

Tidy data -datos ordenados-25

PRISCILA C. O.

2024-03-14

Hecho con gusto por Carla Carolina Pérez Hernández (UAEH)

LABORATORIO - Tidy data -datos ordenados- PARTE 1,2 y 3.

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

En este ejercicio vamos a: 1. Cargar datos (tibbles) 2. Convertir nuestros tibbles en dataframes (para su exportación) 3. Exportar dataframes originales 4. Pivotar tabla 4a 5. Exportar resultado (TABLA PIVOTANTE) 6. Separar y unir tablas

Prerrequisitos instalar paquete tidyverse install.packages("tidyverse") instalar paquete de datos install.packages("remotes") remotes::install_github("cienciadatos/datos") install.packages("datos") Cargar paquete tidyverse

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.3

## Warning: package 'readr' was built under R version 4.3.3

## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✔ dplyr      1.1.4      ✔ readr      2.1.5
## ✔ forcats    1.0.0      ✔ stringr    1.5.1
## ✔ ggplot2    3.5.0      ✔ tibble     3.2.1
## ✔ lubridate  1.9.3      ✔ tidyr      1.3.1
## ✔ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

argar paquete de datos

```
library(datos)

datos::tabla1

## # A tibble: 6 × 4
##   pais      anio  casos  poblacion
##   <chr>    <dbl> <dbl>    <dbl>
## 1 Afganistán 1999    745   19987071
## 2 Afganistán 2000   2666  20595360
## 3 Brasil     1999  37737  172006362
## 4 Brasil     2000  80488  174504898
## 5 China      1999 212258 1272915272
## 6 China      2000 213766 1280428583
```

```
datos::tabla2

## # A tibble: 12 × 4
##   pais      anio tipo      cuenta
##   <chr>    <dbl> <chr>    <dbl>
## 1 Afganistán 1999 casos      745
## 2 Afganistán 1999 población 19987071
## 3 Afganistán 2000 casos      2666
## 4 Afganistán 2000 población 20595360
## 5 Brasil     1999 casos      37737
## 6 Brasil     1999 población 172006362
## 7 Brasil     2000 casos      80488
## 8 Brasil     2000 población 174504898
## 9 China      1999 casos      212258
## 10 China     1999 población 1272915272
## 11 China     2000 casos      213766
## 12 China     2000 población 1280428583
```

```
datos::tabla3

## # A tibble: 6 × 3
##   pais      anio tasa
##   <chr>    <dbl> <chr>
## 1 Afganistán 1999 745/19987071
## 2 Afganistán 2000 2666/20595360
## 3 Brasil     1999 37737/172006362
## 4 Brasil     2000 80488/174504898
## 5 China      1999 212258/1272915272
## 6 China      2000 213766/1280428583
```

```
datos::tabla4a

## # A tibble: 3 × 3
##   pais      `1999` `2000`
##   <chr>    <dbl> <dbl>
## 1 Afganistán    745    2666
## 2 Brasil       37737  80488
## 3 China        212258 213766
```

```
datos::tabla4b

## # A tibble: 3 × 3
##   pais      `1999` `2000`
##   <chr>    <dbl> <dbl>
## 1 Afganistán 19987071  20595360
## 2 Brasil    172006362 174504898
## 3 China    1272915272 1280428583
```

un tibble es similar a un dataframe pero con limitantes ver datos como dataframe

```
df1<-data_frame(tabla1)

## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## i Please use `tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

df2<-data_frame(tabla2)
df3<-data_frame(tabla3)
df4a<-data_frame(tabla4a)
df4b<-data_frame(tabla4b)
```

los datos ordenados tienen 3 características esenciales, 1. cada variable debe tener su propia columna, 2. cada observación debe tener su propia fila y cada valor debe tener su propia celda exportar los dataframes originales(para guardarlos en la computadora)

```
write.csv(df1,file="df1.csv")
write.csv(df2,file="df2.csv")
write.csv(df3,file="df3.csv")
write.csv(df4a,file="df4a.csv")
write.csv(df4b,file="df4b.csv")
```

explicación de tibble

```
vignette("tibble")

## starting httpd help server ... done
```

Ordenar datos con la tabla4a (PIVOTAR)

t4a_PIVOTANTE=tabla4a %>% pivot_longer(cols=c("1999","2000"),names_to="anio",values_to="casos")

Exportar resultado: tabla ordenada

write.csv(t4a_PIVOTANTE,file="t4a_PIVOTANTE.csv")