

María Paula Llano

Departamento de Ciencias de la
Atmósfera y los Océanos

Laboratorio de Procesamiento de Información Meteorológica/Oceanográfica

data.frame - extra

EXTRAS

Formato ancho (wide), para cada variable hay una columna.

Formato largo (long), cada fila es una combinación única de la variable de identificación.

El **formato largo** es el más conveniente para filtrar y realizar algunos tipos de agregaciones, mientras que el **formato ancho** es típico de datos realizados en el tiempo.

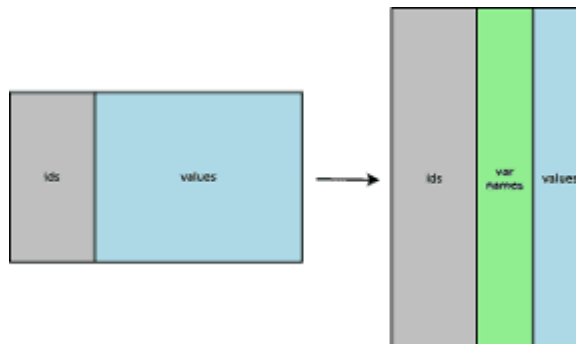
wide

vs

long

ID	a1	a2	a3

ID	ID2	A
1	a1	
2	a1	
3	a1	
1	a2	
2	a2	
3	a2	
1	a3	
2	a3	
3	a3	



library(reshape)

original<-airquality

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

s_2023.R x clase_dataframe_MP_2023.R x original x ejemplo_dataframe_2023.R x

Filter

	Ozone	Solar.R	Wind	Temp	Month	Day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10
11	7	NA	6.9	74	5	11
12	16	256	9.7	69	5	12
13	11	290	9.2	66	5	13

Showing 1 to 13 of 153 entries, 6 total columns

Console Terminal x Background Jobs x

En “original” guarde la info
Tengo 153 observaciones en 6 variables

Project: (None)

Environment History Connections Tutorial

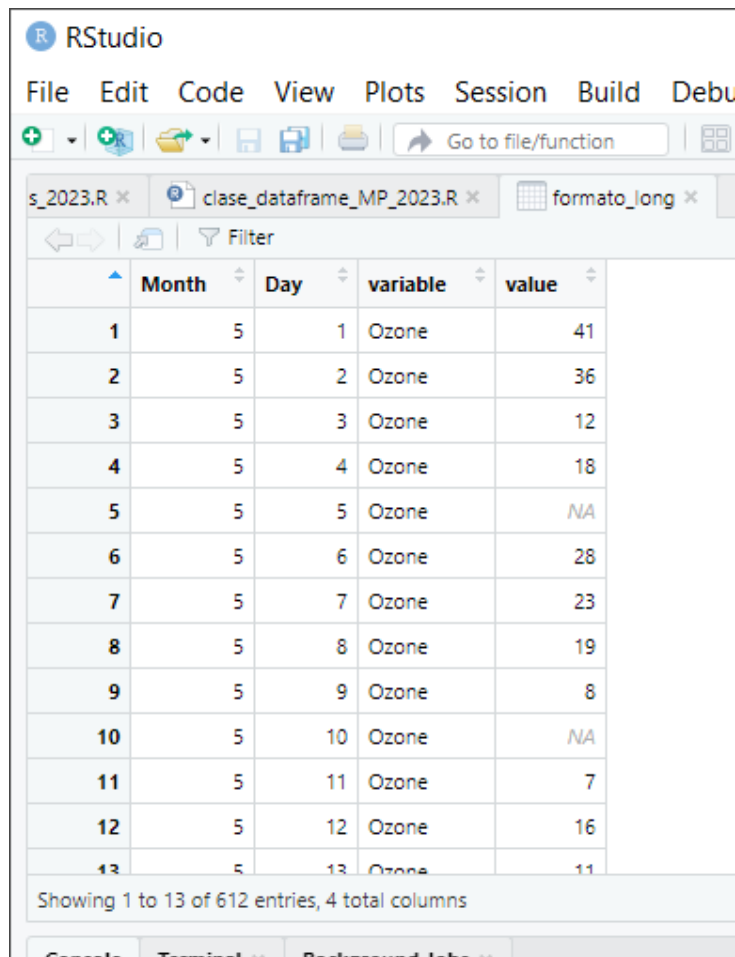
Import Dataset 136 MiB List

R Global Environment

Data

formato_long	612 obs. of 4 variables
formato_lon...	765 obs. of 3 variables
formato_lon...	765 obs. of 3 variables
original	153 obs. of 6 variables

