

Análisis Exploratorio de Datos (EDA)

Elaborado por: Jaqueline Flores Salinas

11 mayo 2022

Remover objetos de la memoria

Después de trabajar durante un tiempo, el entorno de R puede acumular muchos datos en memoria, reteniendo partes significativas de los recursos del ordenador, por ello este punto es importante realizarlo al inicio de un nuevo trabajo.

```
knitr::opts_chunk$set(echo = TRUE)
rm(list=ls())
```

Instalación de paquetes

Aquí procederemos a instalar y cargar los paquetes necesarios para el análisis exploratorio de los datos.

```
#install.packages("rattle")
#install.packages("MASS")
#install.packages("readxl")
#install.packages("magrittr")
#install.packages("corrplot")
#install.packages("mice")
#install.packages("naniar")
#install.packages("tidyverse")
#install.packages("gridExtra")
#install.packages("dplyr")
#install.packages("tidyr")

library(rattle)
```

```
## Warning: package 'rattle' was built under R version 4.1.3
```

```
## Loading required package: tibble
```

```
## Loading required package: bitops
```

```
## Rattle: A free graphical interface for data science with R.
## Versión 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.
## Escriba 'rattle()' para agitar, sacudir y rotar sus datos.
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.1.3
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.1.3
```

```
library(magrittr)
```

```
## Warning: package 'magrittr' was built under R version 4.1.3
```

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 4.1.3
```

```
##
```

```
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
##      filter
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      cbind, rbind
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.1.3
```

```
## corrplot 0.92 loaded
```

```
library(naniar)
```

```
## Warning: package 'naniar' was built under R version 4.1.3
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.1.3
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v dplyr  1.0.7
## v tidyr  1.2.0      v stringr 1.4.0
## v readr  2.0.2      v forcats 0.5.1
## v purrr  0.3.4

## Warning: package 'tidyr' was built under R version 4.1.3

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::combine()   masks gridExtra::combine()
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks mice::filter(), stats::filter()
## x dplyr::lag()       masks stats::lag()
## x dplyr::select()    masks MASS::select()
## x purrr::set_names() masks magrittr::set_names()

#tinytex::install_tinytex()
library(dplyr)
library(tidyr)
```

Cargar nombres de columnas a analizar

Aquí haremos la importación de la base de datos más actualizada que tengamos, para cargar los nombres de columnas. Con base en las columnas de esta se editaron el resto de los documentos de excel, a mano, para asegurar de que todos tuvieran las mismas columnas.

```
nms <- names(readxl::read_excel("C:/Users/jaky_/Documents/GitHub/Tarea_DBT845/Datos/FDL2E1YC19_Padres.xl"))
nms
```

```
## [1] "CEPA"
## [3] "ARCHIVO"
## [5] "FECHA DESOVE MADRE"
## [7] "N° CRUZA"
## [9] "PIT PADRE"
## [11] "CÓDIGO DESOVE M"
## [13] "PESO PADRE"
## [15] "CATEGORÍA HEMBRA"
## [17] "EDAD HEMBRA"
## [19] "FOTOPERÍODO"
## [21] "ESTADO DE CRUZA"
## [23] "FECUNDIDAD TOTAL"
## [25] "FERTILIDAD"
## [27] "PESO OVAS"
## [29] "TAMAÑO OVA"
## [31] "OVAS MUERTAS"
## [33] "OVAS PRODUCCIÓN"
## [35] "FECHA SIEMBRA OVA"
## [37] "FINAL SF"
## [39] "FECHA DESPACHO"
## [41] "FAMILIA"

"LÍNEA"
"CLASE AÑO"
"FECHA DESOVE PADRE"
"PIT MADRE"
"CÓDIGO DESOVE H"
"PESO MADRE"
"PESO PROMEDIO PADRES"
"CATEGORÍA MACHO"
"EDAD MACHO"
"DAÑO OCULAR"
"CAUSA INVIABILIDAD"
"FECUNDIDAD RELATIVA"
"VOLUMEN OVAS"
"N° OVAS EN 30 CM"
"FECHA LIMPIEZA OVAS"
"OVAS PMG"
"PORCENTAJE SOBREVIVENCIA OVA OJO"
"MORTALIDAD A SF"
"PORCENTAJE MORTALIDAD ALEVINES SF"
"N° SF TRASLADADOS"
"OBSERVACIONES"
```



```
dim(FDL2E1YC13Padres);
```

```
## [1] 117 42
```

```
dim(FDL2E1YC16Padres);
```

```
## [1] 290 42
```

```
dim(FDL2E1YC19Padres)
```

```
## [1] 181 42
```

Editar la dimensiones de las bases importadas

Una vez confirmado que las dimensiones o columnas de nuestros archivos, son las mismas, aquí procedemos a fijar la dimensión de los archivos para que contengan 1 sola dimensión.

```
n10 <- dim(FDL2E1YC10Padres)[1]  
n13 <- dim(FDL2E1YC13Padres)[1]  
n16 <- dim(FDL2E1YC16Padres)[1]  
n19 <- dim(FDL2E1YC19Padres)[1]
```

Ordenar archivos según nombre de columnas

En este paso procederemos a organizar cada archivo para que las columnas se re acomoden en orden alfabético, de acuerdo con el nombre de cada columna, esto con la intención de facilitar la búsqueda de columnas dentro del archivo.

```
FDL2E1YC10Padres <-  
  FDL2E1YC10Padres[,sort(names(FDL2E1YC10Padres))]  
  
FDL2E1YC13Padres <-  
  FDL2E1YC13Padres[,sort(names(FDL2E1YC13Padres))]  
  
FDL2E1YC16Padres <-  
  FDL2E1YC16Padres[,sort(names(FDL2E1YC16Padres))]  
  
FDL2E1YC19Padres <-  
  FDL2E1YC19Padres[,sort(names(FDL2E1YC19Padres))]
```

Exploración de ajustes

En este paso procederemos a imprimir los datos de los archivos importados, una vez realizados los ajustes anteriores, para corroborar que los comando anteriores se ejecutaron exitosamente y se logró el objetivo deseado en los archivos a evaluar.

En este primer paso verificamos los nombres de las columnas de los 4 archivos, aquí podremos corroborar una vez mas, que los nombres son los mismos debido al comando del primer paso de este script, donde se realizó la importación de los nombres de las columnas de uno de los archivos.

## [8,]	"CÓDIGO DESOVE M"	"CÓDIGO DESOVE M"
## [9,]	"DAÑO OCULAR"	"DAÑO OCULAR"
## [10,]	"EDAD HEMBRA"	"EDAD HEMBRA"
## [11,]	"EDAD MACHO"	"EDAD MACHO"
## [12,]	"ESTADO DE CRUZA"	"ESTADO DE CRUZA"
## [13,]	"FAMILIA"	"FAMILIA"
## [14,]	"FECHA DESOVE MADRE"	"FECHA DESOVE MADRE"
## [15,]	"FECHA DESOVE PADRE"	"FECHA DESOVE PADRE"
## [16,]	"FECHA DESPACHO"	"FECHA DESPACHO"
## [17,]	"FECHA LIMPIEZA OVAS"	"FECHA LIMPIEZA OVAS"
## [18,]	"FECHA SIEMBRA OVA"	"FECHA SIEMBRA OVA"
## [19,]	"FECUNDIDAD RELATIVA"	"FECUNDIDAD RELATIVA"
## [20,]	"FECUNDIDAD TOTAL"	"FECUNDIDAD TOTAL"
## [21,]	"FERTILIDAD"	"FERTILIDAD"
## [22,]	"FINAL SF"	"FINAL SF"
## [23,]	"FOTOPERÍODO"	"FOTOPERÍODO"
## [24,]	"LÍNEA"	"LÍNEA"
## [25,]	"MORTALIDAD A SF"	"MORTALIDAD A SF"
## [26,]	"N° CRUZA"	"N° CRUZA"
## [27,]	"N° OVAS EN 30 CM"	"N° OVAS EN 30 CM"
## [28,]	"N° SF TRASLADADOS"	"N° SF TRASLADADOS"
## [29,]	"OBSERVACIONES"	"OBSERVACIONES"
## [30,]	"OVAS MUERTAS"	"OVAS MUERTAS"
## [31,]	"OVAS PMG"	"OVAS PMG"
## [32,]	"OVAS PRODUCCIÓN"	"OVAS PRODUCCIÓN"
## [33,]	"PESO MADRE"	"PESO MADRE"
## [34,]	"PESO OVAS"	"PESO OVAS"
## [35,]	"PESO PADRE"	"PESO PADRE"
## [36,]	"PESO PROMEDIO PADRES"	"PESO PROMEDIO PADRES"
## [37,]	"PIT MADRE"	"PIT MADRE"
## [38,]	"PIT PADRE"	"PIT PADRE"
## [39,]	"PORCENTAJE MORTALIDAD ALEVINES SF"	"PORCENTAJE MORTALIDAD ALEVINES SF"
## [40,]	"PORCENTAJE SOBREVIVENCIA OVA OJO"	"PORCENTAJE SOBREVIVENCIA OVA OJO"
## [41,]	"TAMAÑO OVA"	"TAMAÑO OVA"
## [42,]	"VOLUMEN OVAS"	"VOLUMEN OVAS"

En este segundo paso imprimimos un resumen de los datos contenidos en cada uno de los 4 archivos a analizar, aquí podemos corroborar el formato de cada columna o variable, así como el número de observaciones para cada archivo.

```
knitr::knit_print(cbind(str(FDL2E1YC10Padres), str(FDL2E1YC13Padres), str(FDL2E1YC16Padres), str(FDL2E1YC19Padres)))
```

```
## tibble [90 x 42] (S3: tbl_df/tbl/data.frame)
## $ ARCHIVO : chr [1:90] "E1" "E1" "E1" "E1" ...
## $ CATEGORÍA HEMBRA : chr [1:90] NA NA NA NA ...
## $ CATEGORÍA MACHO : chr [1:90] NA NA NA NA ...
## $ CAUSA INVIABILIDAD : chr [1:90] NA NA NA NA ...
## $ CEPA : chr [1:90] "Fanad" "Fanad" "Fanad" "Fanad" ...
## $ CLASE AÑO : num [1:90] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
## $ CÓDIGO DESOVE H : chr [1:90] "HFP-1002" "HFP-1004" "HFP-1005" "HFP-1006" ...
## $ CÓDIGO DESOVE M : chr [1:90] "MFP-1020" "MFP-1003" "MFP-1021" "MFP-1004" ...
## $ DAÑO OCULAR : chr [1:90] NA NA NA NA ...
## $ EDAD HEMBRA : chr [1:90] "2w" "2w" "2w" "2w" ...
```

