

# Recodificación de variables

Etna Carolina Cortés Martínez

30/10/2021

## Importación de la matriz

1.- Lectura de la matriz

```
BD<-read.csv("BD1.csv")
```

2.- Seleccionar las columnas 3 a la 10

```
BD1<-BD[,3:10]
```

*El “[ ]” selecciona y los “:” significan a la o hasta.*

3.- Se obtiene una copia de la matriz

```
BD2<-BD1
```

## Renombrar variables

1.- Visualización del nombre de las variables.

```
colnames(BD1)
```

```
## [1] "species"      "island"        "bill_length_mm"  
## [4] "bill_depth_mm" "flipper_length_mm" "body_mass_g"  
## [7] "sex"          "year"
```

2.- Renombrar las variables en orden de las columnas.

```
names(BD2)=c("Especie", "Isla", "Largo_pico_mm", "Ancho_pico_mm", "Largo_aleta_mm",  
             "Masa_corporal_g", "Sexo", "Año")
```

*c: significa que se abre un vector*

**NOTA:** tenemos que renombrar de manera adecuada por que deben de tener los mismos nombre como viene el la base

3.- Se vuelven a visualizar los nombres de las variables.

```
colnames(BD2)
```

```
## [1] "Especies"      "Isla"           "Largo_pico_mm"  "Ancho_pico_mm"  
## [5] "Largo_aleta_mm" "Masa_corporal_g" "Sexo"           "Año"
```

## Tratamiento de las variables cualitativas

Decir a R que nos identifique factores

1.-Factor\_\_\_\_\_

a) Verificar el orden de la variable

```
str(BD2$Especies)
```

```
## chr [1:344] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" ...
```

```
BD2$Especies
```

```
## [1] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [7] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [13] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [19] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [25] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [31] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [37] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [43] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [49] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [55] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [61] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [67] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [73] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [79] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [85] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [91] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [97] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [103] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [109] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [115] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [121] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [127] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [133] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [139] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [145] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"  
## [151] "Adelie" "Adelie" "Gentoo" "Gentoo" "Gentoo" "Gentoo"  
## [157] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"  
## [163] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"  
## [169] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"  
## [175] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"  
## [181] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
```

```
## [187] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [193] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [199] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [205] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [211] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [217] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [223] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [229] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [235] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [241] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [247] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [253] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [259] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [265] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [271] "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"      "Gentoo"
## [277] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [283] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [289] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [295] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [301] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [307] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [313] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [319] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [325] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [331] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [337] "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"   "Chinstrap"
## [343] "Chinstrap"   "Chinstrap"
```

1) Variable Especie: originalmente está como caracter, pero se requiere convertido a un factor de 3 niveles.

```
BD2$Especies<-factor(BD2$Especies,
                      levels =c("Adelie", "Gentoo", "Chinstrap"))
```

**NOTA:** Voy a decirle a R que en la variable especies me la reconosca como un **factor** (tengo que abrir parentesis y decir que dentro de la matriz que los niveles van a ser el orden).

*Se verifica que la variable sea un factor.*

```
str(BD2$Especies)
```

```
## Factor w/ 3 levels "Adelie","Gentoo",...: 1 1 1 1 1 1 1 1 1 1 ...
```

2) variable Isla

```
str(BD2$Isla)
```

```
## chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
```

```
BD2$Isla
```

[illegible]

```
## [325] "Dream"      "Dream"      "Dream"      "Dream"      "Dream"      "Dream"
## [331] "Dream"      "Dream"      "Dream"      "Dream"      "Dream"      "Dream"
## [337] "Dream"      "Dream"      "Dream"      "Dream"      "Dream"      "Dream"
## [343] "Dream"      "Dream"
```

```
BD2$Isla<-factor(BD2$Isla,
                 levels =c("Torgersen", "Dream", "Biscoe"))
```

*visualización de facor con sus niveles*

```
str(BD2$Isla)
```

```
## Factor w/ 3 levels "Torgersen","Dream",...: 1 1 1 1 1 1 1 1 1 1 ...
```

