



Introducción al Tidy Data

José Luis Texcalac Sangrador

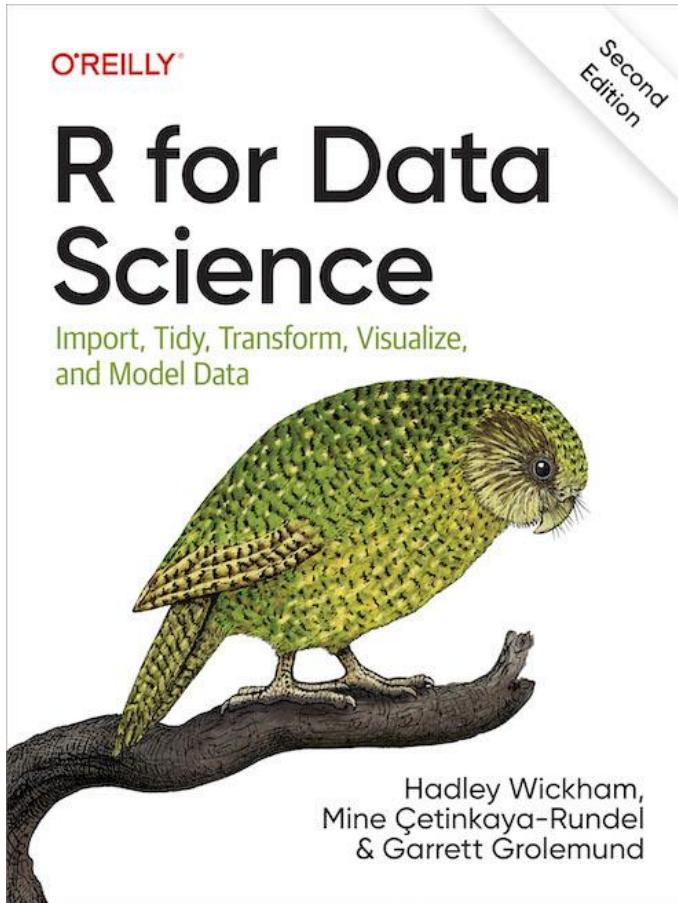
Procesamiento y visualización de datos espaciales en R



Tidyverse

- Actualmente en **R** predominan tres entornos de trabajo, con sus respectivas librerías, que nos permiten programar y manipular la información.
- Cada una tiene sus propias ventajas y desventajas, así como la forma de abordar la manipulación de datos.
- Para este curso, adoptaremos la filosofía del **Tidyverse**.





Tidyverse es una colección de paquetes **R** diseñados para la ciencia de datos.

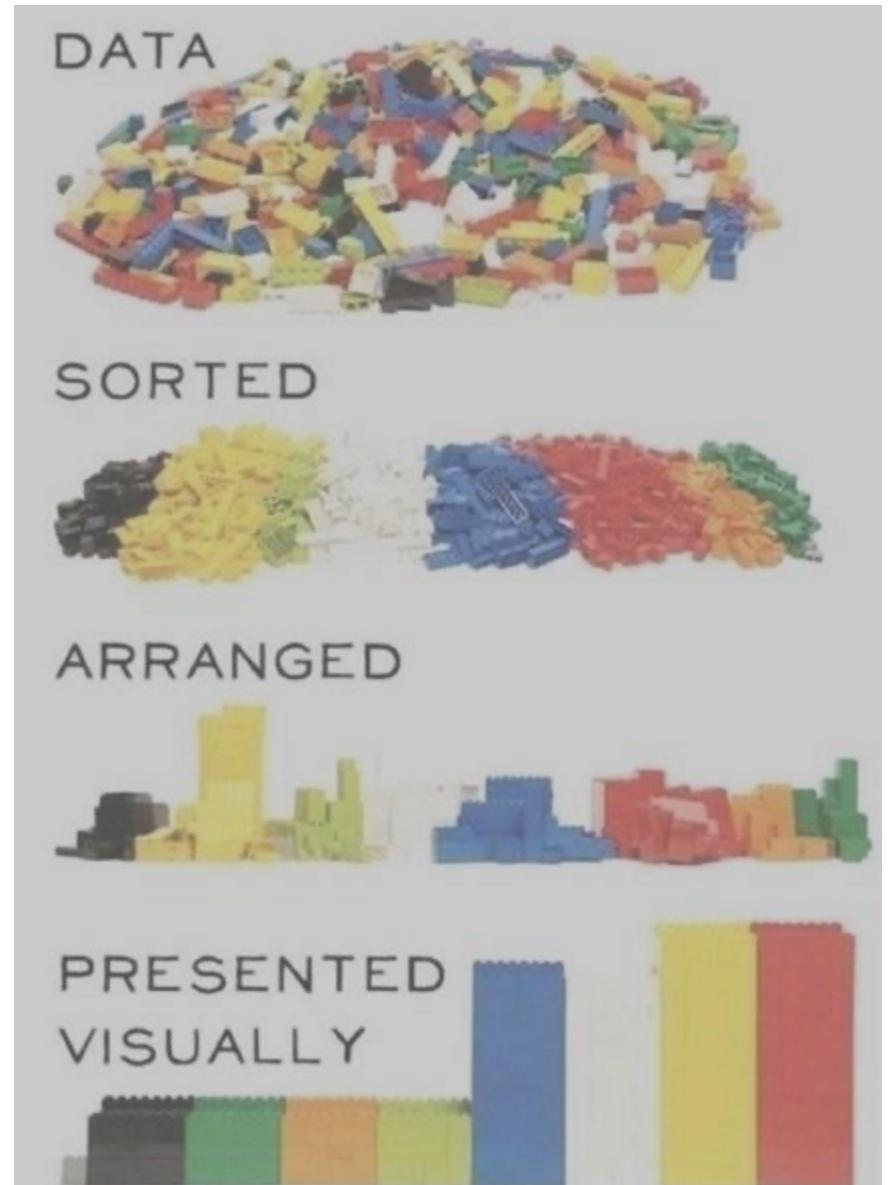
```
> install.packages("tidyverse")
```





Tidy data

Forma estandarizada de vincular la estructura de un conjunto de datos (su disposición física), con su semántica (su significado).