```

## $ EDAD MACHO : chr [1:90] "2w" "2w" "2w" "2w" ...
## $ ESTADO DE CRUZA : chr [1:90] NA NA NA NA ...
## $ FAMILIA : chr [1:90] "N01" "N02" "N03" "N06" ...
## $ FECHA DESOVE MADRE : POSIXct[1:90], format: "2010-05-05" "2010-05-05" ...
## $ FECHA DESOVE PADRE : POSIXct[1:90], format: "2010-05-05" "2010-05-05" ...
## $ FECHA DESPACHO : POSIXct[1:90], format: NA NA ...
## $ FECHA LIMPIEZA OVAS : POSIXct[1:90], format: NA NA ...
## $ FECHA SIEMBRA OVA : POSIXct[1:90], format: NA NA ...
## $ FECUNDIDAD RELATIVA : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ FECUNDIDAD TOTAL : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ FERTILIDAD : num [1:90] 33 100 95 100 95 100 100 94 100 95.2 ...
## $ FINAL SF : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ FOTOPERÍODO : chr [1:90] NA NA NA NA ...
## $ LÍNEA : chr [1:90] "L2" "L2" "L2" "L2" ...
## $ MORTALIDAD A SF : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ N° CRUZA : num [1:90] 1 2 3 4 5 6 7 8 9 10 ...
## $ N° OVAS EN 30 CM : num [1:90] 64 62 62 66 71 66 70 66 62 66 ...
## $ N° SF TRASLADADOS : num [1:90] 400 400 400 400 400 400 400 400 400 400 ...
## $ OBSERVACIONES : chr [1:90] NA NA NA NA ...
## $ OVAS MUERTAS : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ OVAS PMG : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ OVAS PRODUCCIÓN : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ PESO MADRE : num [1:90] 2000 2000 2000 1900 2000 2000 2100 2400 2200 2600 .
## $ PESO OVAS : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ PESO PADRE : num [1:90] 3520 2800 3640 5860 2940 3440 3200 3660 2800 4040 .
## $ PESO PROMEDIO PADRES : num [1:90] 2760 2400 2820 3880 2470 2720 2650 3030 2500 3320 .
## $ PIT MADRE : chr [1:90] "HFP-1002" "HFP-1004" "HFP-1005" "HFP-1006" ...
## $ PIT PADRE : chr [1:90] "MFP-1020" "MFP-1003" "MFP-1021" "MFP-1004" ...
## $ PORCENTAJE MORTALIDAD ALEVINES SF: num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ PORCENTAJE SOBREVIVENCIA OVA OJO : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## $ TAMAÑO OVA : num [1:90] 4.69 4.84 4.84 4.55 4.23 ...
## $ VOLUMEN OVAS : num [1:90] NA NA NA NA NA NA NA NA NA NA ...
## tibble [117 x 42] (S3: tbl_df/tbl/data.frame)
## $ ARCHIVO : chr [1:117] "E1" "E1" "E1" "E1" ...
## $ CATEGORÍA HEMBRA : chr [1:117] "Selección" "Selección" "Selección" "Selección" ..
## $ CATEGORÍA MACHO : chr [1:117] "Selección" "Selección" "Selección" "Selección" ..
## $ CAUSA INVIABILIDAD : chr [1:117] NA NA NA NA ...
## $ CEPA : chr [1:117] "Fanad" "Fanad" "Fanad" "Fanad" ...
## $ CLASE AÑO : num [1:117] 2013 2013 2013 2013 2013 ...
## $ CÓDIGO DESOVE H : chr [1:117] "HFN13-5051" "HFN13-5052" "HFN13-5053" "HFN13-5054
## $ CÓDIGO DESOVE M : chr [1:117] "MFN13-5011" "MFN13-5017" "MFN13-5016" "MFN13-5015
## $ DAÑO OCULAR : chr [1:117] NA NA NA NA ...
## $ EDAD HEMBRA : chr [1:117] "2w" "2w" "2w" "2w" ...
## $ EDAD MACHO : chr [1:117] "2w" "2w" "2w" "2w" ...
## $ ESTADO DE CRUZA : chr [1:117] "Viable" "Viable" "Viable" "Viable" ...
## $ FAMILIA : chr [1:117] "P01" "P02" "P03" "P04" ...
## $ FECHA DESOVE MADRE : POSIXct[1:117], format: "2013-05-15" "2013-05-15" ...
## $ FECHA DESOVE PADRE : POSIXct[1:117], format: "2013-05-15" "2013-05-15" ...
## $ FECHA DESPACHO : POSIXct[1:117], format: "2013-09-25" "2013-09-25" ...
## $ FECHA LIMPIEZA OVAS : POSIXct[1:117], format: "2013-06-17" "2013-06-17" ...
## $ FECHA SIEMBRA OVA : POSIXct[1:117], format: "2013-08-11" "2013-08-11" ...
## $ FECUNDIDAD RELATIVA : num [1:117] 1844 2089 2481 2291 2237 ...
## $ FECUNDIDAD TOTAL : num [1:117] 4411 5882 5095 4753 5882 ...
## $ FERTILIDAD : num [1:117] 91.3 90.2 91.8 100 100 ...

```



```

## $ FINAL SF : num [1:117] 440 451 457 464 470 421 452 473 449 430 ...
## $ FOTOPERÍODO : chr [1:117] "No" "No" "No" "No" ...
## $ LÍNEA : chr [1:117] "L2" "L2" "L2" "L2" ...
## $ MORTALIDAD A SF : num [1:117] 1 0 0 11 1 0 2 1 2 8 ...
## $ N° CRUZA : num [1:117] 1 2 3 4 5 6 7 8 9 10 ...
## $ N° OVAS EN 30 CM : num [1:117] 65 61 63 64 64 62 61 59 58 58 ...
## $ N° SF TRASLADADOS : num [1:117] 260 260 260 260 260 260 260 260 260 260 ...
## $ OBSERVACIONES : chr [1:117] NA NA NA NA ...
## $ OVAS MUERTAS : num [1:117] 465 786 543 249 507 ...
## $ OVAS PMG : num [1:117] 441 451 457 475 471 421 454 474 451 438 ...
## $ OVAS PRODUCCIÓN : num [1:117] 4129 6540 6561 5611 7087 ...
## $ PESO MADRE : num [1:117] 2580 3440 2980 2780 3440 3880 3230 3420 2840 3620
## $ PESO OVAS : num [1:117] 366 678 616 522 652 927 727 847 748 988 ...
## $ PESO PADRE : num [1:117] 3960 3840 3860 3960 3960 3860 4160 4160 3960 3720
## $ PESO PROMEDIO PADRES : num [1:117] 3270 3640 3420 3370 3700 ...
## $ PIT MADRE : chr [1:117] "0006EE247B" "0006F03D3E" "00066E39E3" "00069A9069
## $ PIT PADRE : chr [1:117] "0006F37A46" "0006F3F9B9" "0006F3CF3C" "0006F06DCB
## $ PORCENTAJE MORTALIDAD ALEVINES SF: num [1:117] 0.227 0 0 2.316 0.212 ...
## $ PORCENTAJE SOBREVIVENCIA OVA OJO : num [1:117] 89.5 86.6 89.3 94.8 91.4 ...
## $ TAMAÑO OVA : num [1:117] 4.62 4.92 4.76 4.69 4.69 ...
## $ VOLUMEN OVAS : num [1:117] 1844 2089 2481 2291 2237 ...
## tibble [290 x 42] (S3: tbl_df/tbl/data.frame)
## $ ARCHIVO : chr [1:290] "E1" "E1" "E1" "E1" ...
## $ CATEGORÍA HEMBRA : chr [1:290] "Selección" "Selección" "Selección" "Selección" ..
## $ CATEGORÍA MACHO : chr [1:290] "Selección" "Selección" "Selección" "Selección" ..
## $ CAUSA INVIABILIDAD : chr [1:290] NA NA NA NA ...
## $ CEPA : chr [1:290] "Fanad" "Fanad" "Fanad" "Fanad" ...
## $ CLASE AÑO : num [1:290] 2016 2016 2016 2016 2016 ...
## $ CÓDIGO DESOVE H : chr [1:290] "HFF16-632" "HFF16-633" "HFF16-637" "HFF16-635" ..
## $ CÓDIGO DESOVE M : chr [1:290] "MFF16-019" "MFF16-019" "MFF16-019" "MFF16-026" ..
## $ DAÑO OCULAR : chr [1:290] NA NA NA NA ...
## $ EDAD HEMBRA : chr [1:290] "3w" "3w" "3w" "3w" ...
## $ EDAD MACHO : chr [1:290] "3w" "3w" "3w" "3w" ...
## $ ESTADO DE CRUZA : chr [1:290] "viable" "viable" "viable" "viable" ...
## $ FAMILIA : chr [1:290] "F01" "F02" NA "F04" ...
## $ FECHA DESOVE MADRE : POSIXct[1:290], format: "2016-04-27" "2016-04-27" ...
## $ FECHA DESOVE PADRE : POSIXct[1:290], format: "2016-04-27" "2016-04-27" ...
## $ FECHA DESPACHO : POSIXct[1:290], format: "2017-10-06" "2017-10-06" ...
## $ FECHA LIMPIEZA OVAS : POSIXct[1:290], format: "2016-06-17" "2016-06-17" ...
## $ FECHA SIEMBRA OVA : POSIXct[1:290], format: "2016-07-22" "2016-07-22" ...
## $ FECUNDIDAD RELATIVA : num [1:290] 1206 1896 1540 1432 1614 ...
## $ FECUNDIDAD TOTAL : num [1:290] 18020 20284 11856 10397 13946 ...
## $ FERTILIDAD : num [1:290] 86 86 93 NA 0.72 NA 96 93 84 92 ...
## $ FINAL SF : num [1:290] 493 477 477 498 NA NA 494 482 485 487 ...
## $ FOTOPERÍODO : chr [1:290] "No" "No" "No" "No" ...
## $ LÍNEA : chr [1:290] "L2" "L2" "L2" "L2" ...
## $ MORTALIDAD A SF : num [1:290] 7 23 23 2 NA NA 6 18 15 13 ...
## $ N° CRUZA : num [1:290] 1 2 3 4 5 6 7 8 9 10 ...
## $ N° OVAS EN 30 CM : num [1:290] 55 59 56 55 56 58 60 58 55 56 ...
## $ N° SF TRASLADADOS : num [1:290] 250 250 250 250 NA NA 250 250 250 250 ...
## $ OBSERVACIONES : chr [1:290] "TITULAR" "TITULAR" "NO" "TITULAR" ...
## $ OVAS MUERTAS : num [1:290] 3347 11204 2187 1914 NA ...
## $ OVAS PMG : num [1:290] 500 500 500 500 NA NA 500 500 500 500 ...
## $ OVAS PRODUCCIÓN : num [1:290] 17520 19784 11356 9897 NA ...

```

