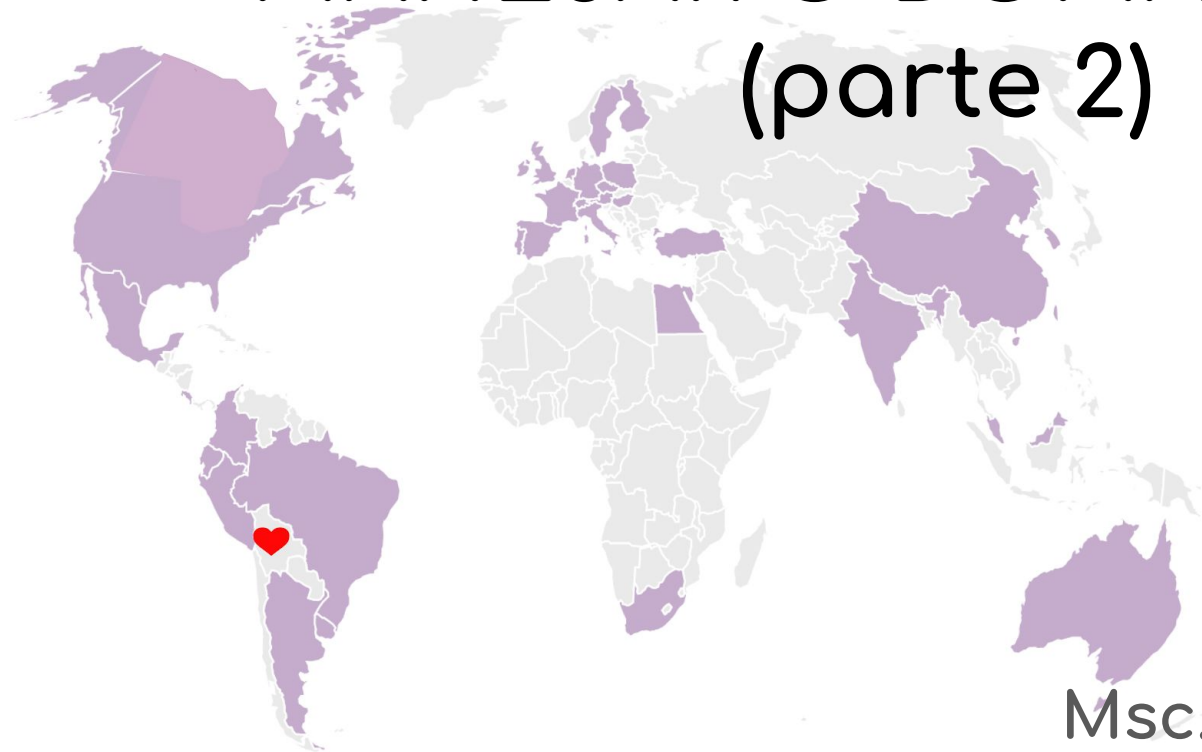


“R para Ciencia de Datos”

MANEJAR O DOMAR DATOS

(parte 2)



R-Ladies La Paz

Msc. Susana Yana Avila

Pre Requisitos

- Pre requisitos:
 - Leer la página Web de Introducción del libro:
<https://es.r4ds.hadley.nz/introducci%C3%B3n.html>
 - Instalar R-base
 - Instalar R-Studio
 - Instalar Tidyverse
 - Instalar paquetes de datos
 - Instalar otros paquetes
 - Instalar dplyr
 - Instalar tidyr
 - Instalar stringr
 - Instalar lubridate
 - [Todo lo que esté en la página Web de Introducción...](#)

Conjuntos de datos

tabla1			
pais	anio	casos	poblacion
Afganistán	1999	745	1998707
Afganistán	2000	2666	2059536
Brasil	1999	37737	17200636
Brasil	2000	80488	17450489
China	1999	212258	127291527
China	2000	213766	128042858

tabla2			
pais	anio	tipo	cuenta
Afganistán	1999	casos	745
Afganistán	1999	población	19987071
Afganistán	2000	casos	2666
Afganistán	2000	población	20595360
Brasil	1999	casos	37737
Brasil	1999	población	172006362
Brasil	2000	casos	80488
Brasil	2000	población	174504898
China	1999	casos	212258
China	1999	población	1272915272

