

Programación con R

Marzo 2019

Edgar Ruiz

 @theotheredgar

 [linkedin.com/in/edgararuiz](https://www.linkedin.com/in/edgararuiz)

 github.com/edgararuiz

El “tidyverse” *... un repaso*



¿Que es el “tidyverse”?



Una colección de paquetes de R que son diseñados para la Ciencia de Datos.

Todos los paquetes utilizan la misma filosofía de diseño, gramática y estructuras de datos.

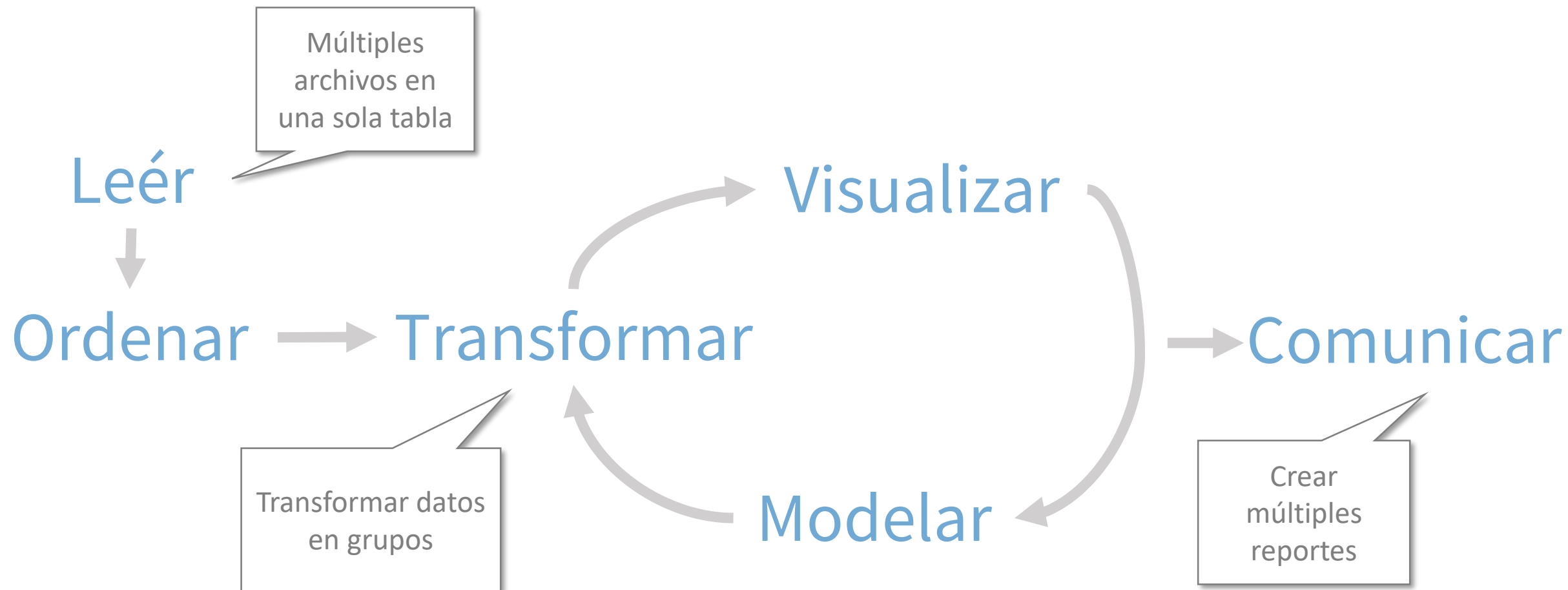
La meta del “tidyverse”



Fóto de [Kelly Sikkema](#) en [Unsplash](#)

Resolver problemas complicados mediante la combinación de diferentes piezas que son consistentes unas con otras

Ciencia de Datos



1. Funciones vectoriales

`x <-`

a	b	c
---	---	---

Repiten sobre un vector

Vector de uno o más

Arguments

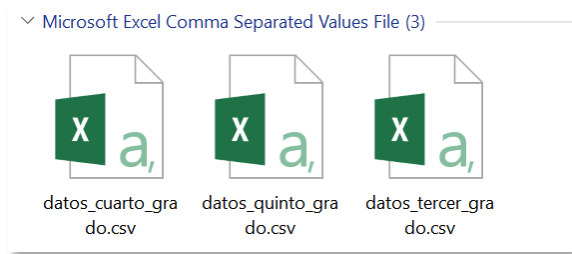
`path` A character vector of one or more paths.
`ext, value` The new file extension.

```
x <- c("home/file1", "user/file2")  
path_file(x)  
## "file1" "file2"
```

