

Ejercicio 2.5: Detección de duplicados

Silvia Pineda

Generación de datos

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.1     v stringr   1.6.0
v ggplot2   4.0.0     v tibble    3.3.0
v lubridate 1.9.4     v tidyr    1.3.1
v purrr    1.2.0

-- Conflicts -----
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
set.seed(123) # Para reproducibilidad
df <- data.frame(
  ID = c(1, 2, 3, 4, 5, 3, 4, 6, 7, 8, 2, 9, 10), # ID con duplicados
  Nombre = c("Ana", "Luis", "Carlos", "Marta", "Sofía", "Carlos", "Marta", "Jorge", "Elena",
  Edad = c(25, 30, 22, 28, 35, 22, 28, 40, 26, 32, 30, 45, 29),
  Ciudad = c("Madrid", "Barcelona", "Sevilla", "Madrid", "Bilbao", "Sevilla", "Madrid", "Val
)
```

Identifica los duplicados en toda las fila

```
df[duplicated(df), ]
```

```
ID Nombre Edad Ciudad
6  3 Carlos  22  Sevilla
7  4 Marta   28  Madrid
11 2 Luis    30  Barcelona
```

Identifica los IDs duplicados

```
df$ID[duplicated(df$ID)]
```

```
[1] 3 4 2
```

Elimina los duplicados manteniendo la primera observación que aparece

```
df_sin_duplicados<-distinct(df)
```