```

## $ PESO MADRE : num [1:290] 14940 10700 7700 7260 8640 ...
## $ PESO OVAS : num [1:290] 2650 2305 1520 1405 1835 ...
## $ PESO PADRE : num [1:290] 15000 15000 15000 14280 14280 ...
## $ PESO PROMEDIO PADRES : num [1:290] 14970 12850 11350 10770 11460 ...
## $ PIT MADRE : chr [1:290] "0006D69D79" "0006D6AA44" "0006D6B37C" "0006952BD8
## $ PIT PADRE : chr [1:290] "0006D6AE53" "0006D6AE53" "0006D6AE53" "0006D6C62C
## $ PORCENTAJE MORTALIDAD ALEVINES SF: num [1:290] 1.4 4.6 4.6 0.4 NA NA 1.2 3.6 3 2.6 ...
## $ PORCENTAJE SOBREVIVENCIA OVA OJO : num [1:290] 81.1 45.4 82.2 81.6 NA ...
## $ TAMAÑO OVA : num [1:290] 5.45 5.08 5.36 5.45 5.36 ...
## $ VOLUMEN OVAS : num [1:290] NA NA NA NA NA NA NA NA NA NA ...
## tibble [181 x 42] (S3: tbl_df/tbl/data.frame)
## $ ARCHIVO : chr [1:181] "E1" "E1" "E1" "E1" ...
## $ CATEGORÍA HEMBRA : chr [1:181] "Selección" "Selección" "Selección" "Selección" ..
## $ CATEGORÍA MACHO : chr [1:181] "Selección" "Selección" "Selección" "Selección" ..
## $ CAUSA INVIABILIDAD : chr [1:181] NA NA NA NA ...
## $ CEPA : chr [1:181] "Fanad" "Fanad" "Fanad" "Fanad" ...
## $ CLASE AÑO : num [1:181] 2019 2019 2019 2019 2019 ...
## $ CÓDIGO DESOVE H : chr [1:181] "HFN19-46" "HFN19-47" "HFN19-48" "HFN19-49" ...
## $ CÓDIGO DESOVE M : chr [1:181] "MFN17-11" "MFN17-11" "MFN17-11" "MFN17-12" ...
## $ DAÑO OCULAR : chr [1:181] NA NA NA NA ...
## $ EDAD HEMBRA : chr [1:181] "2w" "2w" "2w" "2w" ...
## $ EDAD MACHO : chr [1:181] "2w" "2w" "2w" "2w" ...
## $ ESTADO DE CRUZA : chr [1:181] "Viable" "Viable" "Viable" "Viable" ...
## $ FAMILIA : chr [1:181] "F01" "F02" "F03" "F04" ...
## $ FECHA DESOVE MADRE : POSIXct[1:181], format: "2019-04-24" "2019-04-24" ...
## $ FECHA DESOVE PADRE : POSIXct[1:181], format: "2017-04-11" "2017-04-11" ...
## $ FECHA DESPACHO : POSIXct[1:181], format: "2019-08-19" "2019-08-19" ...
## $ FECHA LIMPIEZA OVAS : POSIXct[1:181], format: "2019-05-27" "2019-05-27" ...
## $ FECHA SIEMBRA OVA : POSIXct[1:181], format: "2019-07-10" "2019-07-10" ...
## $ FECUNDIDAD RELATIVA : num [1:181] 2328 2288 2153 2523 2215 ...
## $ FECUNDIDAD TOTAL : num [1:181] 10283 8878 7106 7821 11650 ...
## $ FERTILIDAD : num [1:181] 93.1 90.9 96.4 98 86.2 ...
## $ FINAL SF : num [1:181] 373 387 356 207 261 162 154 146 446 362 ...
## $ FOTOPERÍODO : chr [1:181] "No" "No" "No" "No" ...
## $ LÍNEA : chr [1:181] "L2" "L2" "L2" "L2" ...
## $ MORTALIDAD A SF : num [1:181] 4 6 0 2 4 2 7 5 10 8 ...
## $ N° CRUZA : num [1:181] 1 2 3 4 5 6 7 9 10 11 ...
## $ N° OVAS EN 30 CM : num [1:181] 57 54 55 55 50 52 57 53 58 53 ...
## $ N° SF TRASLADADOS : num [1:181] 130 130 130 130 130 130 130 130 130 130 ...
## $ OBSERVACIONES : chr [1:181] "Causa de mort. Post-eclosión es de normal y doble
## $ OVAS MUERTAS : num [1:181] 441 479 726 354 1760 ...
## $ OVAS PMG : num [1:181] 377 393 356 209 265 164 161 151 456 370 ...
## $ OVAS PRODUCCIÓN : num [1:181] 9507 8102 6052 6778 8446 ...
## $ PESO MADRE : num [1:181] 4417 3880 3300 3100 5260 ...
## $ PESO OVAS : num [1:181] 1055 972 772 823 1397 ...
## $ PESO PADRE : num [1:181] 8200 8200 8200 4620 4620 4620 4380 4380 4040 4040
## $ PESO PROMEDIO PADRES : num [1:181] 6308 6040 5750 3860 4940 ...
## $ PIT MADRE : chr [1:181] "0006D6B049" "00069AA783" "00069AA306" "00069A9EF3
## $ PIT PADRE : chr [1:181] "000695330F" "000695330F" "000695330F" "00066E5096
## $ PORCENTAJE MORTALIDAD ALEVINES SF: num [1:181] 1.061 1.527 0 0.957 1.509 ...
## $ PORCENTAJE SOBREVIVENCIA OVA OJO : num [1:181] 95.7 94.7 89.8 95.2 83.2 ...
## $ TAMAÑO OVA : num [1:181] 4.62 4.92 4.76 4.84 5.08 ...
## $ VOLUMEN OVAS : num [1:181] NA NA NA NA NA NA NA NA NA NA ...
## NULL

```

Matriz para el Exploratorio de datos

Una vez confirmando que los nombres y dimensiones de los archivos coinciden, en este paso procederemos a unir los 4 archivos y crear uno solo, a partir del cual se podrá realizar el análisis exploratorio de los datos.

```
FDL2E1YC <- rbind(FDL2E1YC10Padres,FDL2E1YC13Padres,FDL2E1YC16Padres,FDL2E1YC19Padres)
summary(FDL2E1YC)
```

```
##      ARCHIVO          CATEGORÍA HEMBRA  CATEGORÍA MACHO  CAUSA INVIABILIDAD
## Length:678          Length:678          Length:678          Length:678
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##      CEPA            CLASE AÑO          CÓDIGO DESOVE H  CÓDIGO DESOVE M
## Length:678          Min.   :2010          Length:678          Length:678
## Class :character    1st Qu.:2013          Class :character    Class :character
## Mode  :character    Median :2016          Mode  :character    Mode  :character
##                      Mean   :2015
##                      3rd Qu.:2019
##                      Max.   :2019
##
##      DAÑO OCULAR      EDAD HEMBRA          EDAD MACHO          ESTADO DE CRUZA
## Length:678          Length:678          Length:678          Length:678
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##      FAMILIA          FECHA DESOVE MADRE          FECHA DESOVE PADRE
## Length:678          Min.   :2010-05-05 00:00:00  Min.   :2010-05-05 00:00:00
## Class :character    1st Qu.:2013-06-13 18:00:00  1st Qu.:2013-06-13 18:00:00
## Mode  :character    Median :2016-05-18 00:00:00  Median :2016-05-18 00:00:00
##                      Mean   :2015-11-17 02:32:55  Mean   :2015-04-30 14:30:47
##                      3rd Qu.:2019-05-02 00:00:00  3rd Qu.:2017-04-11 00:00:00
##                      Max.   :2019-05-28 00:00:00  Max.   :2017-05-02 00:00:00
##
##
##      FECHA DESPACHO          FECHA LIMPIEZA OVAS
## Min.   :2013-09-25 00:00:00  Min.   :2013-06-17 00:00:00
## 1st Qu.:2017-10-06 00:00:00  1st Qu.:2016-06-17 00:00:00
## Median :2017-10-06 00:00:00  Median :2016-07-13 00:00:00
## Mean   :2017-06-08 04:27:46  Mean   :2016-10-09 08:25:10
## 3rd Qu.:2019-08-19 00:00:00  3rd Qu.:2019-06-14 00:00:00
## Max.   :2019-08-19 00:00:00  Max.   :2019-06-27 00:00:00
## NA's   :194                  NA's   :125
##      FECHA SIEMBRA OVA          FECUNDIDAD RELATIVA FECUNDIDAD TOTAL
## Min.   :2013-08-11 00:00:00  Min.   :1127          Min.   : 3146
## 1st Qu.:2016-07-05 00:00:00  1st Qu.:2040          1st Qu.: 6010
## Median :2016-07-13 00:00:00  Median :2221          Median : 7897
## Mean   :2016-10-23 19:27:04  Mean   :2259          Mean   : 8481
```

```

## 3rd Qu.:2019-07-04 00:00:00 3rd Qu.:2446 3rd Qu.:10236
## Max. :2019-07-10 00:00:00 Max. :4772 Max. :23569
## NA's :124 NA's :90 NA's :90
## FERTILIDAD FINAL SF FOTOPERÍODO LÍNEA
## Min. : 0.00 Min. :145.0 Length:678 Length:678
## 1st Qu.: 85.83 1st Qu.:398.0 Class :character Class :character
## Median : 92.69 Median :481.0 Mode :character Mode :character
## Mean : 86.79 Mean :444.4
## 3rd Qu.: 96.15 3rd Qu.:494.0
## Max. :100.00 Max. :500.0
## NA's :62 NA's :124
## MORTALIDAD A SF N° CRUZA N° OVAS EN 30 CM N° SF TRASLADADOS
## Min. : 0.00 Min. : 1.0 Min. :37.00 Min. :130.0
## 1st Qu.: 3.00 1st Qu.: 43.0 1st Qu.:55.00 1st Qu.:130.0
## Median : 6.00 Median : 85.0 Median :59.00 Median :250.0
## Mean : 10.76 Mean :102.7 Mean :58.53 Mean :241.3
## 3rd Qu.: 11.00 3rd Qu.:151.0 3rd Qu.:62.00 3rd Qu.:260.0
## Max. :149.00 Max. :290.0 Max. :73.00 Max. :407.0
## NA's :124 NA's :32 NA's :34
## OBSERVACIONES OVAS MUERTAS OVAS PMG OVAS PRODUCCIÓN
## Length:678 Min. : 120.0 Min. :151.0 Min. : 2709
## Class :character 1st Qu.: 596.0 1st Qu.:407.2 1st Qu.: 6016
## Mode :character Median : 959.5 Median :500.0 Median : 7368
## Mean : 1326.0 Mean :455.1 Mean : 7942
## 3rd Qu.: 1655.8 3rd Qu.:500.0 3rd Qu.: 8909
## Max. :11204.0 Max. :500.0 Max. :23069
## NA's :124 NA's :124 NA's :124
## PESO MADRE PESO OVAS PESO PADRE PESO PROMEDIO PADRES
## Min. : 1060 Min. : 264.0 Min. : 2100 Min. : 2080
## 1st Qu.: 2900 1st Qu.: 678.0 1st Qu.: 4000 1st Qu.: 3640
## Median : 3520 Median : 852.5 Median : 4720 Median : 4255
## Mean : 4153 Mean : 945.6 Mean : 5303 Mean : 4728
## 3rd Qu.: 4435 3rd Qu.:1087.8 3rd Qu.: 5635 3rd Qu.: 5010
## Max. :16360 Max. :2940.0 Max. :22360 Max. :17280
## NA's :90
## PIT MADRE PIT PADRE PORCENTAJE MORTALIDAD ALEVINES SF
## Length:678 Length:678 Min. : 0.000
## Class :character Class :character 1st Qu.: 0.600
## Mode :character Mode :character Median : 1.220
## Mean : 2.292
## 3rd Qu.: 2.507
## Max. :32.892
## NA's :125
## PORCENTAJE SOBREVIVENCIA OVA OJO TAMAÑO OVA VOLUMEN OVAS
## Min. : 1.36 Min. :0.000 Min. :1127
## 1st Qu.:80.22 1st Qu.:4.762 1st Qu.:2246
## Median :87.46 Median :4.918 Median :2460
## Mean :84.13 Mean :4.894 Mean :2571
## 3rd Qu.:92.12 3rd Qu.:5.085 3rd Qu.:2851
## Max. :98.54 Max. :6.122 Max. :4772
## NA's :124 NA's :27 NA's :561

```

Transformar los datos

En este punto transformaremos los datos que así lo requieran. Esto debido a que a pesar de haber identificado en un paso inicial el tipo de dato contenido en cada columna, este tipo de dato puede no ser útil para las mediciones o evaluaciones a realizar posteriormente.

