

R

WEEKEND

02 Octubre

UNA INTRODUCCIÓN A TIDYVERSE

CONFERENCISTA:

LINDA
CABRERA ORELLANA

Organiza



SOCIEDAD
ECUATORIANA
ESTADÍSTICA

Colaboran:



R

WEEKEND

02 Octubre

UNA INTRODUCCIÓN A TIDYVERSE

CONFERENCISTA:

DANISSE
CARRASCAL POLO

Organiza



SOCIEDAD
ECUATORIANA
ESTADÍSTICA

Colaboran:





UNA INTRODUCCIÓN A | tidyverse





¡Hola!

Somos Linda y Danisse

Somos co organizadoras de los capítulos de Guayaquil y de Barranquilla, respectivamente. Nos puedes encontrar en twitter como: @linditacabrera y @_danisse__





¿Qué es el tidyverse?

El tidyverse es un conjunto de paquetes diseñados para ciencia de datos en R. Comparten una filosofía de diseño, una gramática y estructuras de datos subyacentes.

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




**Hadley
Wickham**

¿Qué paquetes componen el tidyverse?

Son aproximadamente 30 paquetes los que hacen parte del tidyverse. Entre ellos:

- readr
- tibble
- stringr
- purrr
- dplyr
- forcats
- tidyr
- ggplot2

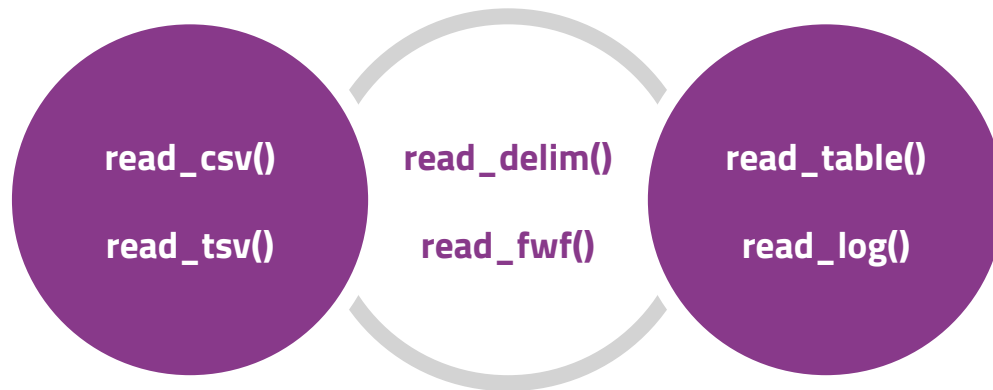


tidyverse_packages()

```
[1] "broom"      "cli"      "crayon"    "dbplyr"    "dplyr"
[6] "dtplyr"     "forcats"  "googledrive" "googlesheets4" "ggplot2"
[11] "haven"      "hms"      "httr"      "jsonlite"  "lubridate"
[16] "magrittr"   "modelr"   "pillar"    "purrr"     "readr"
[21] "readxl"     "reprex"   "rlang"     "rstudioapi" "rvest"
[26] "stringr"    "tibble"   "tidyr"     "xml2"      "tidyverse"
```

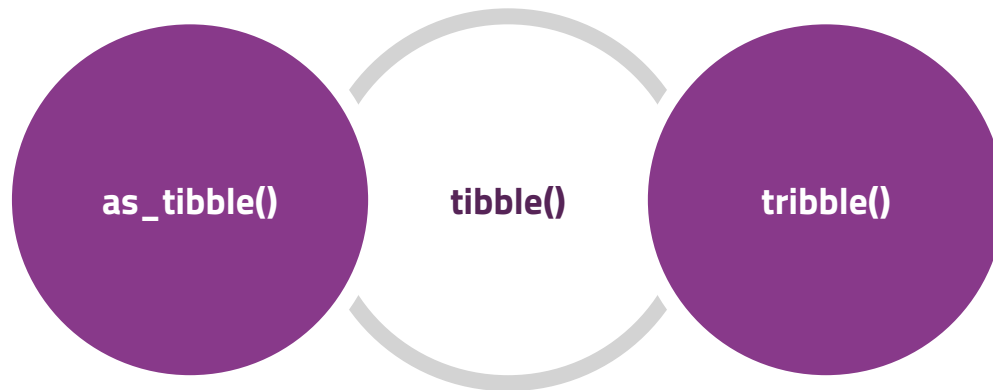

readr

Permite leer datos rectangulares (como csv, tsv y fwf) de una forma rápida y sencilla.



tibble

Provee una versión de data frame que facilita el trabajo con el tidyverse



stringr

Tiene varias funciones que permiten trabajar fácilmente con cadenas de caracteres.