Conjuntos de datos

tabla3		
pais	anio	tasa
Afganistán	1999	745/19987071
Afganistán	2000	2666/20595360
Brasil	1999	37737/172006362
Brasil	2000	80488/174504898
China	1999	212258/1272915272
China	2000	213766/1280428583

tabla4a		
pais	1999	2000
Afganistán	745	2666
Brasil	37737	80488
China	212258	213766

tabla4b		
pais	1999	2000
Afganistán	19987071	20595360
Brasil	172006362	174504898
China	1272915272	1280428583

Datos ordenados

pais	anio	casos	poblacion
Afganistán	1999	745	19987071
Afganistán	2000	2666	20595360
Brasil	1999	37737	172006362
Brasil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

variables

pais	anio	casos	poblacion
Afganistán	1999	745	19987071
Afganistán	2000	2666	20595360
Brasil	1999	37737	172006362
Brasil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

observaciones

pais	anio	casos	poblacion
Afganistán	1999	745	19987071
Afganistán	2000	2666	20595360
Brasil	1999	37737	172006362
Brasil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

valores

Pivotear -> largo y ordenado (pivot_longer)

pais	anio	casos
Afganistán	1999	745
Afganistán	2000	2666
Brasil	1999	37737
Brasil	2000	80488
China	1999	212258
China	2000	213766

pais	1999	2000
Afganistán	745	2666
Brasil	37737	80488
China	212258	213766

Tabla 4

Pivotear -> ancho y ordenado (pivot_wider)

pais	anio	tipo	casos		pais	anio	casos	poblacion
Afganistán	1999	casos	745		Afganistán	1999	745	19987071
Afganistán	1999	población	19987071		Afganistán	2000	2666	20595360
Afganistán	2000	casos	2666		Brasil	1999	37737	172006362
Afganistán	2000	población	20595360		Brasil	2000	80488	17504898
Brasil	1999	casos	37737		China	1999	212258	1272915272
Brasil	1999	población	172006362		China	2000	213766	1280428583
Brasil	2000	casos	80488					
Brasil	2000	población	174504898					
China	1999	casos	212258					
China	1999	población	1272915272					
China	2000	casos	213766					
China	2000	población	1280428583					

Separar > varias columnas y ordenada

pais	anio	tipo	casos
Afganistán	1999	casos	745
Afganistán	1999	población	19987071
Afganistán	2000	casos	2666
Afganistán	2000	población	20595360
Brasil	1999	casos	37737
Brasil	1999	población	172006362
Brasil	2000	casos	80488
Brasil	2000	población	174504898
China	1999	casos	212258
China	1999	población	1272915272
China	2000	casos	213766
China	2000	población	1280428583

pais	anio	casos	poblacion
Afganistán	1999	745	19987071
Afganistán	2000	2666	20595360
Brasil	1999	37737	172006362
Brasil	2000	80488	17504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Unir -> une columns y ordenada



pais	anio	tasa
Afganistán	1999	745 / 19987071
Afganistán	2000	2666 / 20595360
Brasil	1999	37737 / 172006362
Brasil	2000	80488 / 17504898
China	1999	212258 / 1272915272
China	2000	213766 / 1280428583

pais	siglo	casos	tasa
Afganistán	19	99	745 / 19987071
Afganistán	20	0	2666 / 20595360
Brasil	19	99	37737 / 172006362
Brasil	20	0	80488 / 17504898
China	19	99	212258 / 1272915272
China	20	0	213766 / 1280428583