```
dim(FDL2E1YC)
```

```
## [1] 678 42
```

```
FDL2E1YC$`CATEGORÍA HEMBRA` <-as.factor(FDL2E1YC$`CATEGORÍA HEMBRA`)
FDL2E1YC$`CATEGORÍA MACHO` <-as.factor(FDL2E1YC$`CATEGORÍA MACHO`)
FDL2E1YC$`CAUSA INVIABILIDAD` <-as.factor(FDL2E1YC$`CAUSA INVIABILIDAD`)
FDL2E1YC$`CEPA` <-as.factor(FDL2E1YC$`CEPA`)
FDL2E1YC$`CLASE AÑO` <-as.factor(FDL2E1YC$`CLASE AÑO`)
FDL2E1YC$`CÓDIGO DESOVE H` <-as.factor(FDL2E1YC$`CÓDIGO DESOVE H`)
FDL2E1YC$`CÓDIGO DESOVE M` <-as.factor(FDL2E1YC$`CÓDIGO DESOVE M`)
FDL2E1YC$`DAÑO OCULAR` <-as.factor(FDL2E1YC$`DAÑO OCULAR`)
FDL2E1YC$`EDAD HEMBRA` <-as.factor(FDL2E1YC$`EDAD HEMBRA`)
FDL2E1YC$`EDAD MACHO` <-as.factor(FDL2E1YC$`EDAD MACHO`)
FDL2E1YC$`ESTADO DE CRUZA` <-as.factor(FDL2E1YC$`ESTADO DE CRUZA`)
FDL2E1YC$`FAMILIA` <-as.factor(FDL2E1YC$`FAMILIA`)
FDL2E1YC$`FECHA DESOVE MADRE` <-as.Date(FDL2E1YC$`FECHA DESOVE MADRE`,format="%d/%m/%y")
FDL2E1YC$`FECHA DESOVE PADRE` <-as.Date(FDL2E1YC$`FECHA DESOVE PADRE`,format="%d/%m/%y")
FDL2E1YC$`FECHA LIMPIEZA OVAS` <-as.Date(FDL2E1YC$`FECHA LIMPIEZA OVAS`,format="%d/%m/%y")
FDL2E1YC$`FECHA SIEMBRA OVA` <-as.Date(FDL2E1YC$`FECHA SIEMBRA OVA`,format="%d/%m/%y")
FDL2E1YC$`FOTOPERÍODO` <-as.factor(FDL2E1YC$`FOTOPERÍODO`)
FDL2E1YC$`LÍNEA` <-as.factor(FDL2E1YC$`LÍNEA`)
FDL2E1YC$`FOTOPERÍODO` <-as.factor(FDL2E1YC$`FOTOPERÍODO`)
FDL2E1YC$`LÍNEA` <-as.factor(FDL2E1YC$`LÍNEA`)
FDL2E1YC$`N° CRUZA` <-as.factor(FDL2E1YC$`N° CRUZA`)
FDL2E1YC$`N° SF TRASLADADOS` <-as.factor(FDL2E1YC$`N° SF TRASLADADOS`)
FDL2E1YC$`OBSERVACIONES` <-as.factor(FDL2E1YC$`OBSERVACIONES`)
FDL2E1YC$`PIT MADRE` <-as.factor(FDL2E1YC$`PIT MADRE`)
FDL2E1YC$`PIT PADRE` <-as.factor(FDL2E1YC$`PIT PADRE`)
summary(FDL2E1YC)
```

```
##      ARCHIVO      CATEGORÍA HEMBRA  CATEGORÍA MACHO      CAUSA INVIABILIDAD
## Length:678      Selección:588      Selección:588      Afixia          : 4
## Class :character NA's      : 90      NA's      : 90      Alta mortalidad: 6
## Mode  :character                                     Baja fertilidad: 3
##                                                    Eliminada x BKD: 1
##                                                    NA's          :664
##
##
##      CEPA      CLASE AÑO      CÓDIGO DESOVE H      CÓDIGO DESOVE M      DAÑO OCULAR
## Fanad:678      2010: 90      HFF16-1015: 1      MFN13-5065: 4      NA's:678
##                                     2013:117      HFF16-1016: 1      MFN16-084 : 4
##                                     2016:290      HFF16-1017: 1      MFN16-196 : 4
##                                     2019:181      HFF16-1321: 1      MFN16-197 : 4
##                                     HFF16-1490: 1      MFN16-198 : 4
##                                     HFF16-1491: 1      MFN16-199 : 4
```

```

##          (Other) :672      (Other) :654
## EDAD HEMBRA EDAD MACHO ESTADO DE CRUZA FAMILIA FECHA DESOVE MADRE
## 2w:606      2w:638      Inviabile : 21      0      : 17      Min. :2010-05-05
## 3w: 72      3w: 40      No Viable: 14      N01      : 3      1st Qu.:2013-06-13
##          viable :269      N02      : 3      Median :2016-05-18
##          Viable :284      N03      : 3      Mean :2015-11-17
##          NA's : 90      N06      : 3      3rd Qu.:2019-05-02
##          (Other):545      Max. :2019-05-28
##          NA's :104
## FECHA DESOVE PADRE FECHA DESPACHO FECHA LIMPIEZA OVAS
## Min. :2010-05-05 Min. :2013-09-25 00:00:00 Min. :2013-06-17
## 1st Qu.:2013-06-13 1st Qu.:2017-10-06 00:00:00 1st Qu.:2016-06-17
## Median :2016-05-18 Median :2017-10-06 00:00:00 Median :2016-07-13
## Mean :2015-04-30 Mean :2017-06-08 04:27:46 Mean :2016-10-09
## 3rd Qu.:2017-04-11 3rd Qu.:2019-08-19 00:00:00 3rd Qu.:2019-06-14
## Max. :2017-05-02 Max. :2019-08-19 00:00:00 Max. :2019-06-27
##          NA's :194          NA's :125
## FECHA SIEMBRA OVA FECUNDIDAD RELATIVA FECUNDIDAD TOTAL FERTILIDAD
## Min. :2013-08-11 Min. :1127 Min. : 3146 Min. : 0.00
## 1st Qu.:2016-07-05 1st Qu.:2040 1st Qu.: 6010 1st Qu.: 85.83
## Median :2016-07-13 Median :2221 Median : 7897 Median : 92.69
## Mean :2016-10-23 Mean :2259 Mean : 8481 Mean : 86.79
## 3rd Qu.:2019-07-04 3rd Qu.:2446 3rd Qu.:10236 3rd Qu.: 96.15
## Max. :2019-07-10 Max. :4772 Max. :23569 Max. :100.00
## NA's :124 NA's :90 NA's :90 NA's :62
## FINAL SF FOTOPERÍODO LÍNEA MORTALIDAD A SF N° CRUZA
## Min. :145.0 No :529 L2:677 Min. : 0.00 1 : 4
## 1st Qu.:398.0 Si : 59 L3: 1 1st Qu.: 3.00 2 : 4
## Median :481.0 NA's: 90 Median : 6.00 3 : 4
## Mean :444.4 Mean : 10.76 4 : 4
## 3rd Qu.:494.0 3rd Qu.: 11.00 5 : 4
## Max. :500.0 Max. :149.00 6 : 4
## NA's :124 NA's :124 (Other):654
## N° OVAS EN 30 CM N° SF TRASLADADOS
## Min. :37.00 250 :270
## 1st Qu.:55.00 130 :167
## Median :59.00 260 :117
## Mean :58.53 400 : 80
## 3rd Qu.:62.00 310 : 1
## Max. :73.00 (Other): 9
## NA's :32 NA's : 34
## OBSERVACIONES
## Causa de mort. Post-eclosión es de normal y deformes : 26
## Causa de mort. Post-eclosión es de normal y doble saco : 58
## Causa de mort. Post-eclosión es de normal, doble saco y deformes: 13
## NO : 49
## SUPLENTE : 20
## TITULAR :200
## NA's :312
## OVAS MUERTAS OVAS PMG OVAS PRODUCCIÓN PESO MADRE
## Min. : 120.0 Min. :151.0 Min. : 2709 Min. : 1060
## 1st Qu.: 596.0 1st Qu.:407.2 1st Qu.: 6016 1st Qu.: 2900
## Median : 959.5 Median :500.0 Median : 7368 Median : 3520
## Mean : 1326.0 Mean :455.1 Mean : 7942 Mean : 4153

```

```
## 3rd Qu.: 1655.8 3rd Qu.:500.0 3rd Qu.: 8909 3rd Qu.: 4435
## Max. :11204.0 Max. :500.0 Max. :23069 Max. :16360
## NA's :124 NA's :124 NA's :124
## PESO OVAS PESO PADRE PESO PROMEDIO PADRES PIT MADRE
## Min. : 264.0 Min. : 2100 Min. : 2080 0001e6a236: 1
## 1st Qu.: 678.0 1st Qu.: 4000 1st Qu.: 3640 0001e6b5bd: 1
## Median : 852.5 Median : 4720 Median : 4255 000204CB67: 1
## Mean : 945.6 Mean : 5303 Mean : 4728 0006324C1F: 1
## 3rd Qu.:1087.8 3rd Qu.: 5635 3rd Qu.: 5010 0006467C19: 1
## Max. :2940.0 Max. :22360 Max. :17280 00065923B4: 1
## NA's :90 (Other) :672
## PIT PADRE PORCENTAJE MORTALIDAD ALEVINES SF
## 00066E3D33: 4 Min. : 0.000
## 000695022A: 4 1st Qu.: 0.600
## 0006950A21: 4 Median : 1.220
## 0006950F82: 4 Mean : 2.292
## 00069514CC: 4 3rd Qu.: 2.507
## 00069519A4: 4 Max. :32.892
## (Other) :654 NA's :125
## PORCENTAJE SOBREVIVENCIA OVA OJO TAMAÑO OVA VOLUMEN OVAS
## Min. : 1.36 Min. :0.000 Min. :1127
## 1st Qu.:80.22 1st Qu.:4.762 1st Qu.:2246
## Median :87.46 Median :4.918 Median :2460
## Mean :84.13 Mean :4.894 Mean :2571
## 3rd Qu.:92.12 3rd Qu.:5.085 3rd Qu.:2851
## Max. :98.54 Max. :6.122 Max. :4772
## NA's :124 NA's :27 NA's :561
```

Revisar balance de datos

En los siguientes pasos se obtienen tablas de frecuencia para determinar el estado balanceado, o no, de los datos con los que se cuenta.

En esta primera tabla podemos observar las líneas (PREGUNTAR QUE ES ESTO) empleadas por año, y se determina que respecto de esta variable cualitativa los datos no se encuentran balanceados ya que en 2016 se encuentra 1 observación de la línea 3 (L3), siendo que en el resto de los años esto no se observa, solo se encuentran observaciones de la línea 2 (L2).