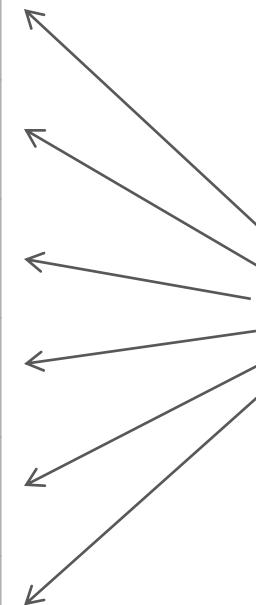
tidy data

Cada celda es un valor e indica la intersección entre una variable y una observación



Cada columna es una variable

nombre	edad	peso
Kevin	17	64.3
Brayan	16	61.7
Kimberly	15	51.9
Britany	16	59.3
Brandon	17	69.1
Melany	16	61.6



Cada fila es una observación



messy data

Los encabezados son
valores, no atributos



nombre	edad	peso	lacio	rizado	ondulado
Kevin	17	64.3	0	1	0
Brayan	16	61.7	0	0	0
Kimberly	15	51.9	1	0	0
Britany	16	59.3	0	0	0
Brandon	17	69.1	0	0	1
Melany	16	61.6	0	0	0



Datos wide y
long

wide

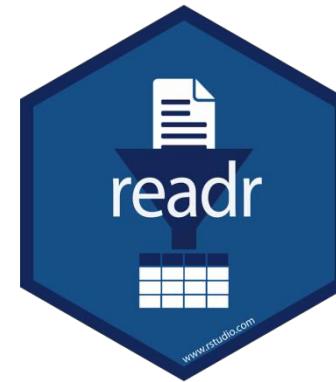
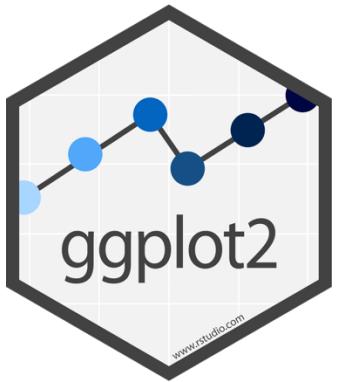
Site	2013	2014	2015
CAM	51.0	42.8	39.9
FAC	48.3	39.0	36.6
IZT	44.6	39.3	35.0

long

Site	Year	PM10
CAM	2013	51.0
FAC	2013	48.3
IZT	2013	44.6
CAM	2014	42.8
FAC	2014	39.0
IZT	2014	39.3
CAM	2015	39.9
FAC	2015	36.6
IZT	2015	35.0



Paquetes base





```
install.packages("tidyverse")
```

Es equivalente a:

```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("tibble")
install.packages("stringr")
install.packages("forcats")
```

```
library(tidyverse)
```

Es equivalente a:

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(readr)
library(purrr)
library(tibble)
library(stringr)
library(forcats)
```



```
install.packages("tidyverse")
```

```
library(tidyverse)
```

Es equivalente a:

```
install.packages("ggplot2")
```

```
install.packages("dplyr")
```

```
install.packages("tidyr")
```

```
install.packages("readr")
```

```
install.packages("purrr")
```

```
install.packages("tibble")
```

```
install.packages("stringr")
```

```
install.packages("forcats")
```

```
R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> library(tidyverse)
— Attaching packages ————— tidyverse 1.3.1 —
✓ ggplot2 3.3.5   ✓ purrr  0.3.4
✓ tibble  3.1.4   ✓ dplyr   1.0.7
✓ tidyr   1.1.3   ✓ stringr 1.4.0
✓ readr   2.0.1   ✓ forcats 0.5.1
— Conflicts ————— tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
```



Su turno...

- Desde la consola de **R** instale los siguientes paquetes del **CRAN**:
 - **tidyverse**, **lubridate**, **readxl**, **haven**, **janitor**.

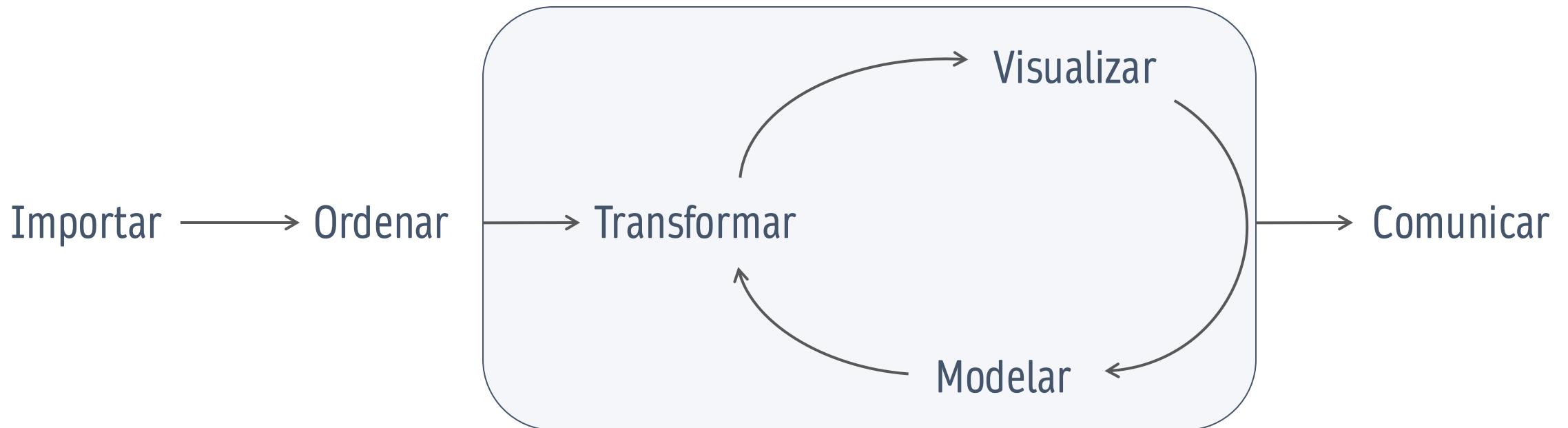
```
> install.packages("tidyverse", dependencies = TRUE)
```



Librerías ampliadas tidyverse

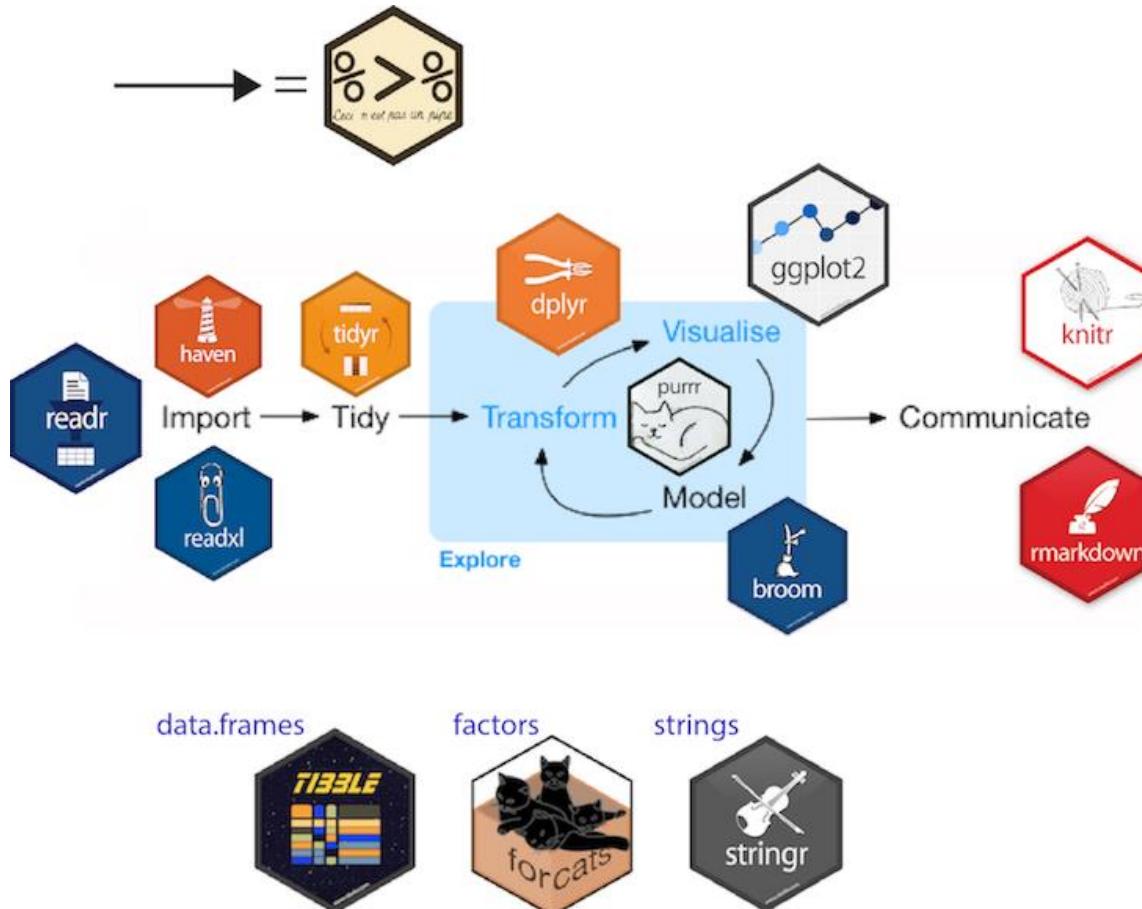


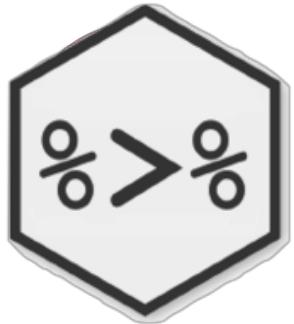
Flujo de trabajo



Programar

Flujo de trabajo usando paquetes





Pipe



El operador `%>%` simplifica y concatena múltiples funciones

```
malla_datos %>%  
  filtro %>%  
  genero_variables %>%  
  agrupo %>%  
  paso_a_wide %>%  
  genero_variables %>%  
  selecciono_columnas
```

```
malla_datos |>  
  filtro |>  
  genero_variables |>  
  agrupo |>  
  paso_a_wide |>  
  genero_variables |>  
  selecciono_columnas
```



1. Click the gear icon in the bottom-left corner of the interface to open the Settings menu.

2. In the Settings menu, click on "Settings [⌘,]" (numbered 2).

3. In the Settings sidebar, click on "Extensions" (numbered 3).

4. In the Extensions list, click on "R" (numbered 4).

5. In the "R" settings section, click the dropdown arrow next to the pipe operator ("|>") and select a value (numbered 5).

The screenshot shows the RStudio IDE interface with the following sections visible:

- Top Bar:** Includes "New", "Open", "File", "Edit", "View", "Project", "Run", "Help", "Session", "Connections", "Help", "Viewer".
- Left Sidebar:** Shows "EXPLORER", "NO FOLDER OPENED" (with a note about opening a folder), "User", "Extensions" (with a list including Npm, PHP, Posit Publisher, Positron Assistant, Pyright, Python, Python Debugger, Quarto, R, Reference Search, Remote SSI, Reticulate), "OUTLINE" (with a note about the active editor), "Command Palette...", "Profiles", and "Check for Updates..." (with a red notification badge).
- Central Area:** Titled "Extensions" with a sub-section for "R". It includes "R > Diagnostics: Enable" (checkbox checked), "R: Pipe" (with a dropdown menu showing "|>"), and a "PROBLEMS" tab.
- Bottom Area:** Shows the R console output:

```
R 4.5.1 started.  
R version 4.5.1 (2025-06-13) -- "Great Square Root"  
Copyright (C) 2025 The R Foundation for Statistical Computing  
Platform: aarch64-apple-darwin20  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
Collaborative project with many contributors.  
Contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
'demo()' for some demos, 'help()' for on-line help, or  
'?start()' for an HTML browser interface to help.  
'q()' to quit R.
```
- Right Sidebar:** Titled "SESSION", "VARIABLES" (empty), and "PLOTS".



Insertar pipe

command

+

shift



+

M



Mac

Ctrl

+

Shift

+

M



Windows



Insertar chunk

command

+

alt

+

|

Ctrl

+

Alt

+

|



Mac



Windows



Insertar operador de asignación <-

option
option

+

-



Mac

Alt

+

-



Windows

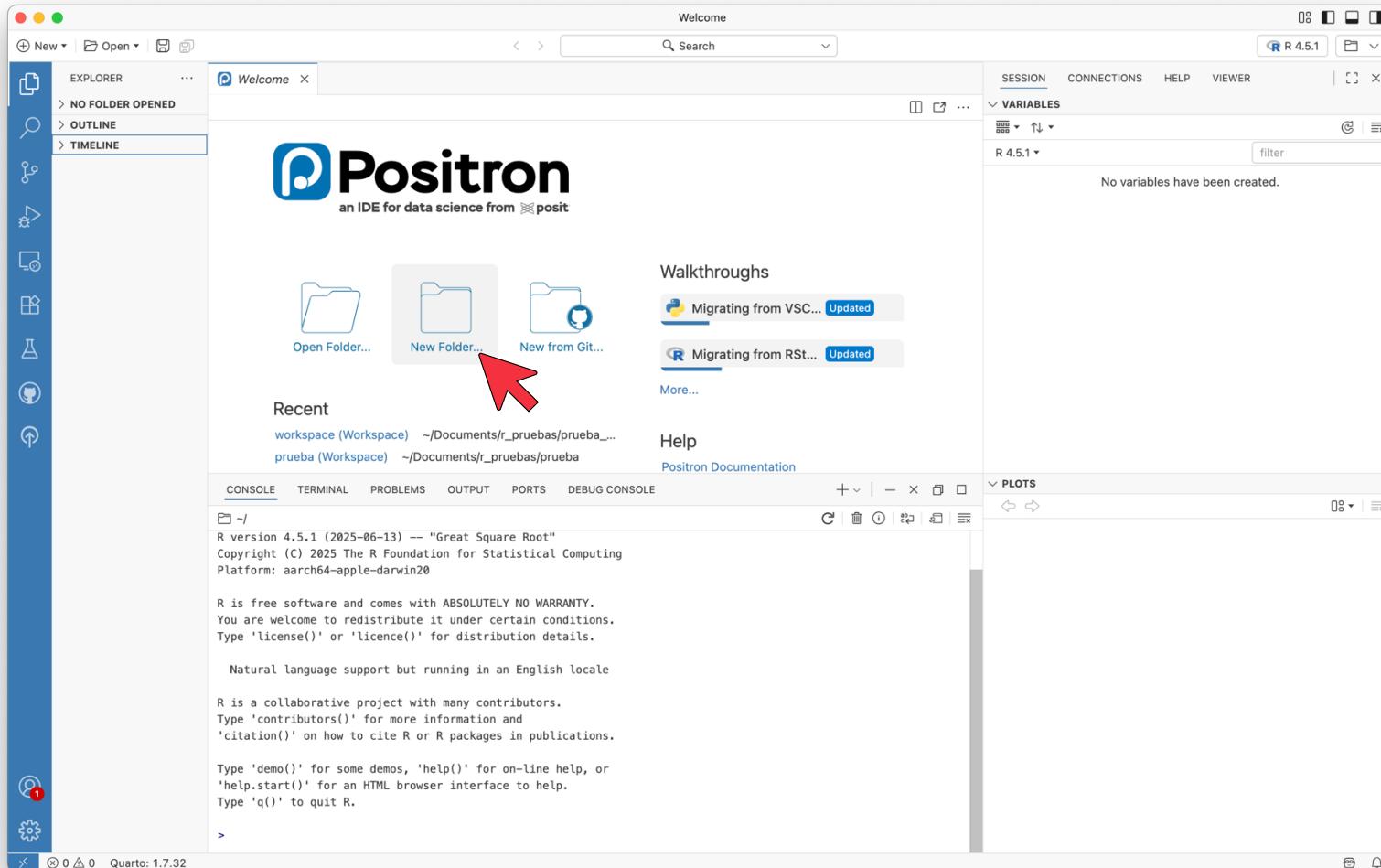


Using = instead of <- for assignment



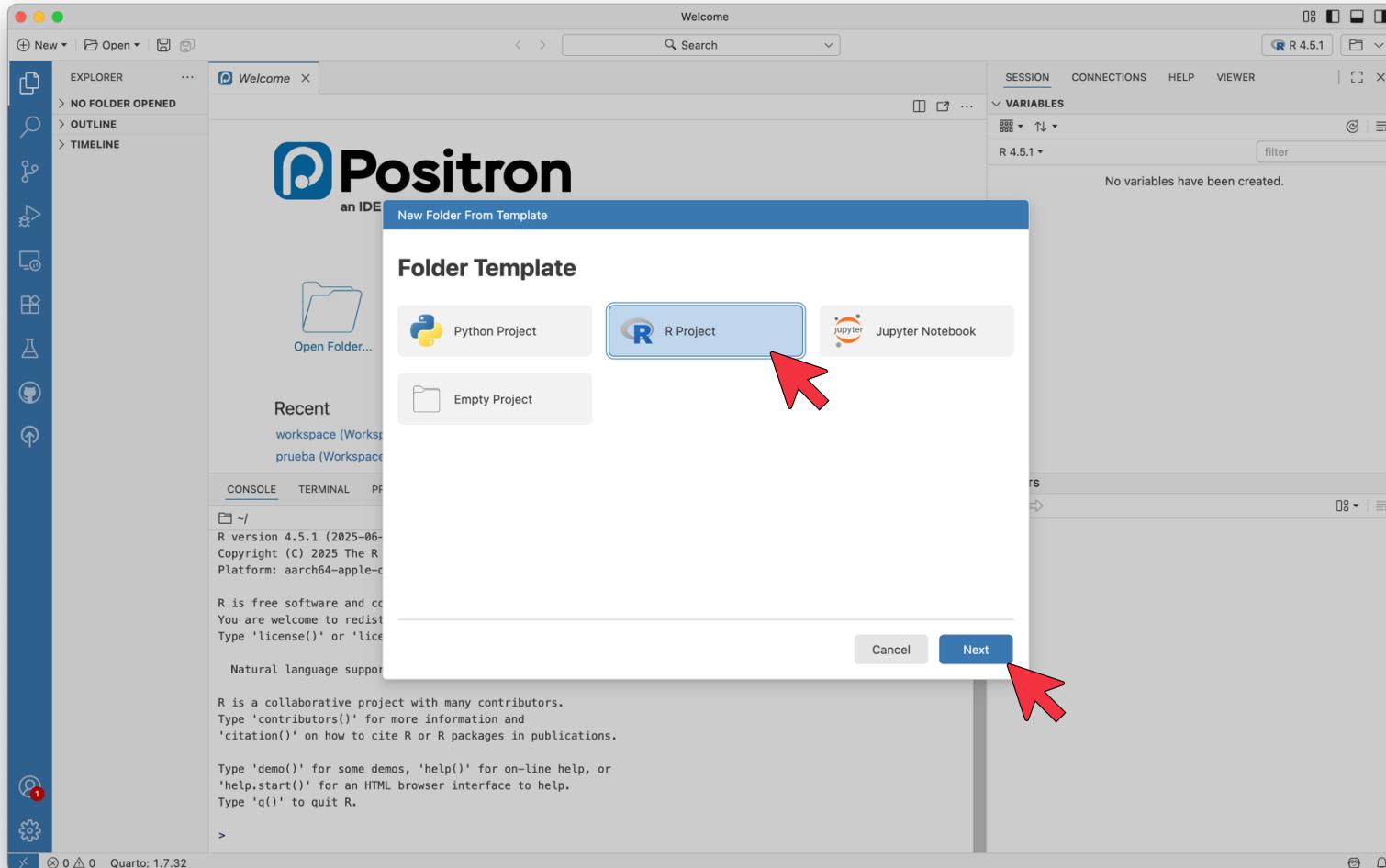