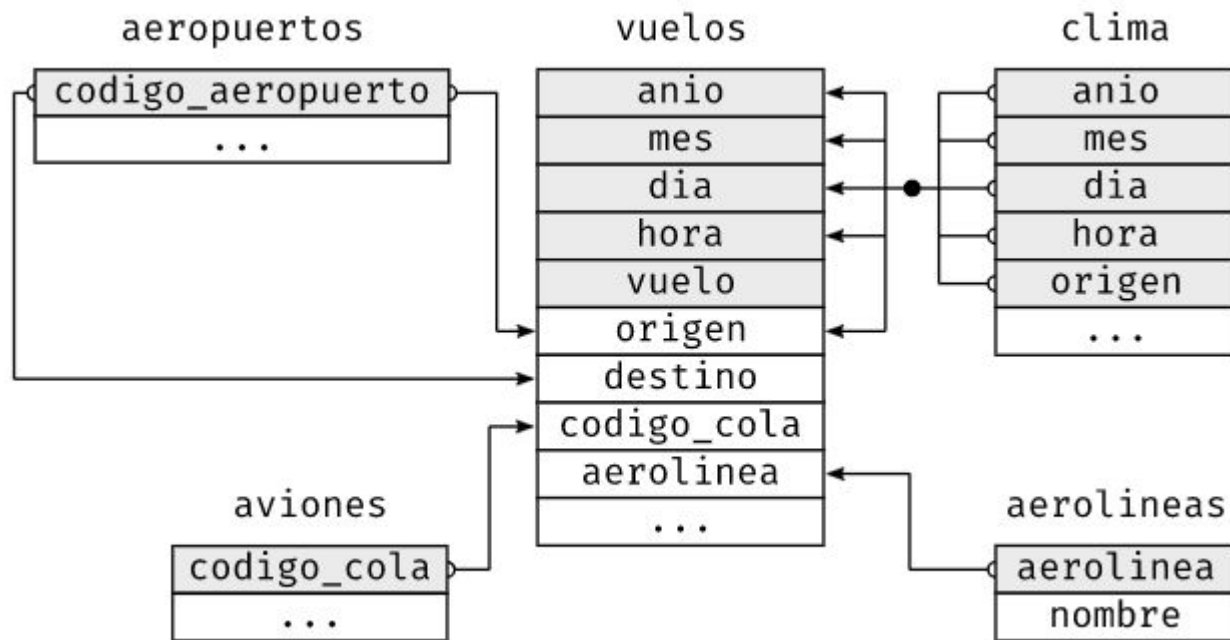
Tabla 6

Valores faltantes

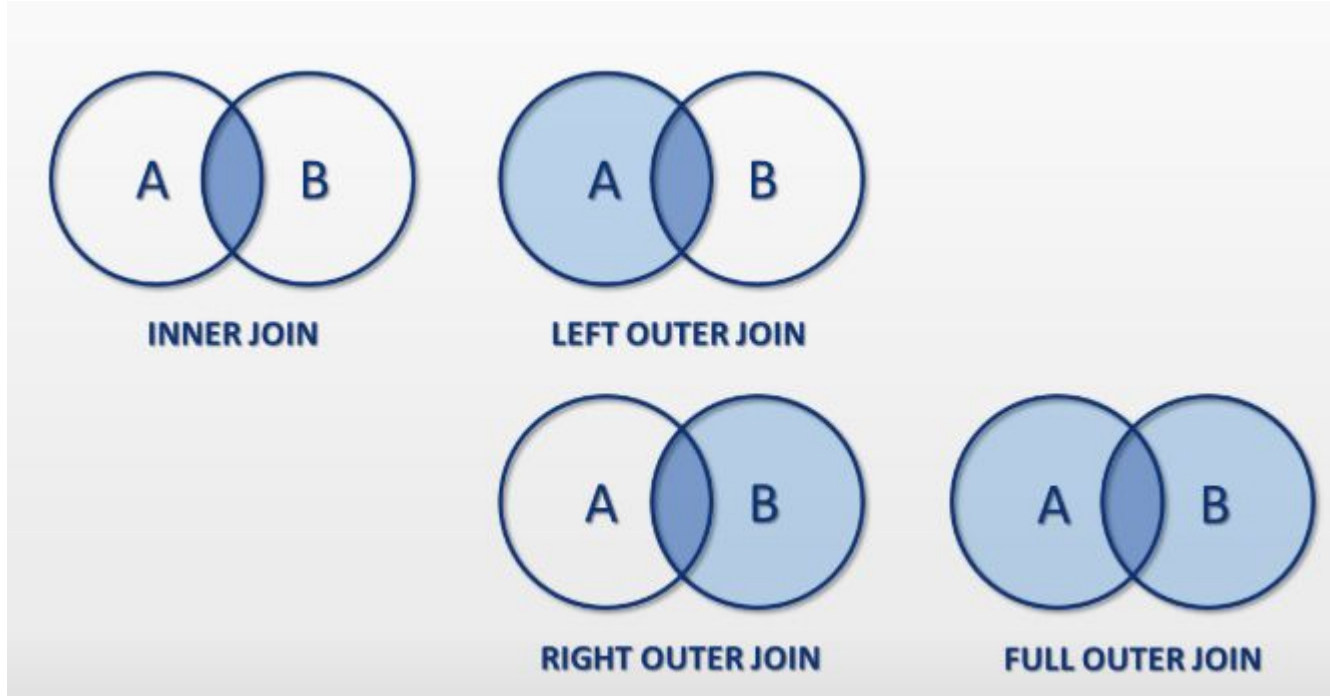
Explícita, esto es, aparece como NA.