2. Múltiples ciclos



Leer múltiples archivos



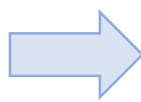
Archivos

`read_csv(`  `)`
datos_tercer_grado.csv

`read_csv(`  `)`
datos_cuarto_grado.csv

`read_csv(`  `)`
datos_quinto_grado.csv

Leer cada uno



día	mes	año	apellido	primer	matricula	origen	grado	materia	puntos
1	1	2015	Hernandez	Rodrigo	100	datos_cuarto_grado.csv	cuarto	ingles	60
1	1	2015	Perez	Roberto	105	datos_cuarto_grado.csv	cuarto	ingles	60
1	1	2015	Ramirez	Alberto	99	datos_cuarto_grado.csv	cuarto	ingles	60
1	3	2015	Perez	Roberto	105	datos_cuarto_grado.csv	cuarto	ingles	62
1	5	2015	Perez	Roberto	105	datos_cuarto_grado.csv	cuarto	ingles	63
1	4	2015	Ramirez	Alberto	99	datos_cuarto_grado.csv	cuarto	ingles	63
1	6	2015	Perez	Roberto	105	datos_cuarto_grado.csv	cuarto	ingles	64
1	3	2015	Ramirez	Alberto	99	datos_cuarto_grado.csv	cuarto	ingles	64
1	2	2015	Ramirez	Alberto	99	datos_cuarto_grado.csv	cuarto	ingles	65
1	2	2015	Perez	Roberto	105	datos_cuarto_grado.csv	cuarto	ingles	66
1	5	2015	Ramirez	Alberto	99	datos_cuarto_grado.csv	cuarto	ingles	66
1	2	2015	Hernandez	Rodrigo	100	datos_cuarto_grado.csv	cuarto	ingles	70
1	4	2015	Perez	Roberto	105	datos_cuarto_grado.csv	cuarto	ingles	70

Combinarlos en
una tabla

Usando purrr

`map (`  `, read_csv)`

Lista o vector Función

Requiere una conjunto y una función. Siempre devuelve un objeto tipo lista.

Funciones purrr

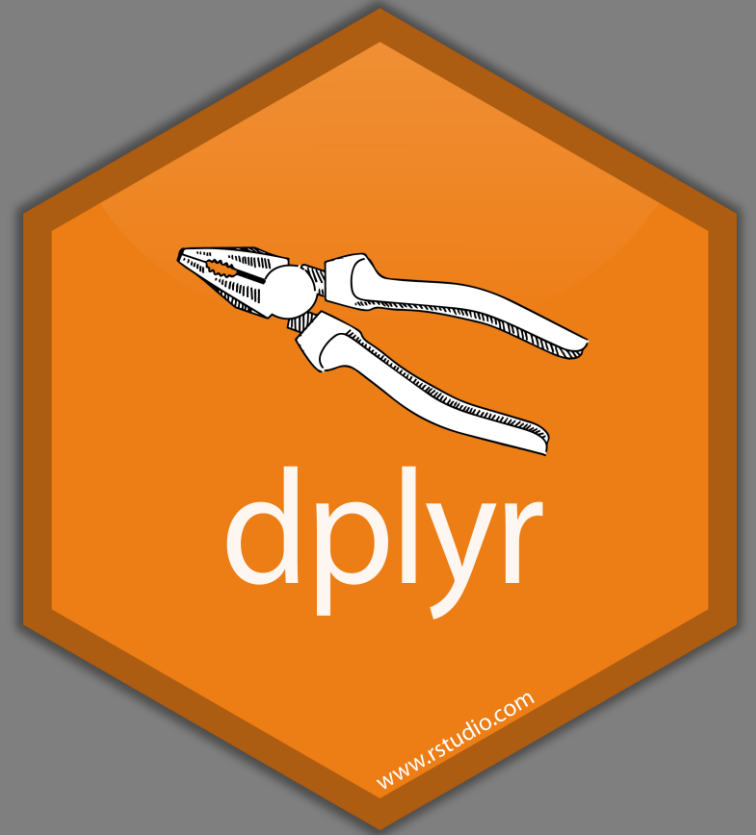
Empieza aquí...

Función	Devuelve
map	Lista
map_chr	Vector alfa-númerico
map_dbl	Vector numérico
map_dfc	Tabla, armada por columna
map_dfr	Tabla, armada por línea
map_int	Vector numérico entero
map_lgl	Vector lógico (TRUE o FALSE)
walk	Lista, invisible

Continúa aquí...