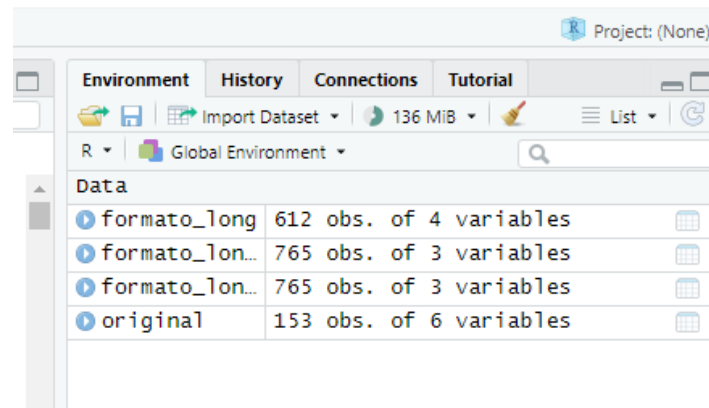
```
formato_long<-melt(airquality,id=c("Month","Day"))
```



RStudio interface showing the 'formato_long' data frame. The table has 4 columns: Month, Day, variable, and value. The first 13 rows are displayed, showing Ozone levels for each day of the month.

	Month	Day	variable	value
1	5	1	Ozone	41
2	5	2	Ozone	36
3	5	3	Ozone	12
4	5	4	Ozone	18
5	5	5	Ozone	NA
6	5	6	Ozone	28
7	5	7	Ozone	23
8	5	8	Ozone	19
9	5	9	Ozone	8
10	5	10	Ozone	NA
11	5	11	Ozone	7
12	5	12	Ozone	16
13	5	13	Ozone	11

Showing 1 to 13 of 612 entries, 4 total columns



RStudio Environment pane showing the 'formato_long' data frame. The table lists the data frame and its dimensions.

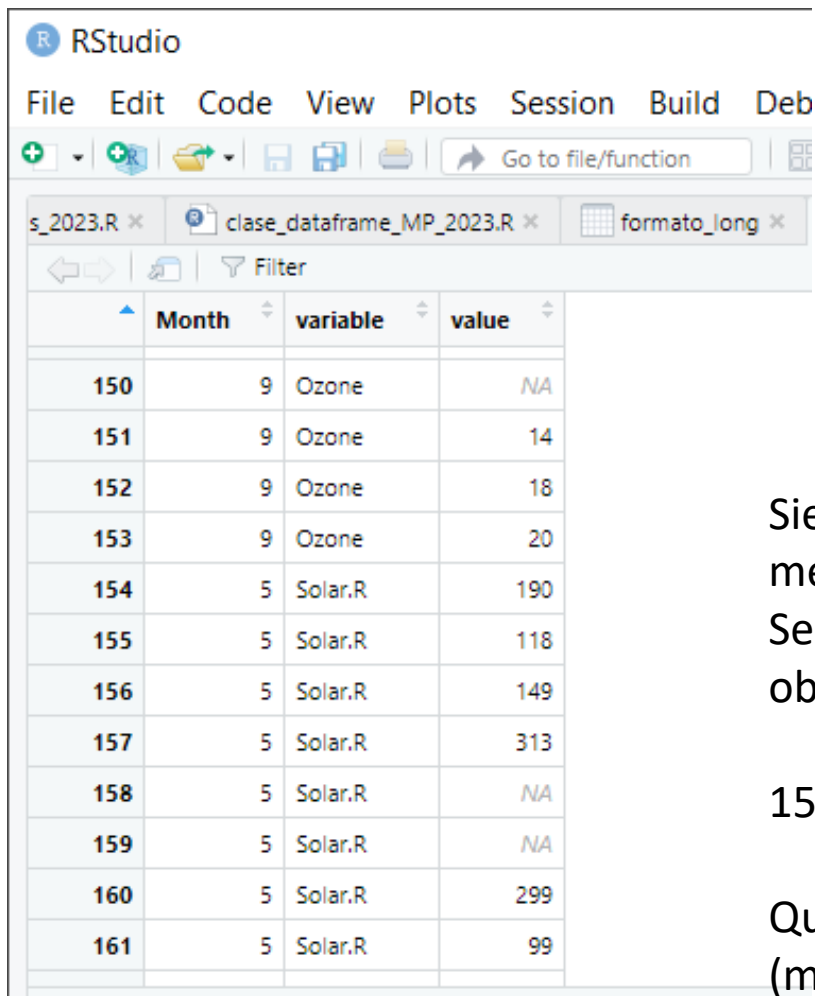
formato_long	612 obs. of 4 variables
formato_lon...	765 obs. of 3 variables
formato_lon...	765 obs. of 3 variables
original	153 obs. of 6 variables