### 3) Variable sexo

```
BD2$Sexo
```

```
## [1] "male" "female" "female" "female" "female" "male" "female" "male"
## [9] "female" "male" "female" "female" "female" "male" "male" "female"
## [17] "female" "male" "female" "male" "female" "male" "female" "male"
## [25] "male" "female" "male" "female" "female" "male" "female" "male"
## [33] "female" "male" "female" "male" "male" "female" "female" "male"
## [41] "female" "male" "female" "male" "female" "male" "male" "female"
## [49] "female" "male" "female" "male" "female" "male" "female" "male"
## [57] "female" "male" "female" "male" "female" "male" "female" "male"
## [65] "female" "male" "female" "male" "female" "male" "female" "male"
## [73] "female" "male" "female" "male" "female" "male" "female" "male"
## [81] "female" "male" "female" "male" "female" "male" "male" "female"
## [89] "male" "female" "female" "male" "female" "male" "female" "male"
## [97] "female" "male" "female" "male" "female" "male" "female" "male"
## [105] "female" "male" "female" "male" "female" "male" "female" "male"
## [113] "female" "male" "female" "male" "female" "male" "female" "male"
## [121] "female" "male" "female" "male" "female" "male" "female" "male"
## [129] "female" "male" "female" "male" "female" "male" "female" "male"
## [137] "female" "male" "female" "male" "female" "male" "female" "male"
## [145] "female" "male" "male" "female" "female" "male" "female" "male"
## [153] "female" "male" "female" "male" "male" "female" "female" "male"
## [161] "female" "male" "female" "male" "female" "male" "female" "male"
## [169] "female" "male" "female" "male" "male" "female" "female" "male"
## [177] "female" "male" "female" "male" "female" "male" "male" "female"
## [185] "female" "male" "female" "male" "female" "male" "female" "male"
## [193] "female" "male" "female" "male" "male" "female" "female" "male"
## [201] "female" "male" "female" "male" "female" "male" "female" "male"
## [209] "female" "male" "female" "male" "female" "male" "female" "male"
## [217] "female" "male" "female" "male" "female" "male" "female" "male"
## [225] "male" "female" "female" "male" "female" "male" "female" "male"
## [233] "female" "male" "female" "male" "female" "male" "female" "male"
## [241] "female" "male" "female" "male" "female" "male" "female" "male"
## [249] "male" "female" "female" "male" "female" "male" "female" "male"
## [257] "female" "male" "female" "male" "female" "male" "female" "male"
## [265] "female" "male" "female" "male" "female" "male" "female" "male"
```

```
## [273] "female" "male" "female" "male" "female" "male" "male" "female"
## [281] "male" "female" "female" "male" "female" "male" "female" "male"
## [289] "female" "male" "female" "male" "male" "female" "female" "male"
## [297] "female" "male" "female" "male" "female" "male" "female" "male"
## [305] "female" "male" "female" "male" "female" "male" "male" "female"
## [313] "female" "male" "female" "male" "male" "female" "male" "female"
## [321] "female" "male" "female" "male" "male" "female" "female" "male"
## [329] "female" "male" "female" "male" "female" "male" "male" "female"
## [337] "male" "female" "female" "male" "female" "male" "male" "female"
```

```
BD2$Sexo<-factor(BD2$Sexo,
                 levels =c("male","female"),
                 labels =c("Masculino","Femenino"))
```

**NOTA:** si cambio el orden de los datos me cambia el orden el factor.

*visualización de factor con sus niveles*

```
str(BD2$Sexo)
```

```
## Factor w/ 2 levels "Masculino","Femenino": 1 2 2 2 2 1 2 1 2 1 ...
```

#### 4) Variable cualitativa “año”

```
BD2$Año
```

```
## [1] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [16] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [31] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [46] 2007 2007 2007 2007 2007 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [61] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [76] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [91] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2009 2009 2009 2009 2009
## [106] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [121] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [136] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [151] 2009 2009 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [166] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [181] 2007 2007 2007 2007 2007 2007 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [196] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [211] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [226] 2008 2008 2008 2008 2008 2008 2008 2008 2009 2009 2009 2009 2009 2009 2009
## [241] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [256] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [271] 2009 2009 2009 2009 2009 2009 2009 2007 2007 2007 2007 2007 2007 2007 2007
## [286] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [301] 2007 2007 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [316] 2008 2008 2008 2008 2008 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [331] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
```

```
BD2$Año<-factor(BD2$Año,
                levels =c("2007","2008","2009"))
```

*visualización de factor con sus niveles*

```
str(BD2$Año)
```

```
## Factor w/ 3 levels "2007","2008",...: 1 1 1 1 1 1 1 1 1 1 ...
```

