

LAB 28 (MD)

AnaGSanjuanM

2023-02-23

LABORATORIO 28

Tidy data: VALORES FALTANTES

Objetivo: Introducción práctica a los datos ordenados (o tidy data) y a las herramientas que provee el paquete tidy.

En este ejercicio vamos a:

1. Cargar datos (tibbles)
2. LLENAR LOS VALORES FALTANTES

Prerrequisitos:

Instalar paquete tidyverse

```
install.packages("tidyverse")
```

Cargar paquete tidyverse

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.0      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr   1.5.0
## ✓ ggplot2     3.4.1      ✓ tibble     3.1.8
## ✓ lubridate   1.9.2      ✓ tidyr      1.3.0
## ✓ purrr       1.0.1
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the `library(help="tidyverse")` to force all conflicts to become errors
```

Cargar datos

Es un tibble llamado acciones. Tiene tres columnas: anio, trimestre y retorno

```
acciones <- tibble(
  anio = c(2015, 2015, 2015, 2015, 2016, 2016, 2016),
  trimestre = c(1, 2, 3, 4, 2, 3, 4),
  retorno = c(1.88, 0.59, 0.35, NA, 0.92, 0.17, 2.66)
)
```

Visualizamos

```
acciones
```

```
## # A tibble: 7 × 3
##   anio trimestre retorno
##   <dbl>     <dbl>   <dbl>
## 1  2015         1     1.88
## 2  2015         2     0.59
## 3  2015         3     0.35
## 4  2015         4      NA
## 5  2016         2     0.92
## 6  2016         3     0.17
## 7  2016         4     2.66
```

DF1

Transformamos este table en un dataframe llamado DF1

```
DF1 = acciones
```

Visualizamos

```
DF1
```

```
## # A tibble: 7 × 3
##   anio trimestre retorno
##   <dbl>     <dbl>   <dbl>
## 1  2015         1     1.88
## 2  2015         2     0.59
## 3  2015         3     0.35
## 4  2015         4     NA
## 5  2016         2     0.92
## 6  2016         3     0.17
## 7  2016         4     2.66
```

Hay un dato faltante explícito: la tasa de retorno, en el cuarto trimestre del año 2015.

También hay un faltante implícito: no está el trimestre 1 del año 2016

DF2

Creamos dataframe DF2 que provenga del DF1

Activamos operador pipe %>% , para dispersar (spread) por un lado el anio y por otro el retorno

```
DF2 = DF1 %>%
  spread(anio, retorno)
```

Visualizamos

DF2

```
## # A tibble: 4 × 3
##   trimestre `2015` `2016`
##   <dbl>   <dbl>   <dbl>
## 1       1     1.88    NA
## 2       2     0.59    0.92
## 3       3     0.35    0.17
## 4       4     NA      2.66
```

Se muestra de manera explícita los datos que se tienen nulos

Se modificó la estructura de la tabla

DF3

Para eliminar los valores faltantes (explícitos e implícitos)

Creamos dataframe3 DF3 que proviene del table acciones

Activamos pipe %>% para generar un pivotaje a lo ancho

Los nombres los tomará de los años y los valores de provengan del retorno

Además, se añadirá otro filtro %>% para pivotear a lo largo (para que la regrese a su estado original), considerando que las columnas estén dadas por los años "2015" y "2016", además de considerar los nombres "anio", y los valores "retorno"

Se debe borrar(values_drop_na) los datos faltantes

```
DF3 = acciones %>%
  pivot_wider(names_from = anio, values_from = retorno) %>%
  pivot_longer(
    cols = c("2015", "2016"),
    names_to = "anio",
    values_to = "retorno",
    values_drop_na = TRUE
  )
```

Para visualizar

DF3

```
## # A tibble: 6 × 3
##   trimestre anio retorno
##   <dbl>   <chr>   <dbl>
## 1       1  2015     1.88
## 2       2  2015     0.59
## 3       2  2016     0.92
## 4       3  2015     0.35
## 5       3  2016     0.17
## 6       4  2016     2.66
```

DF4

Otra herramienta para hacer explícitos lo valores faltantes (conservando la estructura original)

Se crea dataframe 4 DF4 que proviene del DF1

Activamos operador pipe %>% (ctrl shift M)

Añadimos la función complete para anio y trimestre

```
DF4 = DF1 %>%  
  complete(anio, trimestre)
```

Visualizar

DF4

```
## # A tibble: 8 × 3  
##   anio trimestre retorno  
##   <dbl>      <dbl>   <dbl>  
## 1  2015         1     1.88  
## 2  2015         2     0.59  
## 3  2015         3     0.35  
## 4  2015         4     NA  
## 5  2016         1     NA  
## 6  2016         2     0.92  
## 7  2016         3     0.17  
## 8  2016         4     2.66
```

Se ha conservado la estructura original de DF1 pero se ha hecho explícita la ausencia de datos.