Siendo la variable de identificación (ID) el mes y el día.

Se encolumna todo quedando 612 observaciones de 4 variables

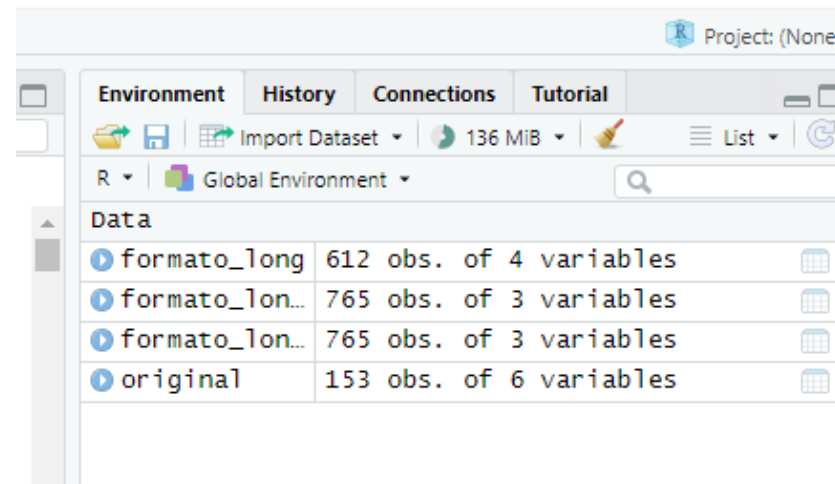
$$153 * 4 = 612$$

```
formato_long_2<-melt(airquality,id=c("Month"))
```



RStudio interface showing the 'formato_long' data frame. The data is displayed in a table with columns: Month, variable, and value. The table shows rows 150 to 161, with 'Month' values 9 and 5, and 'variable' values 'Ozone' and 'Solar.R'.

	Month	variable	value
150	9	Ozone	NA
151	9	Ozone	14
152	9	Ozone	18
153	9	Ozone	20
154	5	Solar.R	190
155	5	Solar.R	118
156	5	Solar.R	149
157	5	Solar.R	313
158	5	Solar.R	NA
159	5	Solar.R	NA
160	5	Solar.R	299
161	5	Solar.R	99



RStudio Environment pane showing the 'Data' section. It lists the following objects:

Object	Size
formato_long	612 obs. of 4 variables
formato_lon...	765 obs. of 3 variables
formato_lon...	765 obs. of 3 variables
original	153 obs. of 6 variables