## Summary de toda la base de datos

```
summary(BD2)
```

```
##      Especies      Isla  Largo_pico_mm  Ancho_pico_mm
## Adelie   :152  Torgersen: 52   Min.    :32.10   Min.    :13.10
## Gentoo   :124  Dream    :124   1st Qu.:39.20   1st Qu.:15.60
## Chinstrap: 68  Biscoe   :168   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  Sexo      Año
## Min.    :172.0  Min.    :2700  Masculino:170  2007:110
## 1st Qu.:190.0  1st Qu.:3550  Femenino :174  2008:114
## Median :197.0  Median :4050              2009:120
## Mean    :200.9  Mean    :4202
## 3rd Qu.:213.2  3rd Qu.:4756
## Max.    :231.0  Max.    :6300
```

## Guardar matriz de datos final

```
BD3<-as.data.frame(BD2)
```

## Realización de gráfico “Bloxpot de variable sexo”

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

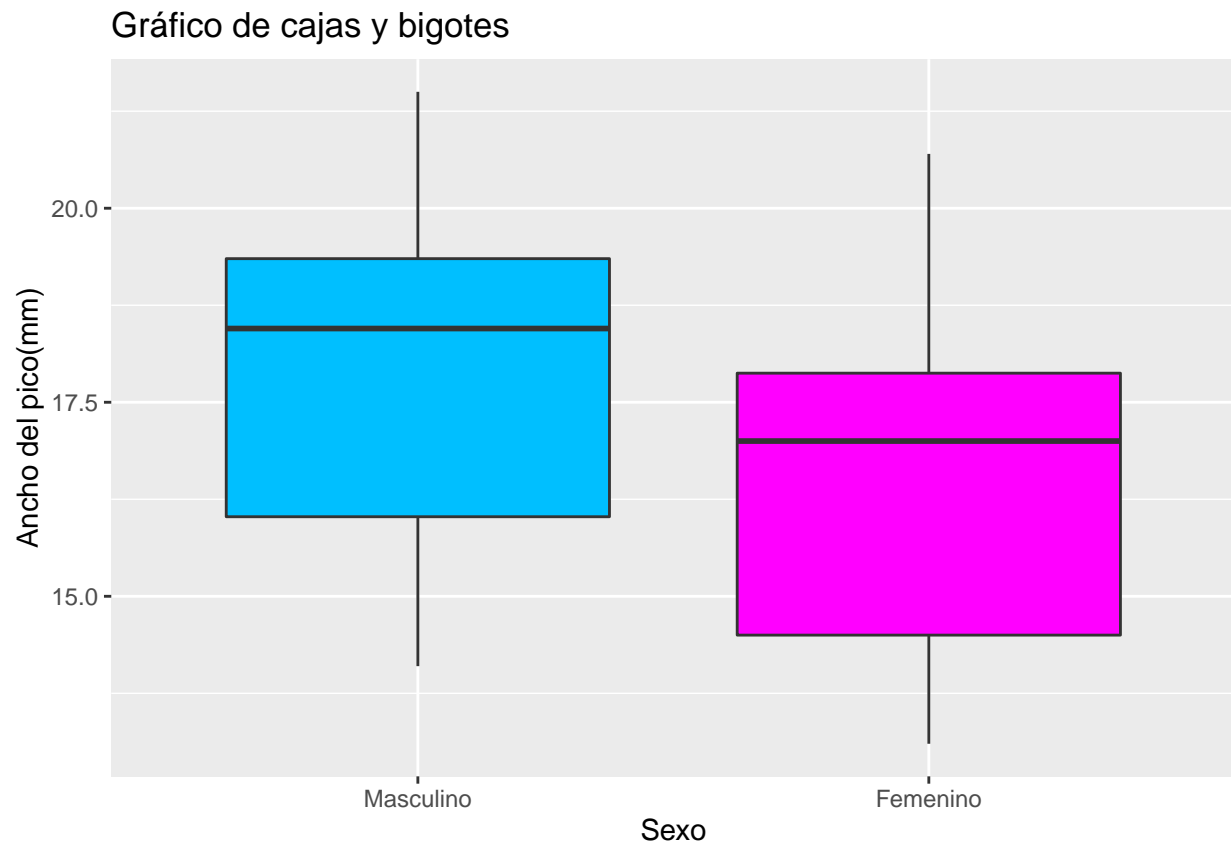
1.- Gráfico boxplot de exploración

```
color=c("deepskyblue1","magenta")
```

```
BX<-ggplot(BD3, aes(x=Sexo, y=Ancho_pico_mm))+
  geom_boxplot(fill=color)+
  ggtitle("Gráfico de cajas y bigotes")+
  xlab("Sexo")+
  ylab("Ancho del pico(mm)")
```

Visualización de gráfico.

