

MANEJO DE DATOS 2

IMPORTAR DATOS

```
songs <- read_csv("data/songs.csv")  
songs
```

```
## # A tibble: 4 x 2  
##   song                name  
##   <chr>              <chr>  
## 1 Across the Universe John  
## 2 Come Together      John  
## 3 Hello, Goodbye      Paul  
## 4 Peggy Sue          Buddy
```

```
artists <- read_csv("data/artists.csv")  
artists
```

```
## # A tibble: 4 x 2  
##   name    plays  
##   <chr>  <chr>  
## 1 George sitar  
## 2 John   guitar  
## 3 Paul   bass  
## 4 Ringo  drums
```

left_join JUNTAR TABLAS

songs		artists					
song	name				song	name	plays
Across the Universe	John	+	George	=	Across the Universe	John	guitar
Come Together	John		John		Come Together	John	guitar
Hello, Goodbye	Paul		Paul		Hello, Goodbye	Paul	bass
Peggy Sue	Buddy		Ringo		Peggy Sue	Buddy	<NA>

```
left_join(songs, artists, by = "name")
```

left_join CÓDIGO

```
left_join(songs, artists, by = "name")
```

```
## # A tibble: 4 x 3
##   song                name plays
##   <chr>              <chr> <chr>
## 1 Across the Universe John  guitar
## 2 Come Together      John  guitar
## 3 Hello, Goodbye      Paul   bass
## 4 Peggy Sue           Buddy <NA>
```

IMPORTAR DATOS 2

```
library(tidyverse)
data("table4a")
table4a
```

```
## # A tibble: 3 x 3
##   country    `1999` `2000`
## * <chr>      <int> <int>
## 1 Afghanistan    745   2666
## 2 Brazil        37737  80488
## 3 China         212258 213766
```

```
data("table2")
table2
```

```
## # A tibble: 12 x 4
##   country      year type      count
##   <chr>      <int> <chr>    <int>
## 1 Afghanistan 1999 cases      745
## 2 Afghanistan 1999 population 19987071
## 3 Afghanistan 2000 cases      2666
## 4 Afghanistan 2000 population 20595360
## 5 Brazil      1999 cases      37737
## 6 Brazil      1999 population 172006362
## 7 Brazil      2000 cases      80488
## 8 Brazil      2000 population 174504898
## 9 China       1999 cases     212258
## 10 China       1999 population 1272915272
## 11 China       2000 cases      213766
## 12 China       2000 population 1280428583
```

```
cases <- read_csv("data/cases.csv")
cases
```

```
## # A tibble: 3 x 4
##   country `2011` `2012` `2013`
##   <chr>    <dbl> <dbl> <dbl>
## 1 FR      7000    6900    7000
## 2 DE      5800    6000    6200
## 3 US     15000   14000   13000
```

¿Como obtendríamos el promedio por país?

¿COMO OBTENDRÍAMOS EL PROMEDIO POR PAÍS?

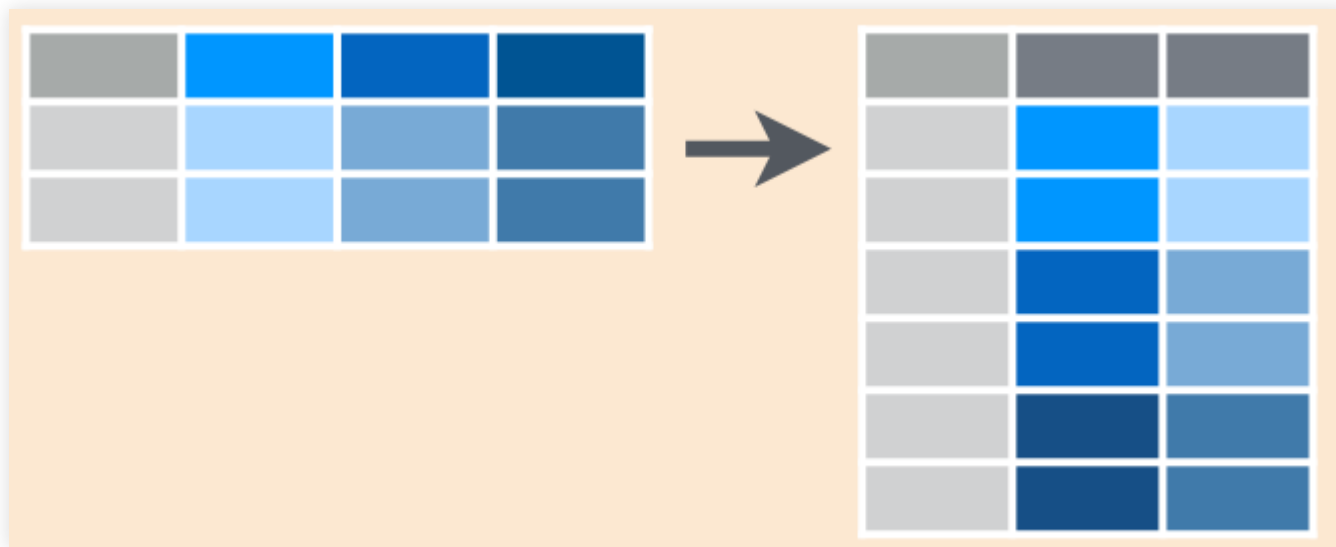
```
cases %>%  
  mutate(promedio = (`2011` + `2012` + `2013`)/3)
```

```
## # A tibble: 3 x 5  
##   country `2011` `2012` `2013` promedio  
##   <chr>    <dbl> <dbl> <dbl>    <dbl>  
## 1 FR      7000    6900    7000    6967.  
## 2 DE      5800    6000    6200    6000  
## 3 US     15000   14000   13000   14000
```

