

Ordenar dataframes

Jorge Loría

Sept 28, 2017

Datos Ordenados

Datos Ordenados

Todas las familias felices se parecen, todas las familias infelices son infelices a su manera - Leo Tolstoy
(parafraseado)

Datos Ordenados

*Todas las familias felices se parecen, todas las familias infelices son infelices a su manera - Leo Tolstoy
(parafraseado)*

Todos los conjuntos de datos ordenados se parecen, todos los conjuntos de datos desordenados son desordenados a su manera - Hadley Wickham

Ordenar dataframes

Para ordenar conjuntos

Principios de los datos ordenados

Principios de los datos ordenados

Los datos ordenados siguen “3” principios básicos:

- ▶ A cada variable le corresponde una columna

Principios de los datos ordenados

Los datos ordenados siguen “3” principios básicos:

- ▶ A cada variable le corresponde una columna
- ▶ A cada observación le corresponde una fila

Principios de los datos ordenados

Los datos ordenados siguen “3” principios básicos:

- ▶ A cada variable le corresponde una columna
- ▶ A cada observación le corresponde una fila
- ▶ A cada unidad de observación le corresponde un `data.frame`

Principios de los datos ordenados

Los datos ordenados siguen “3” principios básicos:

- ▶ A cada variable le corresponde una columna
- ▶ A cada observación le corresponde una fila
- ▶ A cada unidad de observación le corresponde un `data.frame`

Principios de los datos ordenados

Los datos ordenados siguen “3” principios básicos:

- ▶ A cada variable le corresponde una columna
- ▶ A cada observación le corresponde una fila
- ▶ A cada unidad de observación le corresponde un `data.frame`

Corolario de las primeras dos:

- ▶ A cada valor le corresponde una única celda

Principios de los datos ordenados

Los datos ordenados siguen “3” principios básicos:

- ▶ A cada variable le corresponde una columna
- ▶ A cada observación le corresponde una fila
- ▶ A cada unidad de observación le corresponde un `data.frame`

Corolario de las primeras dos:

- ▶ A cada valor le corresponde una única celda

Principios de los datos ordenados

Los datos ordenados siguen “3” principios básicos:

- ▶ A cada variable le corresponde una columna
- ▶ A cada observación le corresponde una fila
- ▶ A cada unidad de observación le corresponde un `data.frame`

Corolario de las primeras dos:

- ▶ A cada valor le corresponde una única celda

A pesar de que estas definiciones son bastante intuitivas y todos los datos que hemos manejado en el curso han sido *ordenados*, muchas veces se presentan desórdenes en los datos crudos.

Desórdenes *más* usuales

- ▶ Los nombres de las columnas son valores, no nombres de variables

Desórdenes *más* usuales

- ▶ Los nombres de las columnas son valores, no nombres de variables
- ▶ Varios valores están guardados en una misma columna

Desórdenes *más* usuales

- ▶ Los nombres de las columnas son valores, no nombres de variables
- ▶ Varios valores están guardados en una misma columna
- ▶ Hay variables que se guardan tanto en filas como en columnas

Desórdenes *más* usuales

- ▶ Los nombres de las columnas son valores, no nombres de variables
- ▶ Varios valores están guardados en una misma columna
- ▶ Hay variables que se guardan tanto en filas como en columnas
- ▶ Varias unidades de observación están en la misma tabla

Desórdenes *más* usuales

- ▶ Los nombres de las columnas son valores, no nombres de variables
- ▶ Varios valores están guardados en una misma columna
- ▶ Hay variables que se guardan tanto en filas como en columnas
- ▶ Varias unidades de observación están en la misma tabla
- ▶ Una unidad de observación está guardada en varias tablas

Funciones para ordenar:

Hay 4 funciones base para ordenar un dataframe, vienen en el paquete `tidyr`.

Funciones para ordenar:

Hay 4 funciones base para ordenar un dataframe, vienen en el paquete `tidyr`. Cargue este paquete usando la función `library`.

Funciones para ordenar:

Hay 4 funciones base para ordenar un dataframe, vienen en el paquete `tidyr`. Cargue este paquete usando la función `library`. Estas funciones son:

función	Operación
<code>spread</code>	pasa filas actuales a columnas
<code>gather</code>	pasa columnas a filas
<code>separate</code>	separa filas
<code>unite</code>	une filas
<code>complete</code>	completa los datos

spread

spread



Figure 1: Otro tipo de spread

spread

Esta función permite pasar de varias columnas que contienen un identificador en su nombre a columnas nuevas con una columna que le corresponda a los valores de los nombres de las columnas, y otro para los valores que toman en esas columnas.

gather

separate

unite

complete