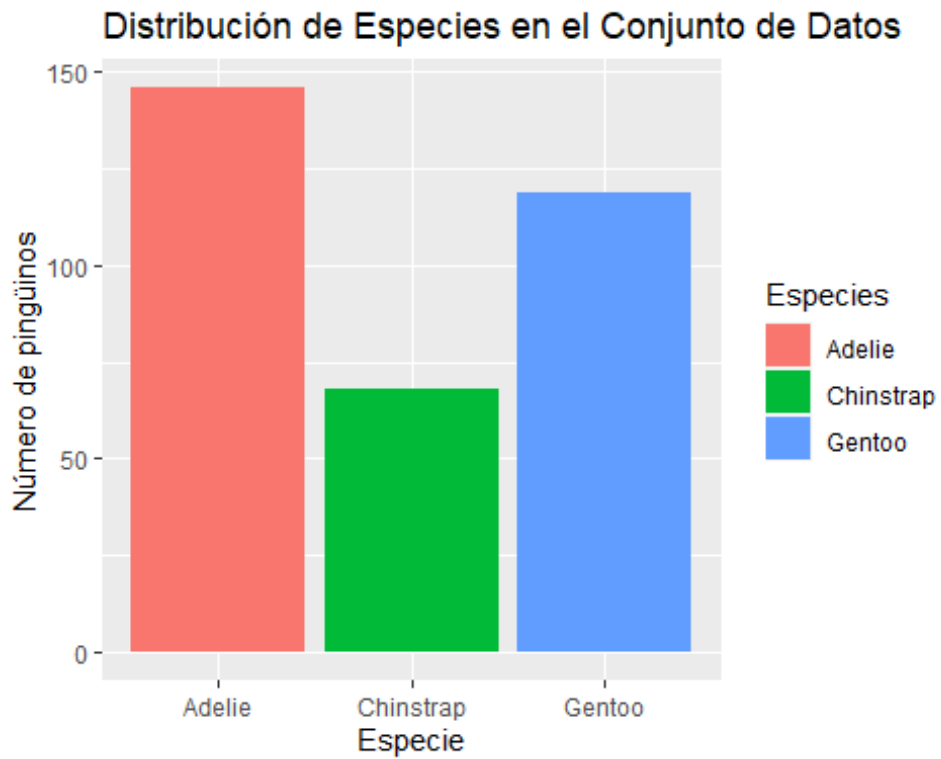


```
ggplot(filtered_penguins, aes(x = body_mass_g, y = bill_depth_mm,
                             size = island)) +
  geom_point(aes(color = sex)) +
  facet_wrap(~species) +
  labs(x = "Masa corporal (mm)", y = "Profundidad de pico",
       title = "Profundidad del pico según la masa corporal",
       size = "Islas") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_color_manual(name = "Sexo", values = c("male" = "blue", "female"
= "red"))
```

```
## Warning: Using size for a discrete variable is not advised.
```



```
ggplot(filtered_penguins, aes(x = species, fill = species)) +  
  geom_bar() +  
  labs(x = "Especie", y = "Número de pingüinos",  
        title = "Distribución de Especies en el Conjunto de Datos", fill =  
        "Especies")
```



```
ggplot(filtered_penguins, aes(x = species, y = body_mass_g, fill = sex))  
+  
  geom_boxplot() +  
  labs(x = "Especie", y = "Masa corporal (g)",  
        title = "Distribución de Masa Corporal por Especie y Sexo", fill =  
        "Sexo")
```

Distribución de Masa Corporal por Especie y Sexo

