

Práctica 3: Parte 2 (Tidyverse: paquete tidyr)

Tabla de contenidos

1	Uso del paquete “tidyr”	1
1.1	Formato largo-formato ancho	1
1.2	Formato ancho a formato largo o combinación: pivot_longer o gather	2
1.3	Formato largo a formato ancho o extensión: pivot_wider o spread	3
1.4	Algunas curiosidades del paquete tidyr: separate() y unite()	3

1 Uso del paquete “tidyr”

1.1 Formato largo-formato ancho

Esta configuración de miniherramientas que usa el paquete **dplyr** invita a otros paquetes a extenderlo. Uno de tales paquetes es el paquete **tidyr** que forma parte del sistema “tidyverse”.

En general, el sistema **tidyverse** supone que cada fila es una observación, y cada columna es una variable. A esta disposición se la conoce como **FORMATO LARGO**.

En el siguiente ejemplo se presenta una situación diferente que aparece habitualmente. Hay una variable, lluvia o cantidad de precipitación, que se extiende sobre tres columnas (“lluvia_estacion01” a “lluvia_estacion03”). Esta disposición de los datos se conoce como **FORMATO ANCHO**.

```
set.seed(24)
dat = data.frame(temper = runif(3, 15, 25),
                 lluvia_estacion01 = runif(3, 1, 3),
                 lluvia_estacion02 = runif(3, 1, 3),
                 lluvia_estacion03 = runif(3, 1, 3))
dat
```

temper	lluvia_estacion01	lluvia_estacion02	lluvia_estacion03
17.92574	2.037794	1.559471	1.509450
17.24891	2.325239	2.527641	2.209778
22.04223	2.840888	2.603261	1.741470

Como se ha comentado anteriormente, el sistema tidyverse espera que los datos se encuentren en “formato largo”. En algunas situaciones se necesita transformar los datos de “formato ancho” a “formato largo”. Es decir, en estos datos se necesita reunir todos los valores de precipitaciones en una única columna, y añadir una columna de identificación adicional que especifique la estación a la que pertenece.

1.2 Formato ancho a formato largo o combinación: `pivot_longer` o `gather`

La **operación de combinación** que convierte datos **de formato ancho a formato largo** puede realizarse usando la función `pivot_longer()` o también con `gather()`.

```
pivot_longer(
  data,
  cols,
  names_to = "name",
  values_to = "value",
  # más argumentos
  ...
)
#gather(data, key, value, ...)
```

donde:

- `data`, es el `data.frame` de entrada.
- `names_to` (`key`), el nombre de la columna con la identificación en el `data.frame` resultante.
- `values_to` (`value`), el nombre de la columna con los valores (precipitacion en nuestro ejemplo) en el `data.frame` resultante.
- `cols` (...), especificación de qué columnas deberían ser reunidas/combinadas. Se pueden usar nombres de columnas con un menos delante para excluir esas columnas de la combinación.

El siguiente código llama a `pivot_longer()` para transformar los datos del ejemplo anterior (aparece comentada el modo equivalente de hacerlo con la función: `gather()`).

```
library(tidyr)
dat_nuevos = dat %>%
  pivot_longer(cols = -temper,
               names_to = "id_estacion",
               values_to = "precipitacion")
# dat_nuevos = gather(dat, key = id_estacion, value = precipitacion, -temper)
head(dat_nuevos)
```

temper	id_estacion	precipitacion
17.92574	lluvia_estacion01	2.037794
17.92574	lluvia_estacion02	1.559471
17.92574	lluvia_estacion03	1.509450
17.24891	lluvia_estacion01	2.325239
17.24891	lluvia_estacion02	2.527641
17.24891	lluvia_estacion03	2.209778

1.3 Formato largo a formato ancho o extensión: pivot_wider o spread

La **operación contraria de extensión** se realiza usando la función `pivot_wider()` o la función `spread()`, las cuales convierten de **formato largo a formato ancho**. En el siguiente ejemplo, se transforman a formato ancho los datos en formato largo creados en el apartado anterior: “dat_nuevos”.

```
pivot_wider(dat_nuevos,
            names_from = "id_estacion",
            values_from = "precipitacion")
```

temper	lluvia_estacion01	lluvia_estacion02	lluvia_estacion03
17.92574	2.037794	1.559471	1.509450
17.24891	2.325239	2.527641	2.209778
22.04223	2.840888	2.603261	1.741470

```
#spread(dat_nuevos, key = id_estacion, value = precipitacion)
```

1.4 Algunas curiosidades del paquete tidyr: separate() y unite()

La función `separate()` permite obtener nuevas columnas a partir de una columna de partida, al indicar una cadena de texto como separador. Por ejemplo, tenemos una columna fecha y queremos separarla en 3 columnas nuevas: día, mes y año.

Su complementario es `unite()`, que construye una nueva columna a partir de unir varias columnas con un separador que indiquemos.

Ejemplo. En la columna “Provincia” del conjunto de datos de población del censo de 2001 empleado anteriormente, aparece el código junto al nombre de la provincia separado por un guión.

```
load("datosPobEspCenso2001.RData", verbose = TRUE)
```

Loading objects:
datos

```
head(datos)
```

Provincia	CCAA	TOTAL	Varon	Mujer
01-Álava	País Vasco	286387	142036	144351
02-Albacete	Castilla-La Mancha	364835	181461	183374
03-Alicante/Alacant	Comunidad Valenciana	1461925	722162	739763
04-Almería	Andalucía	536731	272023	264708
33-Asturias	Asturias (Principado de)	1062998	508995	554003
05-Ávila	Castilla y León	163442	81850	81592

Para separarlo en dos columnas, se haría del siguiente modo:

```
datos_sep = datos %>%
  separate(col = Provincia, into = c("Codigo", "Provincia"), sep="-")
head(datos_sep)
```

Codigo	Provincia	CCAA	TOTAL	Varon	Mujer
01	Álava	País Vasco	286387	142036	144351
02	Albacete	Castilla-La Mancha	364835	181461	183374
03	Alicante/Alacant	Comunidad Valenciana	1461925	722162	739763
04	Almería	Andalucía	536731	272023	264708
33	Asturias	Asturias (Principado de)	1062998	508995	554003
05	Ávila	Castilla y León	163442	81850	81592

Por defecto, elimina la columna de partida, pero esto podría cambiarse (consultar la sintaxis de la función).

Ahora podría ordenarse por código o por el nombre de la provincia.

Ejemplo. Con `unite()` podríamos volver a unirla, pero por ejemplo, poniendo el código detrás, del siguiente modo:

```
datos_sep %>%  
  unite(col="Provincia",sep="-", Provincia,Codigo) %>%  
  head()
```

Provincia	CCAA	TOTAL	Varon	Mujer
Álava-01	País Vasco	286387	142036	144351
Albacete-02	Castilla-La Mancha	364835	181461	183374
Alicante/Alacant-03	Comunidad Valenciana	1461925	722162	739763
Almería-04	Andalucía	536731	272023	264708
Asturias-33	Asturias (Principado de)	1062998	508995	554003
Ávila-05	Castilla y León	163442	81850	81592