BX



## Realización de gráfico de dispersión

```
DD<-ggplot(BD3, aes(Masa_corporal_g, Ancho_pico_mm))+  
  geom_point(aes(color=Isla))+  
  labs(title= "Gráfico de dispersión")+  
  xlab("peso")+  
  ylab("Largo de la aleta (mm)")+  
  theme_light()
```

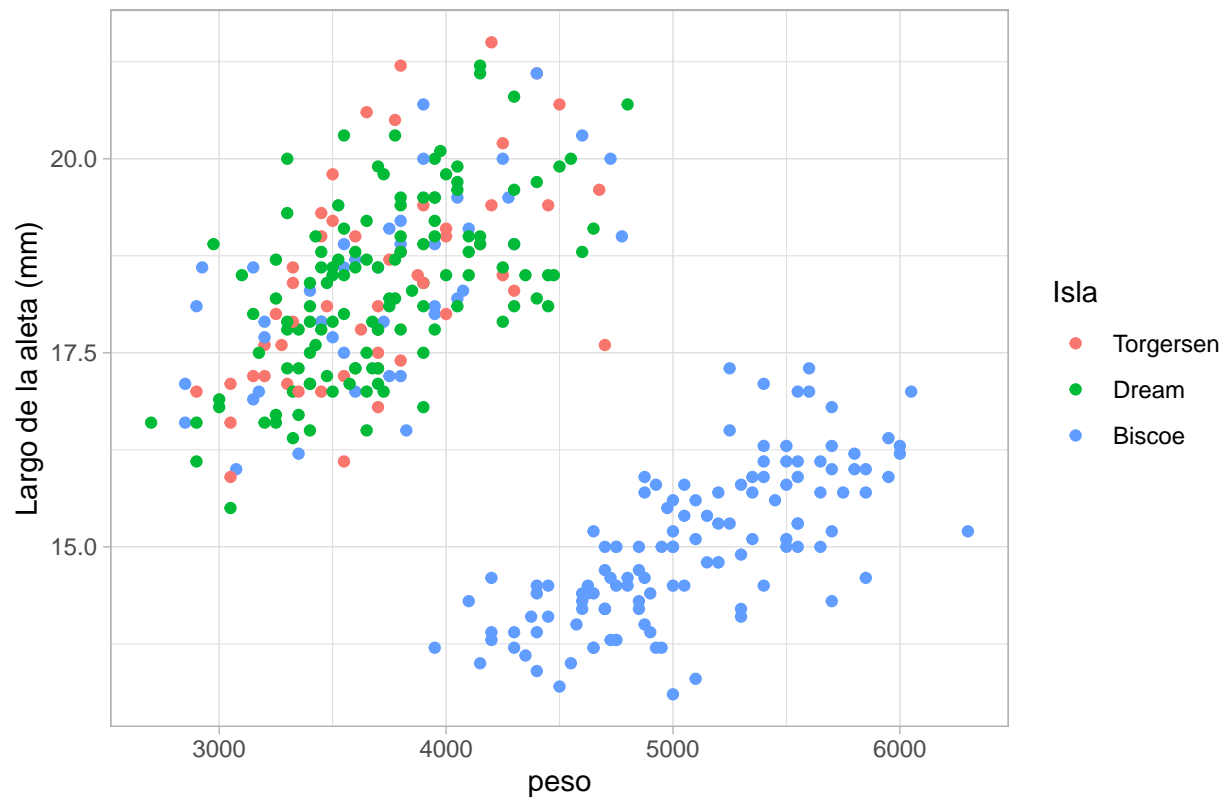
**NOTA:** ggplot reconoce que masa corporal es el eje “x” y el ancho del pico es el eje “y”

Visualización de gráfico.

DD



Gráfico de dispersión



## Guardar la matriz de datos

```
write.csv(BD3, "BD3_penguins.csv")
```