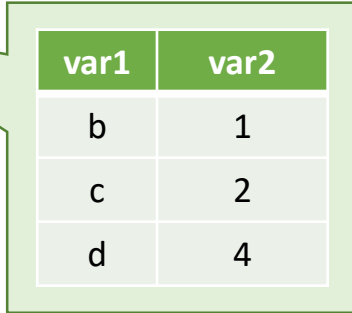
Función	Devuelve
map2	Acepta dos o más listas o vectores
transpose	Convierte los valores de columnas a líneas
pmap	Acepta más de dos listas o vectores
imap	Mantiene el nombre de cada miembro de la lista
set_names	Asigna nombre a cada miembro de la lista
reduce	Combina los resultados de map

3. Grupos de datos



Más allá que tablas, tibbles

letra	datos
a	<tibble>
b	<tibble>
c	<tibble>



var1	var2
b	1
c	2
d	4

tibble permite crear variables de
tipo tibble y otros tipos

Agrupar datos dentro tibbles

`group_nest(nombre, apellido)`

nombre	apellido	mes	valor
Juan	Uno	Enero	10
Sofía	Dos	Febrero	20
Juan	Uno	Febrero	10
Sofía	Dos	Enero	5



nombre	apellido	data
Juan	Uno	<tibble>
Sofía	Dos	<tibble>

Sumariza la table por cada valor único del grupo,
y pone los detalles dentro tibbles individuales

Operaciones por grupo

```
mutate(m = map(lm(x~y, data = .x)))
```

nombre	apellido	data
Juan	Uno	<tibble>
Sofía	Dos	<tibble>



nombre	apellido	data	m
Juan	Uno	<tibble>	<lm>
Sofía	Dos	<tibble>	<lm>

Después de agrupar, se pueden correr operaciones por cada grupo usando `mutate()` y `map()`

Recursos disponibles



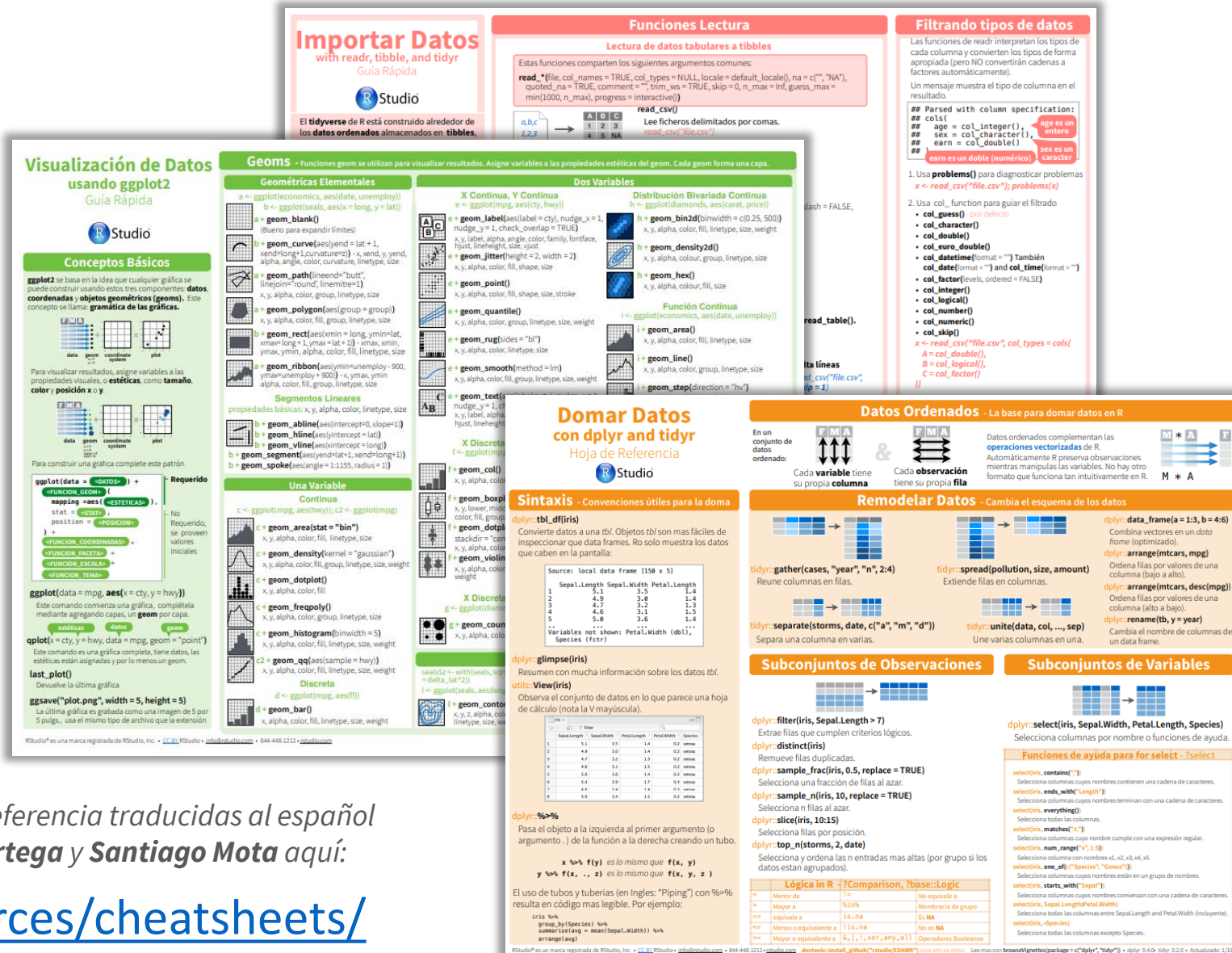
Photo by [Jonathan Simcoe](#) on [Unsplash](#)