Ahora pensemos en una tabla más grande, con más años

gather RECOLECTAR

Coloca nombres de columnas en una variable (columna) **key**, recolectando los valores (**value**) de las columnas en un **sola** columna



gather EJEMPLO

table4a

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

key value

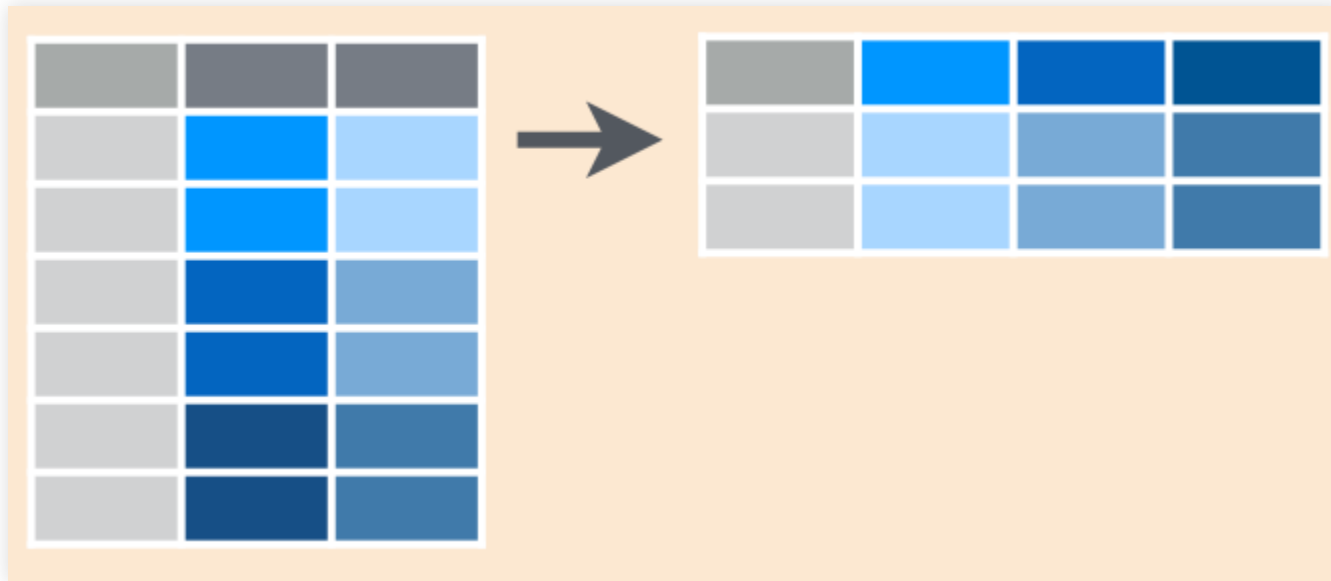
gather CÓDIGO

```
gather(table4a, `1999`, `2000`, key = "year", value = "cases")
```

```
## # A tibble: 6 x 3
##   country    year  cases
##   <chr>      <chr> <int>
## 1 Afghanistan 1999     745
## 2 Brazil      1999   37737
## 3 China       1999  212258
## 4 Afghanistan 2000     2666
## 5 Brazil      2000   80488
## 6 China       2000  213766
```

spread ESPARCIR

Esparece un par de columnas (2, key-value) en multiples columnas



spread EJEMPLO

table2

country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T



country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172M
B	2000	80K	174M
C	1999	212K	1T
C	2000	213K	1T

key value

spread CÓDIGO

```
spread(table2, type, count)
```

```
## # A tibble: 6 x 4
##   country      year  cases population
##   <chr>      <int> <int>      <int>
## 1 Afghanistan 1999     745   19987071
## 2 Afghanistan 2000    2666   20595360
## 3 Brazil      1999   37737   172006362
## 4 Brazil      2000   80488   174504898
## 5 China       1999  212258  1272915272
## 6 China       2000  213766  1280428583
```

EJERCICIOS

Revisemos el archivo `03-script-manipulacion-datos-02.R`