```
knitr::kable(table(FDL2E1YC$LÍNEA, FDL2E1YC$`CLASE AÑO`), caption = "Tabla 1. Línea de cultivo empleada")
```

Table 1: Tabla 1. Línea de cultivo empleada en cada año evaluado.

	2010	2013	2016	2019
L2	90	117	289	181
L3	0	0	1	0

Esta segunda tabla nos permite ver que hay una amplia variedad en el número de cruza durante cada uno de los años evaluados, este número de cruza es igual al número de observaciones por año, por lo que respecto de esta variable, los datos tampoco se encuentran balanceados.

```
#table(FDL2E1YC$`N° CRUZA`, FDL2E1YC$`CLASE AÑO`)
```

Esta tercera tabla nos permite ver la edad que tenían las hembras seleccionadas para cruza durante cada uno de los años evaluados. Nuevamente, respecto de esta variable, se observa que únicamente en el año 2013 hay un balance en los datos, ya que se emplearon únicamente hembras de 2 semanas de edad. Sin embargo, en el conjunto de los datos no existe un balance ya que no se emplea el mismo número de hembras de 2 y 3 semanas de edad.

```
knitr::kable(table(FDL2E1YC$`EDAD HEMBRA`, FDL2E1YC$`CLASE AÑO`), caption = "Tabla 3. Edad de las hembras")
```

Table 2: Tabla 3. Edad de las hembras en cada año evaluado.

	2010	2013	2016	2019
2w	69	117	244	176
3w	21	0	46	5

Esta cuarta tabla, al igual que la anterior, nos permite ver la edad de los machos seleccionados para cruza durante cada uno de los años evaluados. Nuevamente, respecto de esta variable, se observa que únicamente en el año 2013 hay un balance en los datos, ya que se emplearon únicamente hembras de 2 semanas de edad. Sin embargo, en el conjunto de los datos no existe un balance, esto debido a que no se emplea el mismo número de machos de 2 y 3 semanas de edad. Adicionalmente, en los años 2016 y 2019 no existe un balance entre el número de machos y hembras empleados de la misma edad.

```
knitr::kable(table(FDL2E1YC$`EDAD MACHO`, FDL2E1YC$`CLASE AÑO`), caption = "Tabla 4. Edad de los machos")
```

Table 3: Tabla 4. Edad de los machos en cada año evaluado.

	2010	2013	2016	2019
2w	69	117	271	181
3w	21	0	19	0

Esta quinta tabla, nos permite observar el número de individuos que fueron, o no, expuestos a fotoperíodo, en cada uno de los años a evaluar. Respecto de esta variable, los datos tampoco se encuentran balanceados. De forma individual solo están balanceados en los años 2010 y 2013, pero en conjunto no se cuenta con datos balanceados respecto de esta variable, ya que no tenemos el mismo número de de individuos expuestos a fotoperíodo en cada uno de los años.

```
knitr::kable(table(FDL2E1YC$`FOTOPERÍODO`, FDL2E1YC$`CLASE AÑO`), caption = "Tabla 5. Individuos expuestos a fotoperíodo")
```

Table 4: Tabla 5. Individuos expuestos a fotoperíodo en cada año evaluado.

	2010	2013	2016	2019
No	0	117	236	176
Si	0	0	54	5

Esta última tabla nos permite identificar el número de cruzas que fueron viables e inviables por cada uno de los años a evaluar. Podemos percatarnos de que las 4 opciones que aparecen en la tabla representan solo 2

opciones pero escritas de forma distinta, por lo que se detectan como 4. Respecto de esta variable podemos determinar que los datos no se encuentran balanceados, ya que el número de cruza viables e inviables por año es distinto.

```
knitr::kable(table(FDL2E1YC$`ESTADO DE CRUZA`, FDL2E1YC$`CLASE AÑO`), caption = "Tabla 6. Viabilidad de
```

Table 5: Tabla 6. Viabilidad de las cruza en cada año evaluado.

	2010	2013	2016	2019
Inviaible	0	0	21	0
No Viable	0	0	0	14
viable	0	0	269	0
Viable	0	117	0	167

Revisar variables de estudio

Variables relacionadas con los reproductores.

Aquí se observa que la mayoría de las hembras o madres tienen un peso por debajo de los 5kg, mientras que la mayoría de los machos o padres tienen un peso entre los 5kg y los 7kg. Siendo que, respecto del promedio del peso de la pareja reproductora, la mayoría se ubican desde los 5kg y por debajo de este valor.

```
par(mfrow=c(2,2))
((hist(FDL2E1YC$`PESO MADRE`,main="PESO MADRE", col = "aquamarine3")))
```

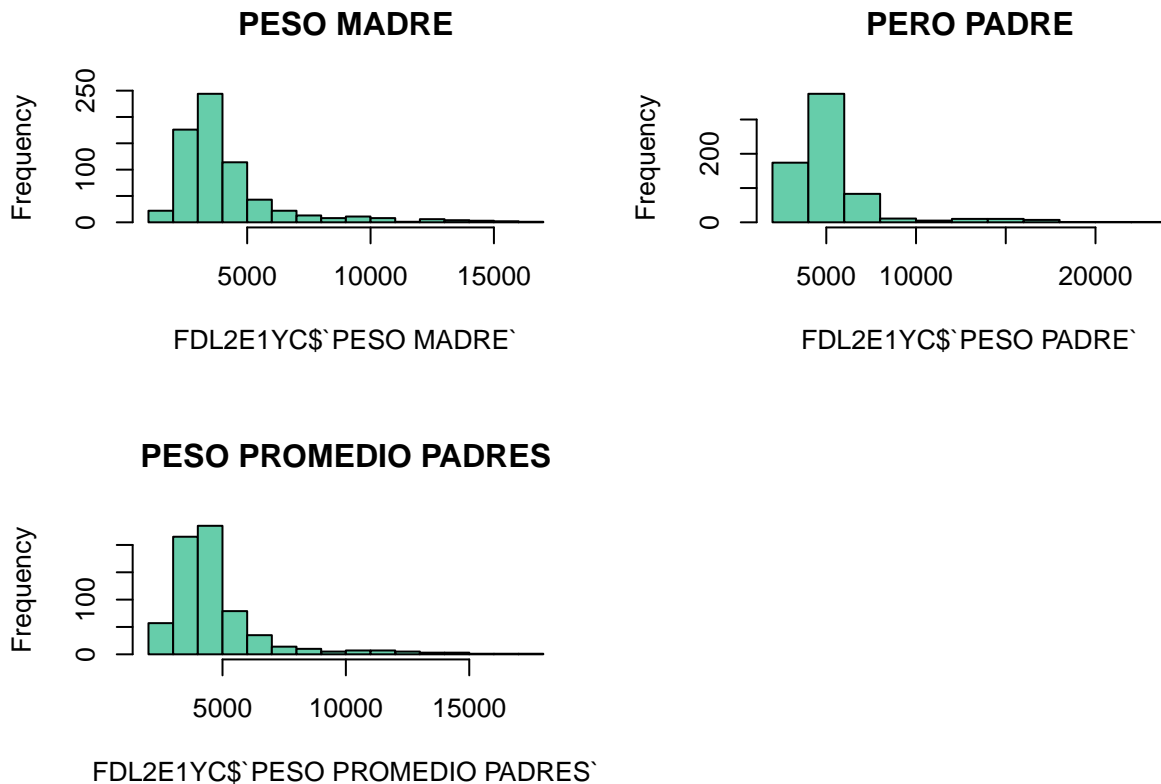
```
## $breaks
## [1] 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000
## [13] 13000 14000 15000 16000 17000
##
## $counts
## [1] 22 176 244 114 43 22 13 8 11 8 1 6 4 3 2 1
##
## $density
## [1] 3.244838e-05 2.595870e-04 3.598820e-04 1.681416e-04 6.342183e-05
## [6] 3.244838e-05 1.917404e-05 1.179941e-05 1.622419e-05 1.179941e-05
## [11] 1.474926e-06 8.849558e-06 5.899705e-06 4.424779e-06 2.949853e-06
## [16] 1.474926e-06
##
## $mids
## [1] 1500 2500 3500 4500 5500 6500 7500 8500 9500 10500 11500 12500
## [13] 13500 14500 15500 16500
##
## $xname
## [1] "FDL2E1YC$`PESO MADRE`"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`PESO PADRE`,main="PERO PADRE", col = "aquamarine3")))
```

```
## $breaks
## [1] 2000 4000 6000 8000 10000 12000 14000 16000 18000 20000 22000 24000
##
## $counts
## [1] 174 375 83 11 5 10 10 7 1 1 1
##
## $density
## [1] 1.283186e-04 2.765487e-04 6.120944e-05 8.112094e-06 3.687316e-06
## [6] 7.374631e-06 7.374631e-06 5.162242e-06 7.374631e-07 7.374631e-07
## [11] 7.374631e-07
##
## $mids
## [1] 3000 5000 7000 9000 11000 13000 15000 17000 19000 21000 23000
##
## $xname
## [1] "FDL2E1YC$`PESO PADRE`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`PESO PROMEDIO PADRES`,main="PESO PROMEDIO PADRES", col = "aquamarine3")))
```

```
## $breaks
## [1] 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000 13000
## [13] 14000 15000 16000 17000 18000
##
## $counts
## [1] 57 215 235 79 35 14 10 5 7 7 5 3 3 1 1 1
##
## $density
## [1] 8.407080e-05 3.171091e-04 3.466077e-04 1.165192e-04 5.162242e-05
## [6] 2.064897e-05 1.474926e-05 7.374631e-06 1.032448e-05 1.032448e-05
## [11] 7.374631e-06 4.424779e-06 4.424779e-06 1.474926e-06 1.474926e-06
## [16] 1.474926e-06
##
## $mids
## [1] 2500 3500 4500 5500 6500 7500 8500 9500 10500 11500 12500 13500
## [13] 14500 15500 16500 17500
##
## $xname
## [1] "FDL2E1YC$`PESO PROMEDIO PADRES`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```



Variables relacionadas con la reproducción.