Recursos en español

Para aprender como usar los paquetes en práctica, las Hojas de Referencia, o *Cheatsheets*, son los mejores recursos, no importa el idioma

Tenemos disponibles varias Hojas de Referencia traducidas al español cortesía de **Frans van Dunné**, **Carlos Ortega** y **Santiago Mota** aquí:

rstudio.com/resources/cheatsheets/



Recursos en español

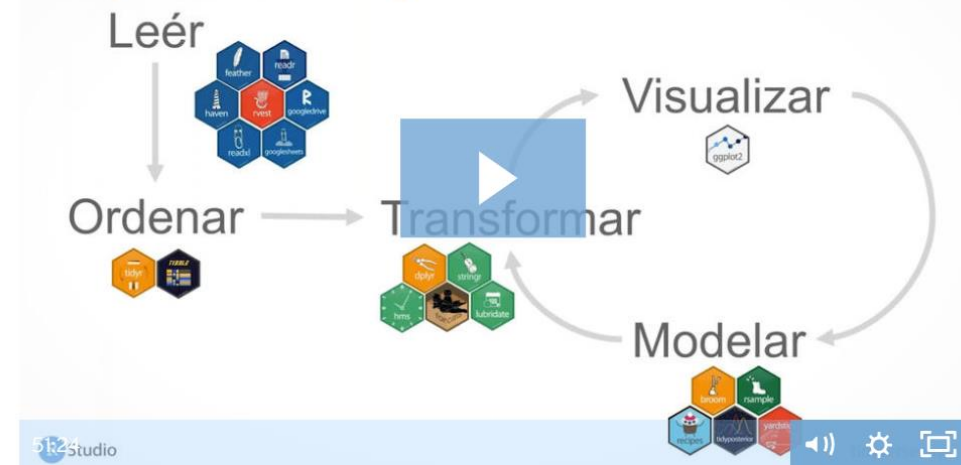
Este, y el webinar anterior, son una nueva base para aprender a utilizar R de manera efectiva en nuestros análisis. Estos webinars proveen tres cosas:

1. Ejemplos de código
2. Presentaciones
3. Video de la sesión

<https://resources.rstudio.com/espanol>

Usando R para la Ciencia de Datos

Paquetes del “tidyverse”



RStudio



DOWNLOAD MATERIALS

Recursos en ingles

1. Reporte - **R Notebook**: <https://bookdown.org/yihui/rmarkdown/notebook.html>
2. Presentaciones – **Xaringan**: <https://github.com/yihui/xaringan>
3. Tablas de datos – **DT**: <https://rstudio.github.io/DT/>
4. **Parameters** en RMarkdown: https://rmarkdown.rstudio.com/developer_parameterized_reports
5. Programe ciclos - **map()**: <https://purrr.tidyverse.org/reference/map.html>
6. Generar reportes con usando código- **render()**: <https://rmarkdown.rstudio.com/lesson-9.html>
7. Dashboards – **flexdashboard**: <https://rmarkdown.rstudio.com/flexdashboard/>
8. Dashboards – **shinydashboard**: <https://rstudio.github.io/shinydashboard/>

¡Gracias por
su atención!

Foto de [Gary Bendig](#) en [Unsplash](#)