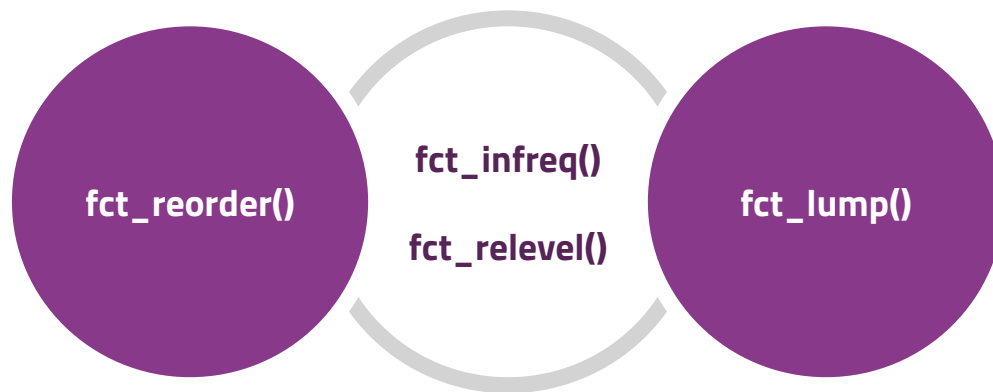


`str_...()`



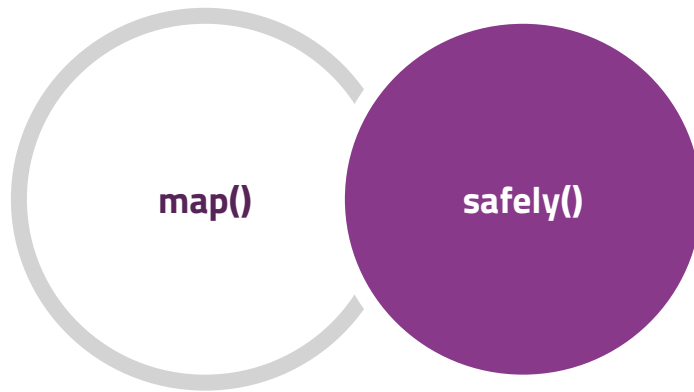
forcats

Proporcionan un conjunto de herramientas que facilita el manejo de factores.



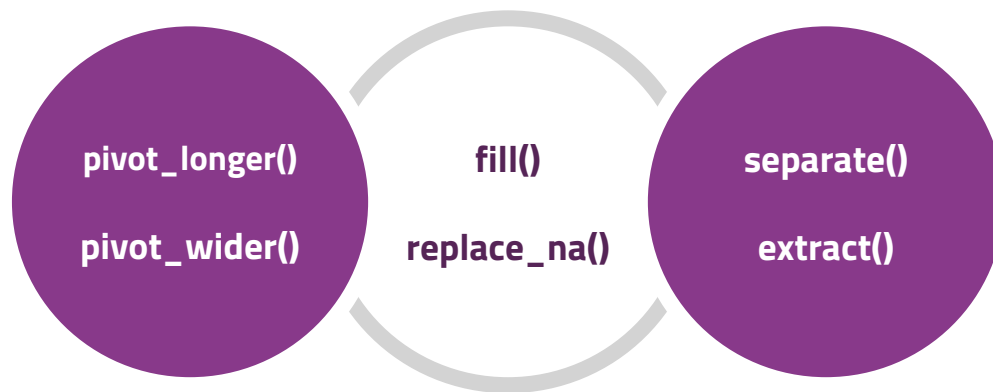
purrr

Permite realizar bucles de una manera no convencional, diferente, quizás más fácil de entender, al tratarlos como funciones naturales.