La fecundidad total de los reproductores, medida respecto del número de ovas obtenidas, se encuentra en su mayoría entre 5-10mil ovas. Mientras que la fecundidad relativa, medida como el número de ovas sobre el peso de las mismas, se encuentra entre las 2mil ovas. Vemos además que las cruza o parejas reproductoras presentan un porcentaje de fertilidad superior al 60%, dónde la mayoría de las parejas presenta entre un 90-100% de fertilidad.

```
par(mfrow=c(2,2))
((hist(FDL2E1YC$`FECUNDIDAD TOTAL`,main="FECUNDIDAD TOTAL", col = "coral")))
```

```
## $breaks
## [1] 2000 4000 6000 8000 10000 12000 14000 16000 18000 20000 22000 24000
##
## $counts
## [1] 10 137 155 136 76 39 12 10 8 3 2
##
## $density
## [1] 8.503401e-06 1.164966e-04 1.318027e-04 1.156463e-04 6.462585e-05
## [6] 3.316327e-05 1.020408e-05 8.503401e-06 6.802721e-06 2.551020e-06
## [11] 1.700680e-06
##
## $mids
## [1] 3000 5000 7000 9000 11000 13000 15000 17000 19000 21000 23000
```

```
##
## $xname
## [1] "FDL2E1YC$`FECUNDIDAD TOTAL`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

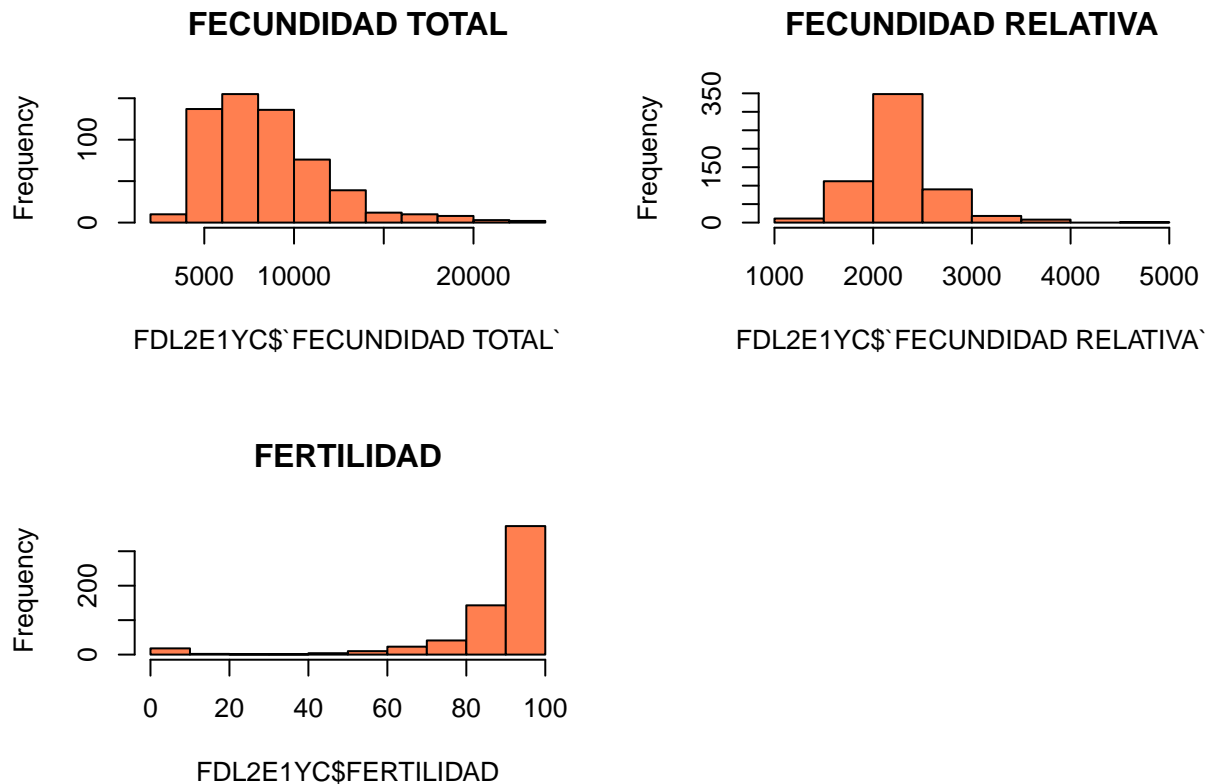
```
((hist(FDL2E1YC$`FECUNDIDAD RELATIVA`,main="FECUNDIDAD RELATIVA", col = "coral")))
```

```
## $breaks
## [1] 1000 1500 2000 2500 3000 3500 4000 4500 5000
##
## $counts
## [1] 11 112 348 90 18 8 0 1
##
## $density
## [1] 3.741497e-05 3.809524e-04 1.183673e-03 3.061224e-04 6.122449e-05
## [6] 2.721088e-05 0.000000e+00 3.401361e-06
##
## $mids
## [1] 1250 1750 2250 2750 3250 3750 4250 4750
##
## $xname
## [1] "FDL2E1YC$`FECUNDIDAD RELATIVA`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`FERTILIDAD`,main="FERTILIDAD", col = "coral")))
```

```
## $breaks
## [1] 0 10 20 30 40 50 60 70 80 90 100
##
## $counts
## [1] 18 2 1 1 4 10 23 41 143 373
##
## $density
## [1] 0.0029220779 0.0003246753 0.0001623377 0.0001623377 0.0006493506
## [6] 0.0016233766 0.0037337662 0.0066558442 0.0232142857 0.0605519481
##
## $mids
## [1] 5 15 25 35 45 55 65 75 85 95
##
## $xname
## [1] "FDL2E1YC$FERTILIDAD"
##
## $equidist
```

```
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```



Variables relacionadas con la calidad de las ovas.

Respecto de la calidad de las ovas, podemos observar que el tamaño de estas, medido por el diámetro de la ova, oscila entre 4-6mm. Asimismo, la mayor frecuencia para el volumen de la ova se encuentra entre los 2-3L. El peso de las ovas, por otro lado, presenta en su mayoría valores de entre los 0.5-1.5kg. Por último, el número de ovas de 30cm oscila mayormente entre las 55-65 ovas.

```
par(mfrow=c(2,2))
((hist(FDL2E1YC$`TAMAÑO OVA`,main="TAMAÑO OVA", col = "seagreen")))
```

```
## $breaks
## [1] 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5
##
## $counts
## [1] 5 0 0 0 0 0 0 0 0 29 417 180 16 4
##
## $density
## [1] 0.01536098 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## [7] 0.00000000 0.00000000 0.08909370 1.28110599 0.55299539 0.04915515
```

```
## [13] 0.01228879
##
## $mids
## [1] 0.25 0.75 1.25 1.75 2.25 2.75 3.25 3.75 4.25 4.75 5.25 5.75 6.25
##
## $xname
## [1] "FDL2E1YC$`TAMAÑO OVA`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`VOLUMEN OVAS`,main="VOLUMEN OVAS", col = "seagreen")))
```

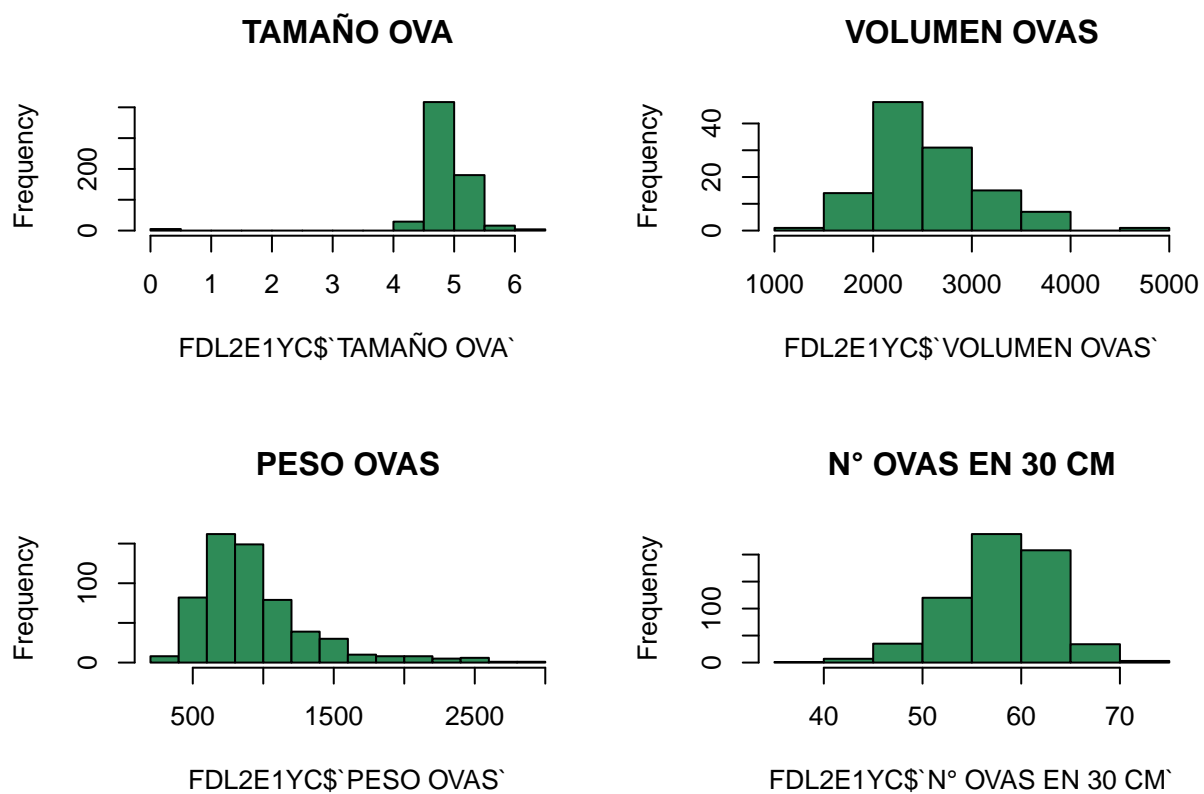
```
## $breaks
## [1] 1000 1500 2000 2500 3000 3500 4000 4500 5000
##
## $counts
## [1] 1 14 48 31 15 7 0 1
##
## $density
## [1] 1.709402e-05 2.393162e-04 8.205128e-04 5.299145e-04 2.564103e-04
## [6] 1.196581e-04 0.000000e+00 1.709402e-05
##
## $mids
## [1] 1250 1750 2250 2750 3250 3750 4250 4750
##
## $xname
## [1] "FDL2E1YC$`VOLUMEN OVAS`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`PESO OVAS`,main="PESO OVAS", col = "seagreen")))
```

```
## $breaks
## [1] 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000
##
## $counts
## [1] 8 82 162 149 79 39 30 10 8 8 5 6 1 1
##
## $density
## [1] 6.802721e-05 6.972789e-04 1.377551e-03 1.267007e-03 6.717687e-04
## [6] 3.316327e-04 2.551020e-04 8.503401e-05 6.802721e-05 6.802721e-05
## [11] 4.251701e-05 5.102041e-05 8.503401e-06 8.503401e-06
##
## $mids
## [1] 300 500 700 900 1100 1300 1500 1700 1900 2100 2300 2500 2700 2900
```

```
##
## $xname
## [1] "FDL2E1YC$`PESO OVAS`"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`N° OVAS EN 30 CM`,main="N° OVAS EN 30 CM", col = "seagreen")))
```



```
## $breaks
## [1] 35 40 45 50 55 60 65 70 75
##
## $counts
## [1] 1 7 35 120 238 208 34 3
##
## $density
## [1] 0.0003095975 0.0021671827 0.0108359133 0.0371517028 0.0736842105
## [6] 0.0643962848 0.0105263158 0.0009287926
##
## $mids
## [1] 37.5 42.5 47.5 52.5 57.5 62.5 67.5 72.5
##
```

```
## $xname
## [1] "FDL2E1YC$`N° OVAS EN 30 CM`"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

Variables relacionadas con la viabilidad de las ovas.

Respecto de la viabilidad de las ovas, un número mínimo de ovas murieron durante los 4 años evaluados, con valores oscilando mayormente entre las 0-1mil ovas muertas. De las ovas sobrevivientes la mayoría fueron seleccionadas para el Programa de Mejora Genética. En cuanto a la producción de ovas, en su mayoría los padres produjeron entre 5-10mil ovas. Por último, respecto de la supervivencia de las ovas, medida con base en la visualización del ojo del embrión, se observa que en su mayoría sobreviven entre el 60-100% de las ovas.

