

"Graficos Básicos"

Etna Carolina Cortés Martínez

13/10/2021

Unidad 5. Analisis y visualización de datos

Gráficos Básicos

Dataset

```
install.packages("palmerpenguins")
```

```
library(palmerpenguins)
```

Paquetes de trabajo

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

```
library(ggplot2)
```

```
install.packages("RColorBrewer")
```

```
library(RColorBrewer)
```

Importación de la matriz

```
BD<-penguins
```

*si le pongo < le asigno un nuevo nombre a la base de datos

Exploración de la matriz

1.- Dimensión

```
dim(BD)
```

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

son 344 datos y 8 variables 2.- Nombre de las columnas.

```
colnames(BD)
```

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

3.- Clase a la que pertenece la matriz

```
class(BD)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

Es el formato en la que esta la base 4.- Estructura interna.

```
str(BD)
```

```
## tibble [344 x 8] (S3: tbl_df/tbl/data.frame)
## $ species      : Factor w/ 3 levels "Adelie","Chinstrap",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ island       : Factor w/ 3 levels "Biscoe","Dream",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ bill_length_mm : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
## $ bill_depth_mm : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
## $ flipper_length_mm: int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
## $ body_mass_g    : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
## $ sex           : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
## $ year          : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

Visualizar variables cuantitativas y cualitativas **Nota Integer (int)**: adquieren valores enteros.(Variables cuantitativas discretas) **numeric (num)**: adquieren valores y con decimales. (Variables cuantitativas continuas)

5.- Estadística descriptiva básica

```
summary(BD)
```

```
##      species      island  bill_length_mm  bill_depth_mm
## Adelie   :152  Biscoe   :168  Min.      :32.10  Min.      :13.10
## Chinstrap: 68  Dream    :124  1st Qu.:39.23  1st Qu.:15.60
## Gentoo   :124  Torgersen: 52  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      year
## Min.      :172.0    Min.      :2700  female:165  Min.      :2007
```

```
## 1st Qu.:190.0      1st Qu.:3550    male :168    1st Qu.:2007
## Median :197.0      Median :4050    NA's  : 11    Median :2008
## Mean   :200.9      Mean   :4202                Mean   :2008
## 3rd Qu.:213.0      3rd Qu.:4750                3rd Qu.:2009
## Max.   :231.0      Max.   :6300                Max.   :2009
## NA's   :2          NA's   :2
```

6.- Identificación de datos faltantes (NA)

```
anyNA(BD)
```

```
## [1] TRUE
```

7.- Tratamiento de **NA's** mediante el remplazo del dato por la media.

7.1 Trabajar sobre una nueva matriz de datos (BD1)

```
BD1<-BD
```

Asignación de nuevo nombre a la base 7.2.- Conocer la media aritmetica de bill_length

```
mean(BD$bill_length_mm, na.rm = TRUE)
```

```
## [1] 43.92193
```

Media 7.3.- Remplazar el valor perdido por la media.

```
BD1$bill_length_mm_medias<-ifelse(is.na(BD1$bill_length_mm), mean(BD1$bill_length_mm,na.rm = TRUE), BD1$bill_length_mm)
```

Todos los datos faltantes se van a remplazar por el valor de la media 7.4.-Visualización de la nueva columna. (bil_length_mn)

```
str(BD1)
```

```
## tibble [344 x 9] (S3: tbl_df/tbl/data.frame)
## $ species      : Factor w/ 3 levels "Adelie","Chinstrap",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ island       : Factor w/ 3 levels "Biscoe","Dream",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ bill_length_mm : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
## $ bill_depth_mm : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
## $ flipper_length_mm : int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
## $ body_mass_g    : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
## $ sex           : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
## $ year          : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
## $ bill_length_mm_medias: num [1:344] 39.1 39.5 40.3 43.9 36.7 ...
```

7.5 Repetir el paso 7.2, en las variables que tengan datos perdidos **media aritmetica de flipper_length_mm**

```
mean(BD$flipper_length_mm, na.rm = TRUE)
```

```
## [1] 200.9152
```

Remplazo de los datos perdidos de flipper_leng por la media aritmética

```
BD1$flipper_length_mm_medias<-ifelse(is.na(BD1$flipper_length_mm), mean(BD1$flipper_length_mm,na.rm = T
```

Media aritmética de bill_depth

```
mean(BD$bill_depth_mm, na.rm = TRUE)
```

```
## [1] 17.15117
```

Reemplazo de los datos perdidos de bill_depth por la media aritmética

```
BD1$bill_depth_mm_medias<-ifelse(is.na(BD1$bill_depth_mm), mean(BD1$bill_depth_mm,na.rm = TRUE), BD1$bi
```

Media aritmética de body_mass

```
mean(BD$body_mass_g, na.rm = TRUE)
```

```
## [1] 4201.754
```

Reemplazo de los datos perdidos de body_mass_g por la media artmética

```
BD1$body_mass_g_medias<-ifelse(is.na(BD1$body_mass_g), mean(BD1$body_mass_g,na.rm = TRUE), BD1$body_mas
```