Organizar nuestro trabajo en “folders” (1)



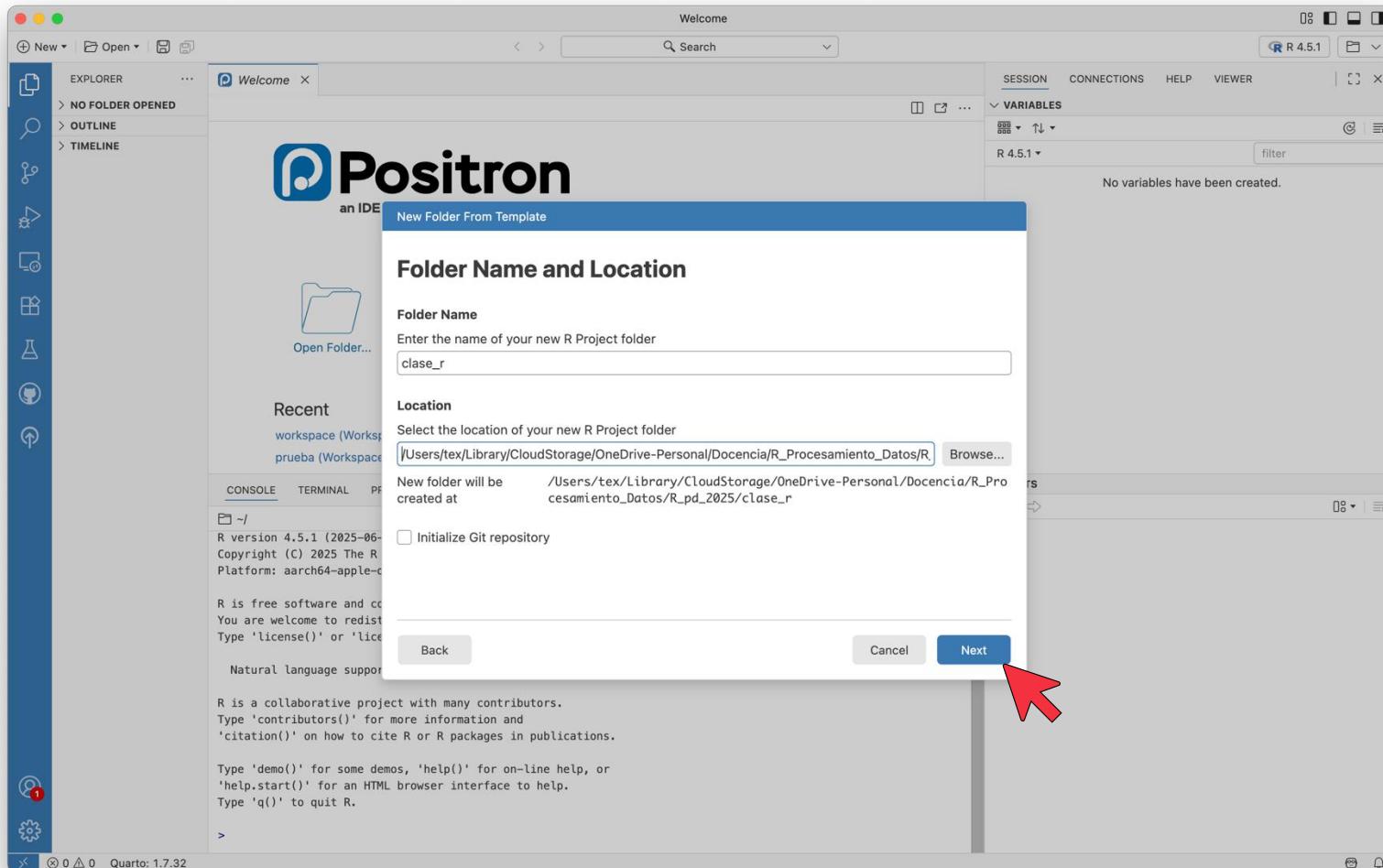


Organizar nuestro trabajo en “folders” (2)



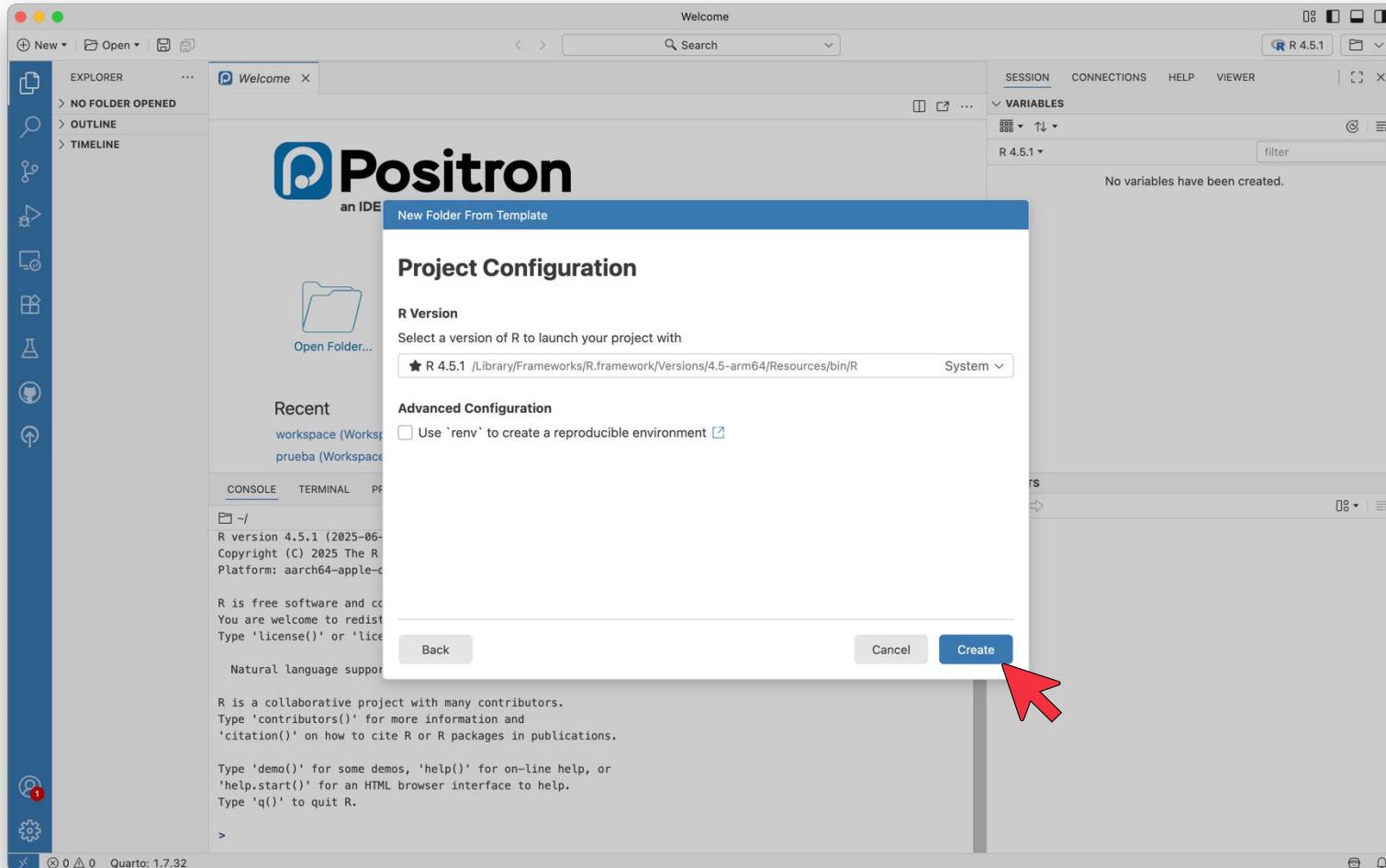


Organizar nuestro trabajo en “folders” (3)



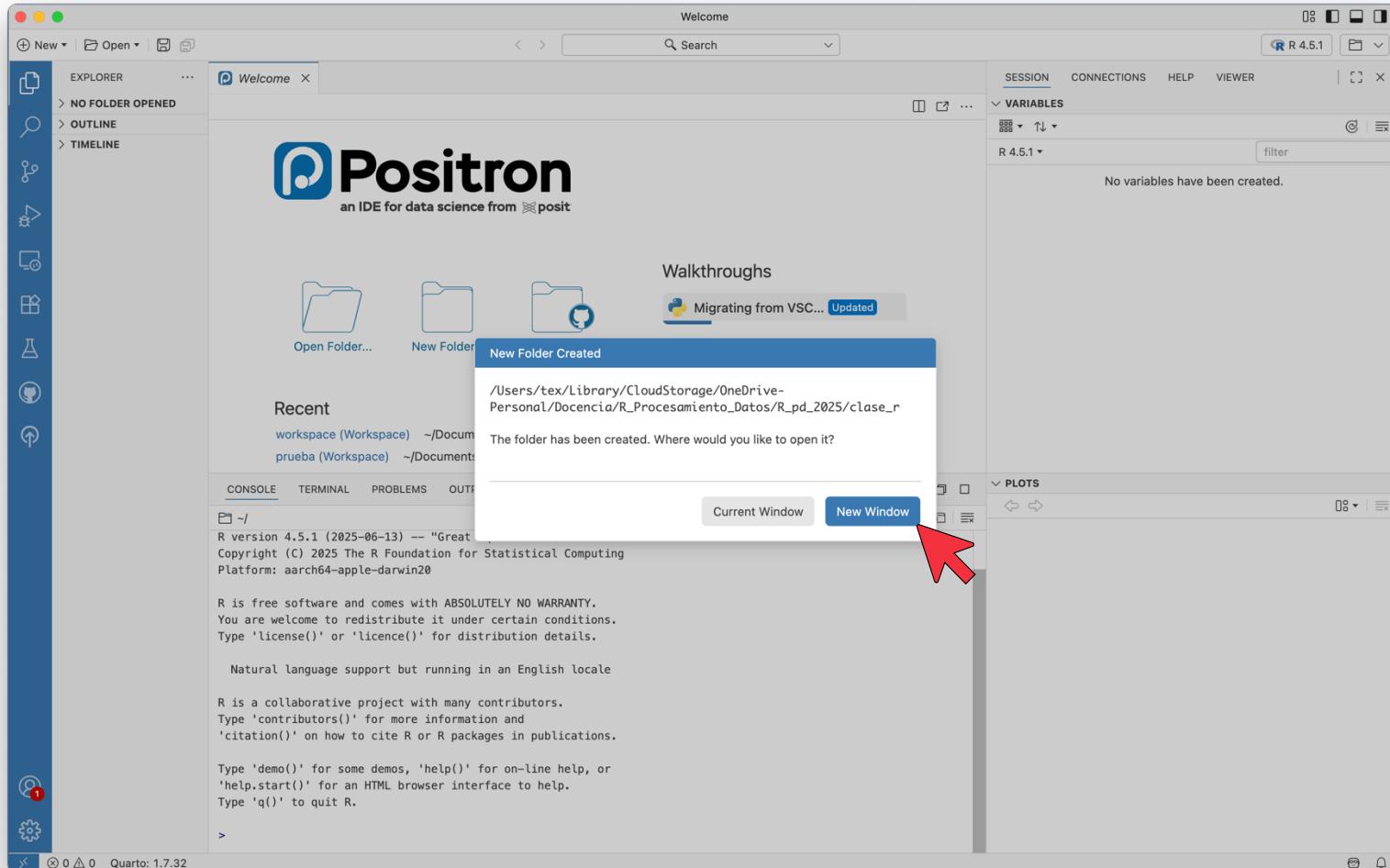


Organizar nuestro trabajo en “folders” (4)



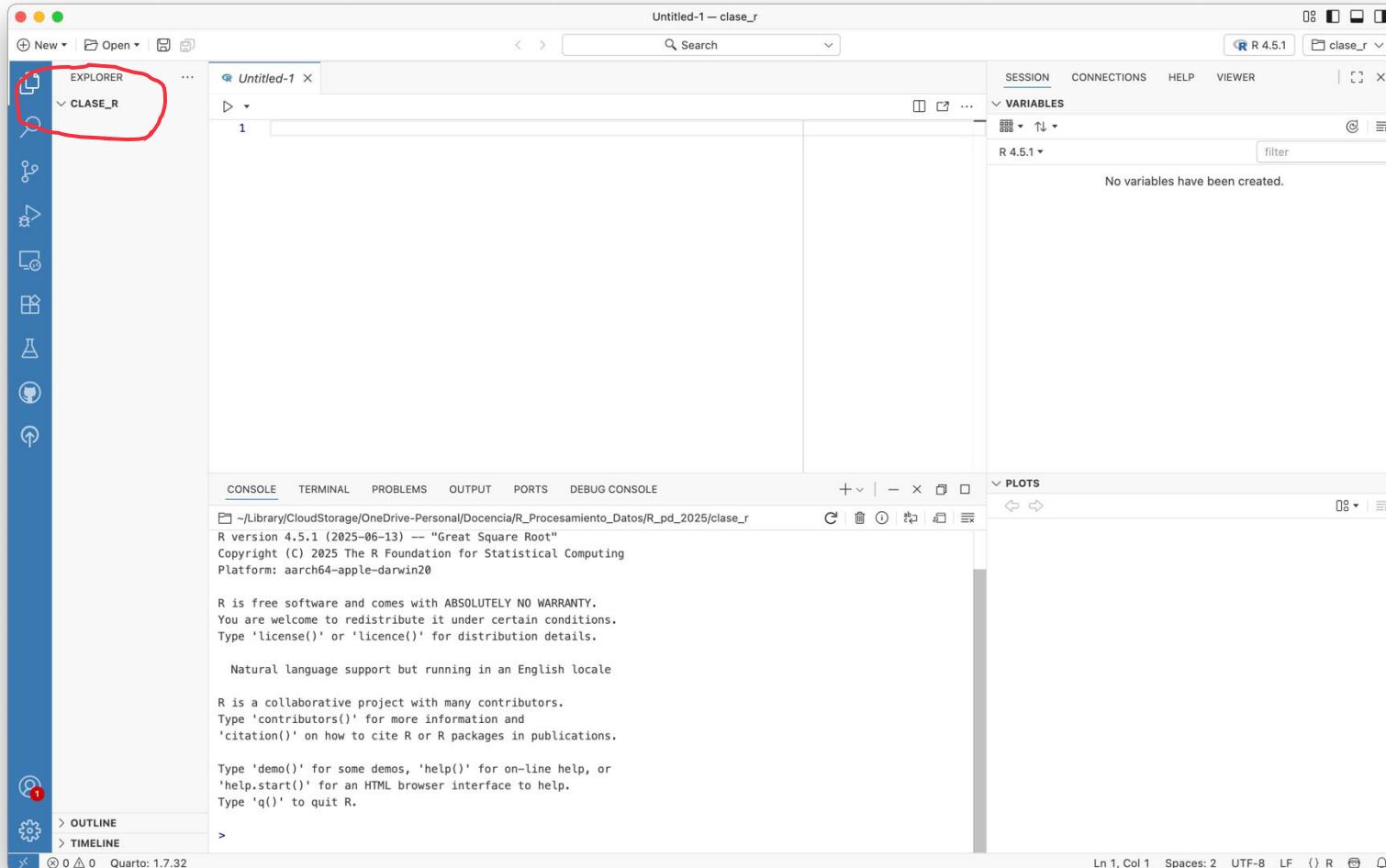


Organizar nuestro trabajo en “folders” (5)



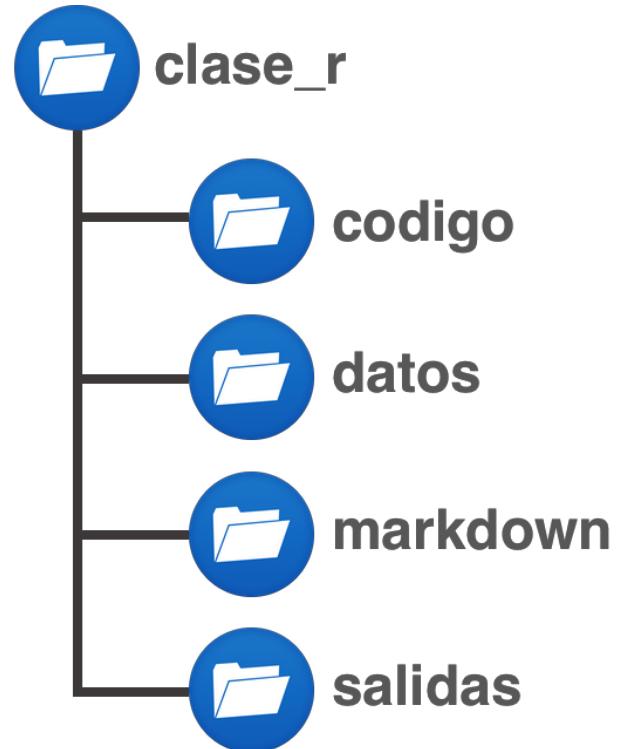