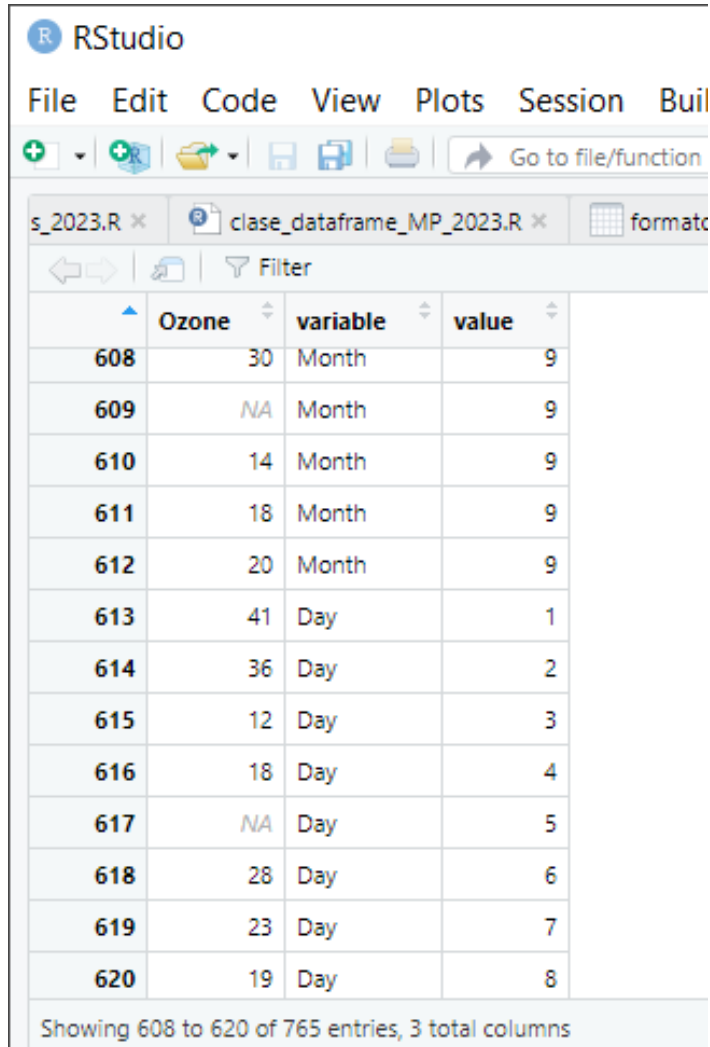
Siendo la variable de identificación (ID) el mes.

Se encolumna todo quedando 765 observaciones de 3 variables

$$153 * 5 = 765$$

Quedando también como variable los días (miren al final del nuevo dato frame)

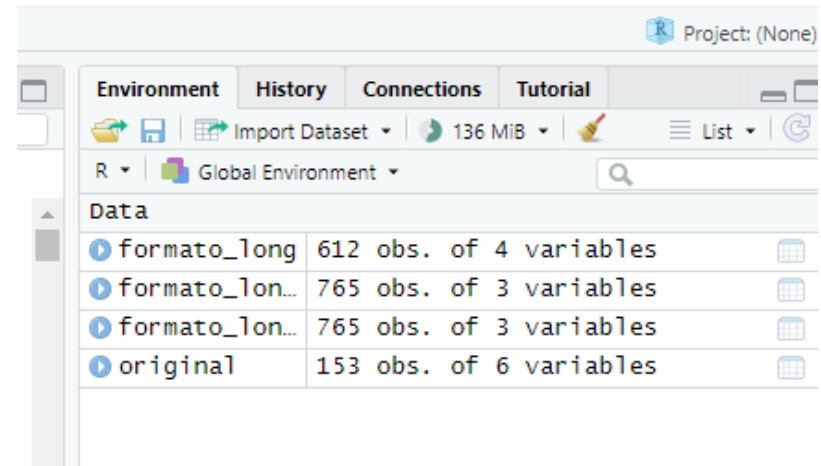
```
formato_long_3<-melt(airquality,id=c("Ozone"))
```



RStudio interface showing the `airquality` dataset being melted into a long format. The `formato_long_3` object is displayed in the Environment pane, showing 765 observations of 3 variables: `Ozone`, `variable`, and `value`.

	Ozone	variable	value
608	30	Month	9
609	NA	Month	9
610	14	Month	9
611	18	Month	9
612	20	Month	9
613	41	Day	1
614	36	Day	2
615	12	Day	3
616	18	Day	4
617	NA	Day	5
618	28	Day	6
619	23	Day	7
620	19	Day	8

Showing 608 to 620 of 765 entries, 3 total columns



RStudio Environment pane showing the `formato_long_3` object. The Data section displays the following information:

Object	Observations	Variables
<code>formato_long</code>	612 obs.	4 variables
<code>formato_lon...</code>	765 obs.	3 variables
<code>formato_lon...</code>	765 obs.	3 variables
<code>original</code>	153 obs.	6 variables

Siendo la variable de identificación (ID) el ozono.

Se encolumna todo quedando 765 observaciones de 4 variables

$$153 \times 4 = 765$$

Quedando también como variable los días y los meses (se puede ver en esta captura del nuevo dato frame)