Implícita, esto es, simplemente no aparece en los datos.

Datos Relacionales



Datos Relacionales



Equivalentes de dplyr - SQL

dplyr

SQL

```
inner_join(x, y, by = "z")
```

```
SELECT * FROM x INNER JOIN y USING (z)
```

```
left_join(x, y, by = "z")
```

```
SELECT * FROM x LEFT OUTER JOIN y USING (z)
```

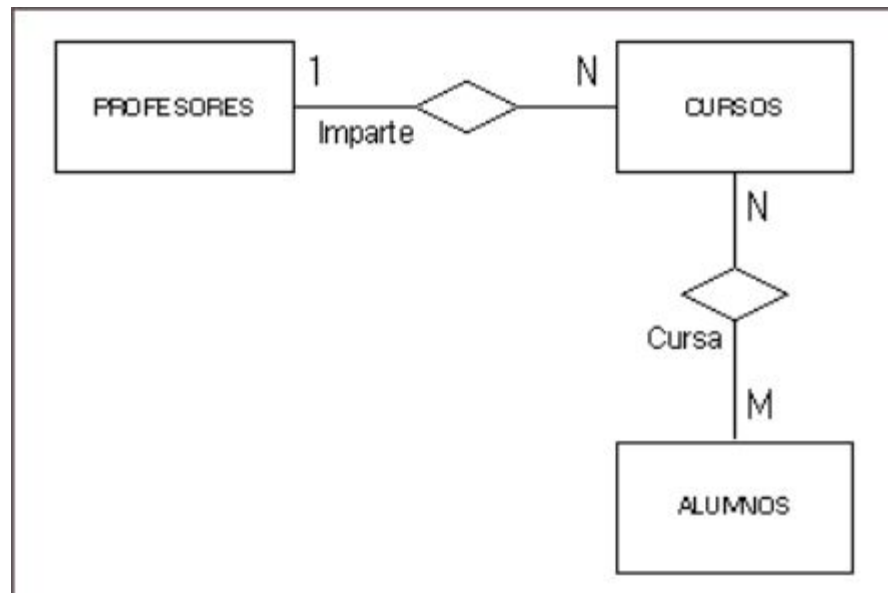
```
right_join(x, y, by = "z")
```

```
SELECT * FROM x RIGHT OUTER JOIN y USING (z)
```

```
full_join(x, y, by = "z")
```

```
SELECT * FROM x FULL OUTER JOIN y USING (z)
```

Tipos de Relaciones



Laboratorio