

# LAB1

Kalos Lazo

2023-02-08

## R para archivos de datos

Para limpiar el ambiente usamos `rm(list=ls())`

## Carga de los datos

Para cargar una librería:

```
library(readr)
```

Sólo usamos la librería **readr** nos permite leer los archivos de datos. Debemos evitar usar la librería **read.csv** (errores comunes con tildes).

## Lectura de archivos

1. Para leer un archivo de datos en extensión **.csv**:

```
DF <- read_csv("DatosDiversos.csv")
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,  
## e.g.:  
##   dat <- vroom(...)  
##   problems(dat)
```

```
## Rows: 2038 Columns: 21  
## -- Column specification -----  
## Delimiter: ","  
## chr (14): Período, Apellido, Nombre, Peso (kg), Signo del zodiaco, Sexo, Car...  
## dbl (5): Código, Sección, Ciclo, Edad, Tiempo de reacción  
## num (1): Estatura (cm)  
## lgl (1): Indicaciones  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

2. Consultar cuáles son las variables:

```
names(DF)
```

```
## [1] "Período"          "Código"           "Apellido"
## [4] "Nombre"           "Sección"          "Ciclo"
## [7] "Edad"             "Estatura (cm)"    "Peso (kg)"
## [10] "Signo del zodiaco" "Sexo"             "Carrera"
## [13] "Tiempo de reacción" "Juega videojuegos" "Hace deporte"
## [16] "Color favorito"    "Talla de zapato"   "Talla de polo"
## [19] "Sistema operativo" "Fecha de nacimiento" "Indicaciones"
```

3. Instalar librería para corregir nombres variables:

```
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
```

4. Renombrar las variables con tildes a nombres compuestos y guardar en la variable nueva DFN:

```
DF %>% rename(Periodo = Período, Código = Código) -> DFN
```

5. Para ver cuántas filas (número de observaciones) tiene el dataframe usamos:

```
nrow(DFN) # 2038 Filas
```

```
## [1] 2038
```

6. Para ver cuántas columnas (número de variables):

```
ncol(DFN) # 21
```

```
## [1] 21
```

7. Acceder a una posición del DataFrame:

```
DFN[3, 4] # Fila 3, columna 4
```

```
## # A tibble: 1 x 1
##   Nombre
##   <chr>
## 1 Diego Walter
```

## Limpieza de datos

1. Para que R nos de un resumen del DataFrame, usamos la función:

```
summary(DFN)
```

```
##      Periodo          Codigo      Apellido      Nombre
## Length:2038      Min.   :2.018e+07 Length:2038      Length:2038
## Class :character 1st Qu.:2.019e+08 Class :character Class :character
## Mode  :character Median :2.020e+08 Mode  :character Mode  :character
##                Mean  :2.065e+08
##                3rd Qu.:2.021e+08
##                Max.   :2.022e+09
##                NA's   :673
##      Sección      Ciclo      Edad      Estatura (cm)
## Min.   : 1.000      Min.   :1.000      Min.   :16.00      Min.   : 1.2
## 1st Qu.: 2.000      1st Qu.:3.000      1st Qu.:18.00      1st Qu.:157.0
## Median : 4.000      Median :3.000      Median :19.00      Median :168.0
## Mean   : 4.772      Mean   :3.879      Mean   :19.36      Mean   :140.6
## 3rd Qu.: 6.010      3rd Qu.:5.000      3rd Qu.:20.00      3rd Qu.:175.0
## Max.   :16.010      Max.   :9.000      Max.   :53.00      Max.   :203.0
## NA's   :678        NA's   :671      NA's   :659      NA's   :676
##      Peso (kg)      Signo del zodiaco      Sexo      Carrera
## Length:2038      Length:2038      Length:2038      Length:2038
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      Tiempo de reacción Juega videojuegos Hace deporte      Color favorito
## Min.   : 0.187      Length:2038      Length:2038      Length:2038
## 1st Qu.: 237.000      Class :character  Class :character  Class :character
## Median : 267.000      Mode  :character  Mode  :character  Mode  :character
## Mean   : 287.403
## 3rd Qu.: 308.000
## Max.   :3665.000
## NA's   :829
##      Talla de zapato      Talla de polo      Sistema operativo      Fecha de nacimiento
## Length:2038      Length:2038      Length:2038      Length:2038
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      Indicaciones
## Mode :logical
## FALSE:1790
## TRUE :247
## NA's :1
##
##
##
```

2. Cantidad de datos faltantes:

```
sum(is.na(DFN)) # 17350
```

```
## [1] 17350
```

3. Tabla de datos faltantes:

```
table(is.na(DFN)) # 17350
```

```
##  
## FALSE TRUE  
## 25448 17350
```

4. Registros/datos completos:

```
sum(complete.cases(DFN)) # 209
```

```
## [1] 209
```

5. Casos no completos:

```
sum(!complete.cases(DFN)) # 1829
```

```
## [1] 1829
```