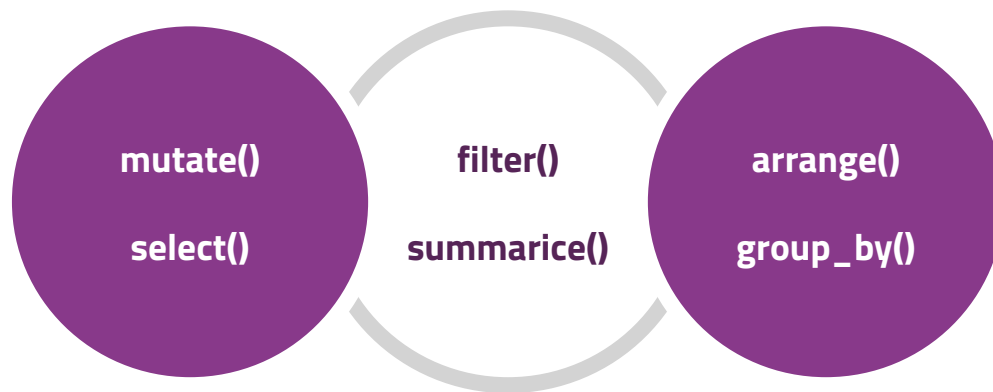
tidyr

Facilita la manipulación y organización de un dataset.



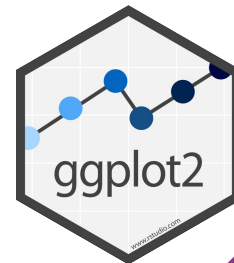
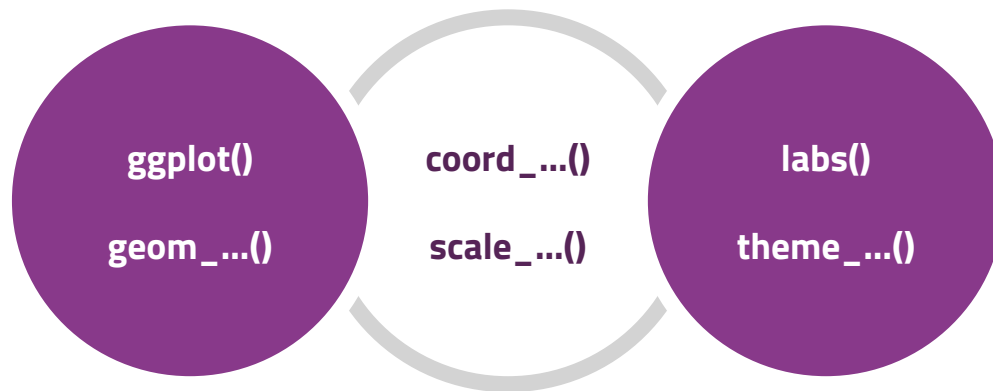
dplyr

Útil para realizar diferentes acciones sobre una base de datos:
filtrar, seleccionar columnas, ordenar, añadir variables nuevas,
etc.



ggplot2

Permite construir distintos gráficos a partir de un conjunto de datos.



remote R learners,



Hagamos un poco de



Datos ordenados con tidyr

Existen tres reglas interrelacionadas que hacen que un conjunto de datos sea ordenado:

1. Cada variable debe tener su propia columna.
2. Cada observación debe tener su propia fila.
3. Cada valor debe tener su propia celda.

pais	anio	casos	poblacion
Afganistán	1999	745	19987071
Afganistán	2000	2666	20595360
Brasil	1999	37737	172006362
Brasil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

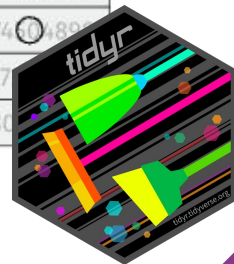
variables

pais	anio	casos	poblacion
Afganistán	1999	745	19987071
Afganistán	2000	2666	20595360
Brasil	1999	37737	172006362
Brasil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

observaciones

pais	anio	casos	poblacion
Afganistán	1999	745	19987071
Afganistán	2000	2666	20595360
Brasil	1999	37737	172006362
Brasil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

valores





Manipular datos con `dplyr`

Dplyr nos da una serie de herramientas para manipular datos.

Los principales verbos o funciones de **dplyr**, son:

- `select()`, para seleccionar columnas
- `filter()`, para filtrar filas
- `mutate()`, para crear o modificar columnas
- `summarise()`, para resumir información de las columnas

Las 3 últimas funciones se pueden usar con `group_by()` para aplicar operaciones por grupos.





Recursos

- [Tidyverse](#)
- [R para ciencia de datos](#)
- [Introducción a la Ciencia de Datos](#)