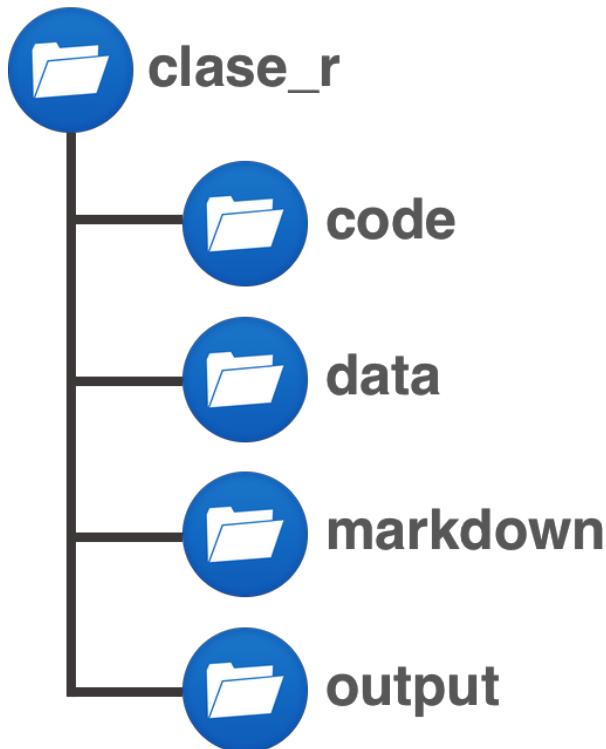


Organizar nuestro trabajo en “folders” (6)



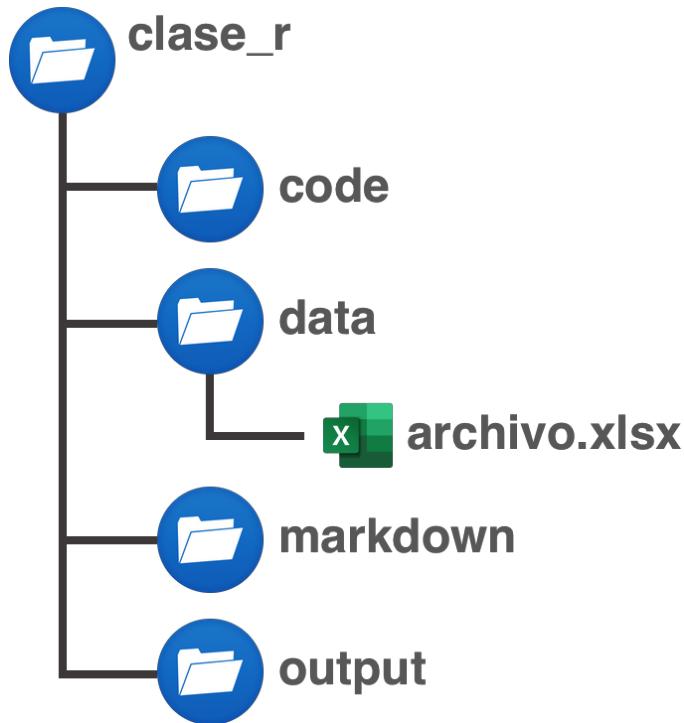


Configuración básica de un "folder"





Configuración básica de un "folder"



Llamar un archivo a una sesión sin un proyecto...

```
> read_xlsx("/Users/tex/Documents/insp/clase_r/data/archivo.xlsx")
```

Llamar un archivo a una sesión dentro un proyecto...

```
> read_xlsx("./data/archivo.xlsx")
```

El “**“folder”** es mutable, es posible compartir la carpeta sin necesidad de modificar el código para redirigir rutas de lectura y/o salida.



Su turno...

- Abra Positron y genere un “folder” de nombre **clase_r** con los siguientes subdirectorios
 - code, data, markdown, output
- Ahora "**clase_r**" será el directorio de su trabajo durante el semestre, revise en el explorador.
- Genere un folder para cada proyecto, clase, análisis u otro tema que usted trabaje.