```
par(mfrow=c(2,2))
((hist(FDL2E1YC$`OVAS MUERTAS`,main="OVAS MUERTAS", col = "skyblue")))
```

```
## $breaks
## [1] 0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000
## [13] 12000
##
## $counts
## [1] 290 165 55 26 9 5 2 1 0 0 0 1
##
## $density
## [1] 5.234657e-04 2.978339e-04 9.927798e-05 4.693141e-05 1.624549e-05
## [6] 9.025271e-06 3.610108e-06 1.805054e-06 0.000000e+00 0.000000e+00
## [11] 0.000000e+00 1.805054e-06
##
## $mids
## [1] 500 1500 2500 3500 4500 5500 6500 7500 8500 9500 10500 11500
##
## $xname
## [1] "FDL2E1YC$`OVAS MUERTAS`"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`OVAS PMG`,main="OVAS PMG", col = "skyblue")))
```

```
## $breaks
## [1] 150 200 250 300 350 400 450 500
##
## $counts
## [1] 6 9 16 15 91 35 382
##
```



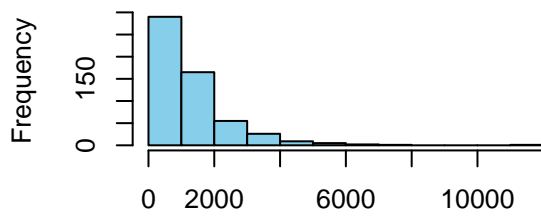
```
## $density
## [1] 0.0002166065 0.0003249097 0.0005776173 0.0005415162 0.0032851986
## [6] 0.0012635379 0.0137906137
##
## $mids
## [1] 175 225 275 325 375 425 475
##
## $xname
## [1] "FDL2E1YC$`OVAS PMG`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`OVAS PRODUCCIÓN`,main="OVAS PRODUCCIÓN", col = "skyblue")))
```

```
## $breaks
## [1] 2000 4000 6000 8000 10000 12000 14000 16000 18000 20000 22000 24000
##
## $counts
## [1] 14 124 204 112 58 16 11 5 6 3 1
##
## $density
## [1] 1.263538e-05 1.119134e-04 1.841155e-04 1.010830e-04 5.234657e-05
## [6] 1.444043e-05 9.927798e-06 4.512635e-06 5.415162e-06 2.707581e-06
## [11] 9.025271e-07
##
## $mids
## [1] 3000 5000 7000 9000 11000 13000 15000 17000 19000 21000 23000
##
## $xname
## [1] "FDL2E1YC$`OVAS PRODUCCIÓN`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

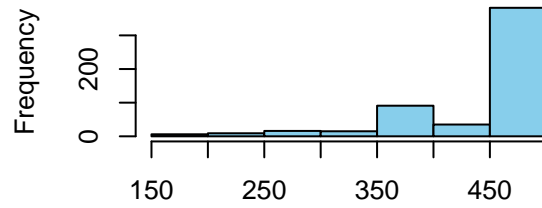
```
((hist(FDL2E1YC$`PORCENTAJE SOBREVIVENCIA OVA OJO`,main="% SOBREVIVENCIA OVA", col = "skyblue")))
```

OVAS MUERTAS



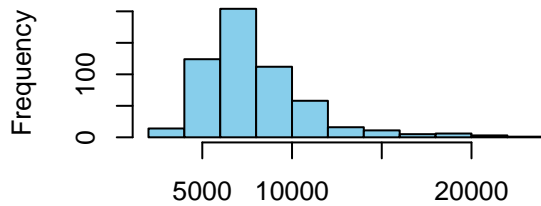
FDL2E1YC\$`OVAS MUERTAS`

OVAS PMG



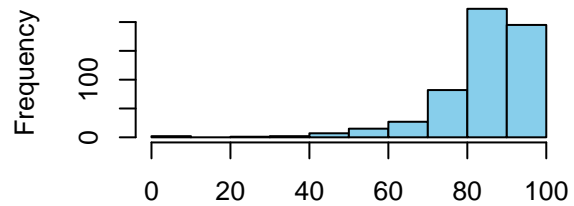
FDL2E1YC\$`OVAS PMG`

OVAS PRODUCCIÓN



FDL2E1YC\$`OVAS PRODUCCIÓN`

% SOBREVIVENCIA OVA



FDL2E1YC\$`PORCENTAJE SOBREVIVENCIA OVA

```
## $breaks
## [1] 0 10 20 30 40 50 60 70 80 90 100
##
## $counts
## [1] 2 0 1 2 7 15 27 82 223 195
##
## $density
## [1] 0.0003610108 0.0000000000 0.0001805054 0.0003610108 0.0012635379
## [6] 0.0027075812 0.0048736462 0.0148014440 0.0402527076 0.0351985560
##
## $mids
## [1] 5 15 25 35 45 55 65 75 85 95
##
## $xname
## [1] "FDL2E1YC$`PORCENTAJE SOBREVIVENCIA OVA OJO`"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

Variables relacionadas con la supervivencia de alevines.

Respecto de la supervivencia de los alevines, se puede observar que la mortalidad, desde la observación de supervivencia de la ova por el ojo del embrión hasta la etapa que este comienza a comer, tiene una baja frecuencia, ya que en los años evaluados solo se observa una mortalidad de entre 0-30 individuos. Por ende, se obtienen, en su mayoría, cada año evaluado, alrededor de 500 individuos sanos y que sobreviven hasta después de la etapa en que comienzan a comer. Por lo que el porcentaje de mortalidad de alevines, es también muy bajo, entre el 0-5%.

```
par(mfrow=c(2,2))
((hist(FDL2E1YC$`MORTALIDAD A SF`,main="MORTALIDAD A SF", col = "salmon1")))
```

```
## $breaks
## [1] 0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
##
## $counts
## [1] 404 91 31 8 2 3 3 0 2 0 2 5 1 0 2
##
## $density
## [1] 0.0729241877 0.0164259928 0.0055956679 0.0014440433 0.0003610108
## [6] 0.0005415162 0.0005415162 0.0000000000 0.0003610108 0.0000000000
## [11] 0.0003610108 0.0009025271 0.0001805054 0.0000000000 0.0003610108
##
## $mids
## [1] 5 15 25 35 45 55 65 75 85 95 105 115 125 135 145
##
## $xname
## [1] "FDL2E1YC$`MORTALIDAD A SF`"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

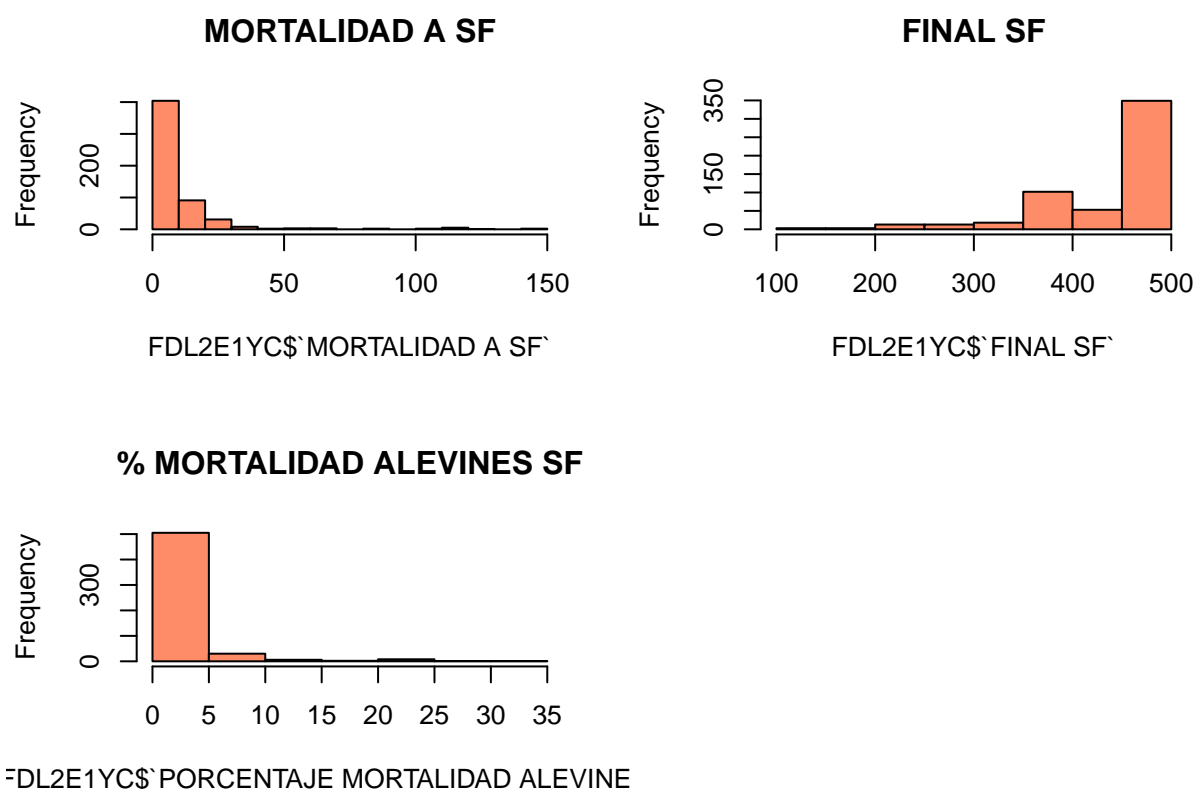
```
((hist(FDL2E1YC$`FINAL SF`,main="FINAL SF", col = "salmon1")))
```

```
## $breaks
## [1] 100 150 200 250 300 350 400 450 500
##
## $counts
## [1] 3 3 13 13 18 102 53 349
##
## $density
## [1] 0.0001083032 0.0001083032 0.0004693141 0.0004693141 0.0006498195
## [6] 0.0036823105 0.0019133574 0.0125992780
##
## $mids
## [1] 125 175 225 275 325 375 425 475
##
## $xname
## [1] "FDL2E1YC$`FINAL SF`"
##
```

```
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```

