

Curso R / Rstudio 2022

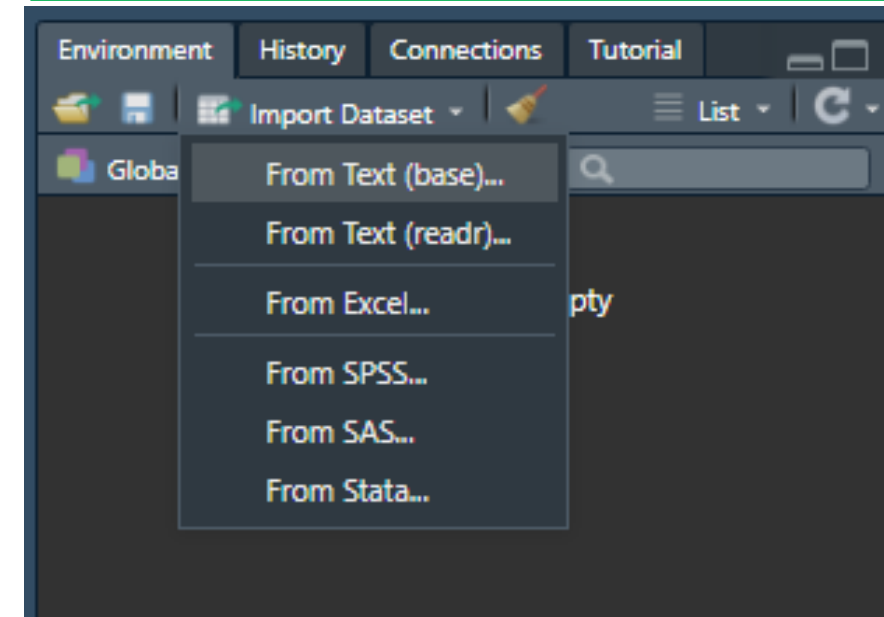
MANEJO DE DATOS

Importar y Exportar datos

- Base R tiene la función `read.csv()` para importar archivos de tipo csv
- Hay paqueterías que te permiten descargar datos desde:
 - Excel
 - dta (Stata)
 - SPSS, SAS
 - Fuentes oficiales (FRED, Banxico, ...)
- Para exportar está la función `write.csv()` en Base R
 - Podemos descargar paqueterías para exportar en otros formatos

➤ Vamos al código “Importar_exportar.R”

La herramienta más directa para importar



read_excel()

- Esta es la función más usada (existen varias) para importar datos desde excel
- Por default, la base se importa en una estructura llamada tibble data frame

➤ Vamos al código “readxl.R”

Stata user trying to do a simple data cleaning task in R
be like



Tidyverse



- [Tidyverse](#) es una colección de paquetes disponibles en R y orientados a la manipulación, importación, exploración y visualización de datos y que se utiliza exhaustivamente en ciencia de datos.
- Usaremos:
 - dplyr
 - tidyr
- Mas adelante estudiaremos ggplot

Nota: Pipes

dplyr introduce a R los pipes `%>%` para enfatizar la secuencia de funciones (o acciones). Hacen al código más fácil de leer.

- $f(g(h(x)))$ se cambia por `x %>% h() %>% g() %>% f()`
- `mean(rnorm(10))` se escribe mejor como `rnorm(10) %>% mean()`

Tamaño del data frame

- Quitar una variable
- Subset de variables (`dplyr::select()`)
- Agregar variables (`dplyr::mutate()`)
- Agregar filas
- Combinar dos data frames (`dplyr::joins`)

➤ Vamos al código “manejo de dataframes.R”

➤ Vamos al código “plots base.R”

tidyr::pivot_longer()

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

➤ Vamos al código “wide_long.R”

tidyr::pivot_wider()

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

Nota: antes estas funciones se llamaban gather() y spread()

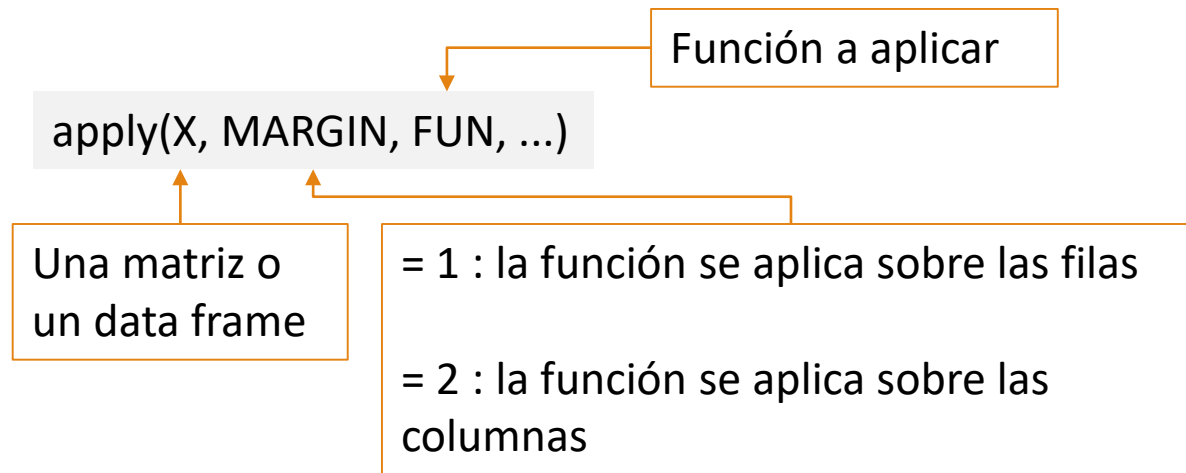
dplyr::filter() / dplyr::summarise() / count()

- filter(): sirve para quedarse con las observaciones que cumplen con cierta característica
 - Habíamos visto: `países[países$continente == "SA",]`
 - Es lo mismo usar: `filter(países, continente == "SA")`
- summarise(): reduce las observaciones de un dataframe en un solo estadístico
 - └ En Stata esto es lo mismo que el comando collapse
- count(): devuelve la frecuencia en que aparecen observaciones de un vector

➤ Vamos al código "filter_summarise_count.R"

Programación funcional: apply()

- In computer science, **functional programming** is a programming paradigm where programs are constructed by applying and composing functions (Wikipedia).



➤ Vamos al código “apply.R”

