

Solución ejercicios - Sesión 1

Capacitación en R

Capacitadores DET, en base a PE SCPE

Solución ejercicios - Sesión 1

Vamos a utilizar la base de datos `guaguas`, que conocimos en la primera sesión de este ciclo de capacitaciones. Recuerda que si ya la instalaste, con la función `install.packages("guaguas")`, no necesitas volver a hacerlo, solo deberás cargarla con `library(guaguas)`.

```
#install.packages("guaguas")  
library(guaguas)
```

```
## Warning: package 'guaguas' was built under R version 3.6.3
```

Cargamos los datos en un objeto que podamos ver en el entorno de trabajo.

```
data <- guaguas
```

Ejercicio 1

1- ¿Cuántas personas **con tu nombre** nacieron **el mismo año que tú**?

Vamos a utilizar el nombre **Ignacio** para responder a la pregunta. En primer lugar exploramos los datos, para conocer nombres y tipos de variables.

```
str(data)  
  
## tibble [84,565 x 5] (S3: tbl_df/tbl/data.frame)  
##   $ anio      : num [1:84565] 1920 1920 1920 1920 1920 1920 1920 1920 1920 1920 ...  
##   $ sexo      : chr [1:84565] "F" "M" "M" "M" ...  
##   $ nombre    : chr [1:84565] "María" "José" "Juan" "Luis" ...  
##   $ n         : int [1:84565] 2128 982 636 631 426 340 326 289 277 269 ...  
##   $ proporcion: num [1:84565] 0.1045 0.0482 0.0312 0.031 0.0209 ...
```

Para responder el ejercicio debemos usar las variables `anio` (numeric) y `nombre` (character), para construir un filtro que solo mantenga las observaciones que cumplen la condición. Adicionalmente, mantenemos solo algunas variables en el nuevo objeto `mi_nombre` que creamos (la variable `n` indica la cantidad de personas que nacieron en cada año según cada nombre).

```
mi_nombre = data[data$nombre == "Nicolás" &
                  data$anio == 1993, c("nombre", "anio", "n")]
mi_nombre
```

```
## # A tibble: 1 x 3
##   nombre   anio     n
##   <chr>   <dbl> <int>
## 1 Nicolás  1993   4180
```

Ejercicio 2

¿Cuál fue el **nombre más usado** el año que tú naciste?

Esto se puede resolver de diversas maneras. Mostraremos 2 formas, una creando objetos intermedios y otra resolviendo todo el ejercicio en solo una línea de código.

Debemos cargar la librería `tidyverse`, que instalamos durante la primera sesión.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
mas_usado <- data[data$anio==1993, c("nombre", "anio", "n")]
mas_usado <- arrange(mas_usado, desc(n))
```

O en una sola instrucción.

```
mas_usado <- arrange(data[data$anio==1993, c("nombre", "anio", "n")])
## Vemos las 10 primeras filas del data frame.
head(mas_usado, 10)
```

```
## # A tibble: 10 x 3
##   nombre   anio     n
##   <chr>   <dbl> <int>
## 1 María    1993   7330
## 2 Camila   1993   6045
## 3 José     1993   5143
```

```
## 4 Felipe      1993  5073
## 5 Daniela     1993  4613
## 6 Juan        1993  4522
## 7 Diego       1993  4352
## 8 Luis        1993  4224
## 9 Nicolás     1993  4180
## 10 Sebastián  1993  3942
```

En este caso al filtrar por año R entrega automáticamente ordenados los nombres según frecuencia, pero podría no ser así, o podríamos querer ordenarlos en sentido inverso.

Ejercicio 3 (Bonus)

3. ¿Cuál es el nombre más usado en Chile en todos los tiempos?

Este ejercicio es un *bonus* porque aún no contamos con todas las herramientas para realizar esta consulta a la base de datos. **Estos son contenidos de la Sesión 2.**

```
data %>%
  group_by(nombre) %>%
  summarise(total = sum(n)) %>%
  arrange(desc(total)) %>%
  head()
```

```
## # A tibble: 6 x 2
##   nombre total
##   <chr>   <int>
## 1 María  959108
## 2 José   586877
## 3 Juan   580185
## 4 Luis   505292
## 5 Carlos 293174
## 6 Jorge  238719
```

Como podrían imaginar, el nombre más popular en Chile en todos los tiempos es “María”.