```
((hist(FDL2E1YC$`PORCENTAJE MORTALIDAD ALEVINES SF`,main="% MORTALIDAD ALEVINES SF", col = "salmon1")))
```

```
## $breaks
## [1] 0 5 10 15 20 25 30 35
##
## $counts
## [1] 505 30 6 2 8 1 1
##
## $density
## [1] 0.1826401447 0.0108499096 0.0021699819 0.0007233273 0.0028933092
## [6] 0.0003616637 0.0003616637
##
## $mids
## [1] 2.5 7.5 12.5 17.5 22.5 27.5 32.5
##
## $xname
## [1] "FDL2E1YC$`PORCENTAJE MORTALIDAD ALEVINES SF`"
##
## $equidist
## [1] TRUE
##
## attr("class")
## [1] "histogram"
```



Relación entre variables cuantitativas y factores

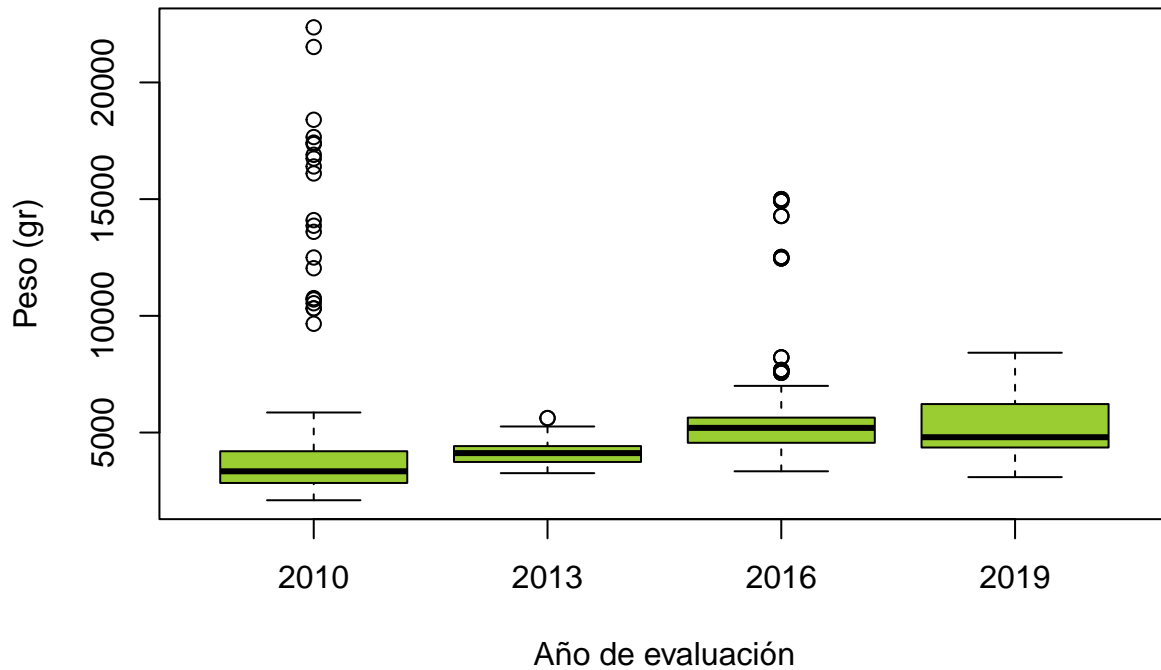
En este punto procederemos a realizar evaluaciones visuales entre las variables y factores de nuestra base de datos.

Peso de los padres VS año de evaluación

A partir del gráfico y las tablas de resumen presentadas a continuación para la variable de peso de los padres, podemos ver que para el año 2019 no se observan valores o datos atípicos como se sucede en los años anteriores. Siendo que la mayor variabilidad se observa en el año 2010, año que corresponde al inicio del PMG y por ende a la población inicial, lo que puede ser la causa de esa variabilidad entre los datos, donde se perciben valores atípicos como individuos con peso de hasta más de 20kg. Aquí resalta que en el año 2016 la media fue 400gr superior a la media observada en 2019, sin embargo el peso máximo para el año 2016 no superó los 7kg, esto sin considerar los datos atípicos que incluyen una minoría de individuos de hasta 15kg, mientras que en 2019 se observa una máxima de 8.4kg, sin incluir o considerar datos atípicos debido a la uniformidad del peso en la población. Esto es un indicio de que el Programa de Mejora Genética ha permitido incrementar el peso promedio de los padres.

```
boxplot(FDL2E1YC$`PESO PADRE`~FDL2E1YC$`CLASE AÑO`, main="Peso padres por año", ylab="Peso (gr)", xlab=
```

Peso padres por año



Resumen: Peso de los padres VS año de evaluación

```
PePad10 <- FDL2E1YC %>% filter(`CLASE AÑO` == 2010) %>% summarise("N" = n(), "Promedio"= mean(`PESO PADRE`), "Mediana"= median(`PESO PADRE`), "Min."= min(`PESO PADRE`), "Max."= max(`PESO PADRE`), "DS"= sd(`PESO PADRE`))
PePad13 <- FDL2E1YC %>% filter(`CLASE AÑO` == 2013) %>% summarise("N" = n(), "Promedio"= mean(`PESO PADRE`), "Mediana"= median(`PESO PADRE`), "Min."= min(`PESO PADRE`), "Max."= max(`PESO PADRE`), "DS"= sd(`PESO PADRE`))
PePad16 <- FDL2E1YC %>% filter(`CLASE AÑO` == 2016) %>% summarise("N" = n(), "Promedio"= mean(`PESO PADRE`), "Mediana"= median(`PESO PADRE`), "Min."= min(`PESO PADRE`), "Max."= max(`PESO PADRE`), "DS"= sd(`PESO PADRE`))
PePad19 <- FDL2E1YC %>% filter(`CLASE AÑO` == 2019) %>% summarise("N" = n(), "Promedio"= mean(`PESO PADRE`), "Mediana"= median(`PESO PADRE`), "Min."= min(`PESO PADRE`), "Max."= max(`PESO PADRE`), "DS"= sd(`PESO PADRE`))
knitr::kable(PePad10, caption = "Tabla 7. Resumen de peso de los padres año 2010.")
```

Table 6: Tabla 7. Resumen de peso de los padres año 2010.

N	Promedio	Mediana	Min.	Max.	DS
90	5876.333	3340	2100	22360	5249.637

```
knitr::kable(PePad13, caption = "Tabla 8. Resumen de peso de los padres año 2013.")
```

Table 7: Tabla 8. Resumen de peso de los padres año 2013.

N	Promedio	Mediana	Min.	Max.	DS
117	4125.128	4120	3260	5620	492.8394

```
knitr::kable(PePad16, caption = "Tabla 9. Resumen de peso de los padres año 2016.")
```

Table 8: Tabla 9. Resumen de peso de los padres año 2016.

N	Promedio	Mediana	Min.	Max.	DS
290	5627.862	5200	3340	15000	2144.34

```
knitr::kable(PePad19, caption = "Tabla 10. Resumen de peso de los padres año 2019.")
```

Table 9: Tabla 10. Resumen de peso de los padres año 2019.

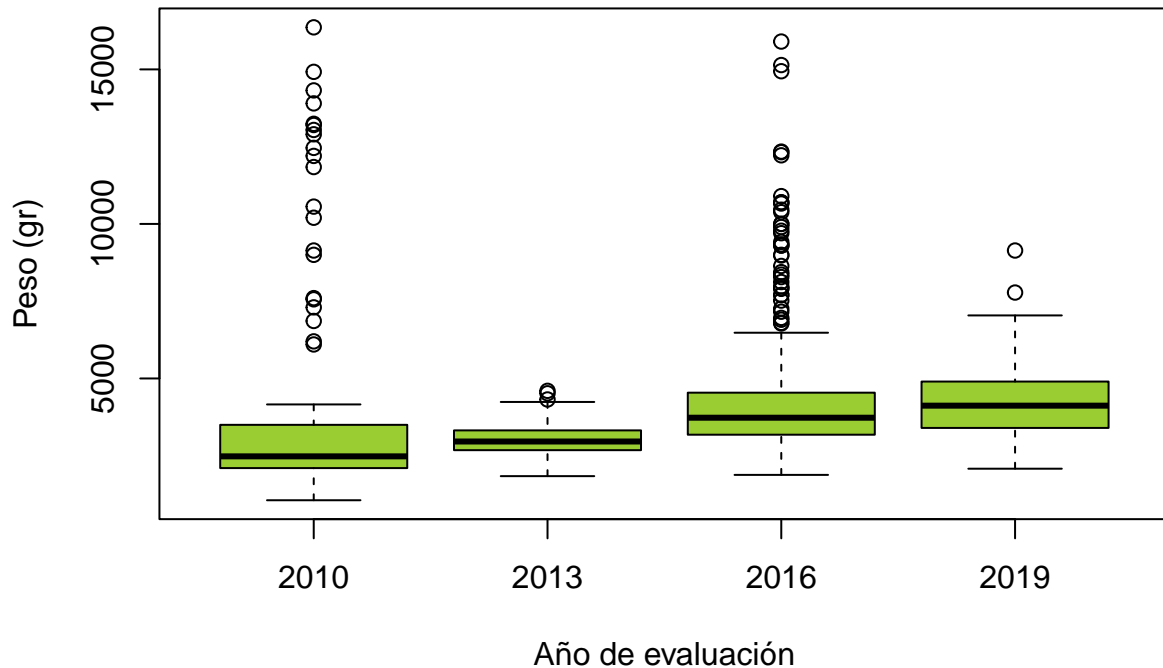
N	Promedio	Mediana	Min.	Max.	DS
181	5258.508	4800	3090	8420	1331.475

Peso de las madres VS año de evaluación

Al igual que como sucede respecto de los padres, las madres también muestran una menor variabilidad o menor proporción de datos por fuera de la distribución normal, en el año 2019. Vemos además que el peso en las madres se ha mantenido por debajo de los 5kg, sin embargo, respecto de la media, resulta evidente que dicho promedio ha ido aumentando durante los años. Esto indica también que el Programa de Mejora Genética ha permitido incrementar el peso promedio de las madres.

```
boxplot(FDL2E1YC$`PESO MADRE`~FDL2E1YC$`CLASE AÑO`, main="Peso madres por año", ylab="Peso (gr)", xlab=
```

Peso madres por año



Peso promedio de ambos padres VS año de evaluación

Aquí vemos nuevamente la alta variabilidad en los pesos de los padres del año 2010, año de comienzo del PMG. Vemos también como en el año 2019 la media del peso de ambos padres es la mayor de entre los 4 años evaluados, con una distribución entre los 3-7kg, y una media entre los 4.5-5kg. Confirmándose así que el PMG ha permitido el incremento en el peso de los padres, por ende mejorando su crecimiento.

```
boxplot(FDL2E1YC$`PESO PROMEDIO PADRES`~FDL2E1YC$`CLASE AÑO`, main="Peso promedio de padres por año", y
```


Peso promedio de padres por año

