



Introducción al



Sábado 08 junio - 10:30 UTC-4

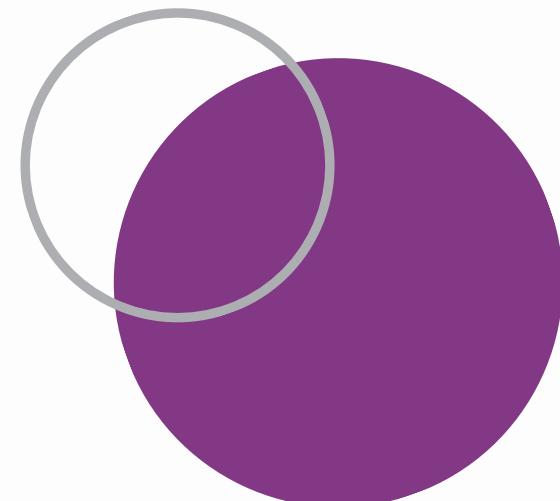


Una actividad organizada por el capítulo de R-Ladies Concepción



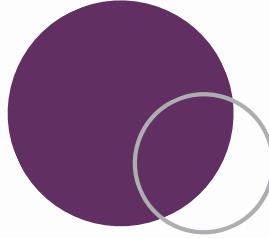
TEMAS PRINCIPALES

PUNTOS QUE SE ABORDARÁN



- ¿Qué es R?
- ¿Para qué sirve?
- ¿Qué es R-Ladies?
- ¿Qué es Tidyverse?

¿Qué es R?



- **R es un entorno y lenguaje de programación con un enfoque en el análisis estadístico.**
- **Licencia gratuita y libre.**
- **En sus orígenes se basó en el lenguaje S, desarrollado por Becker, Chambers y Wilks.**
- **S posteriormente evolucionó a dos tipos de softwares paralelos: Splus y R.**

¿Para qué sirve?

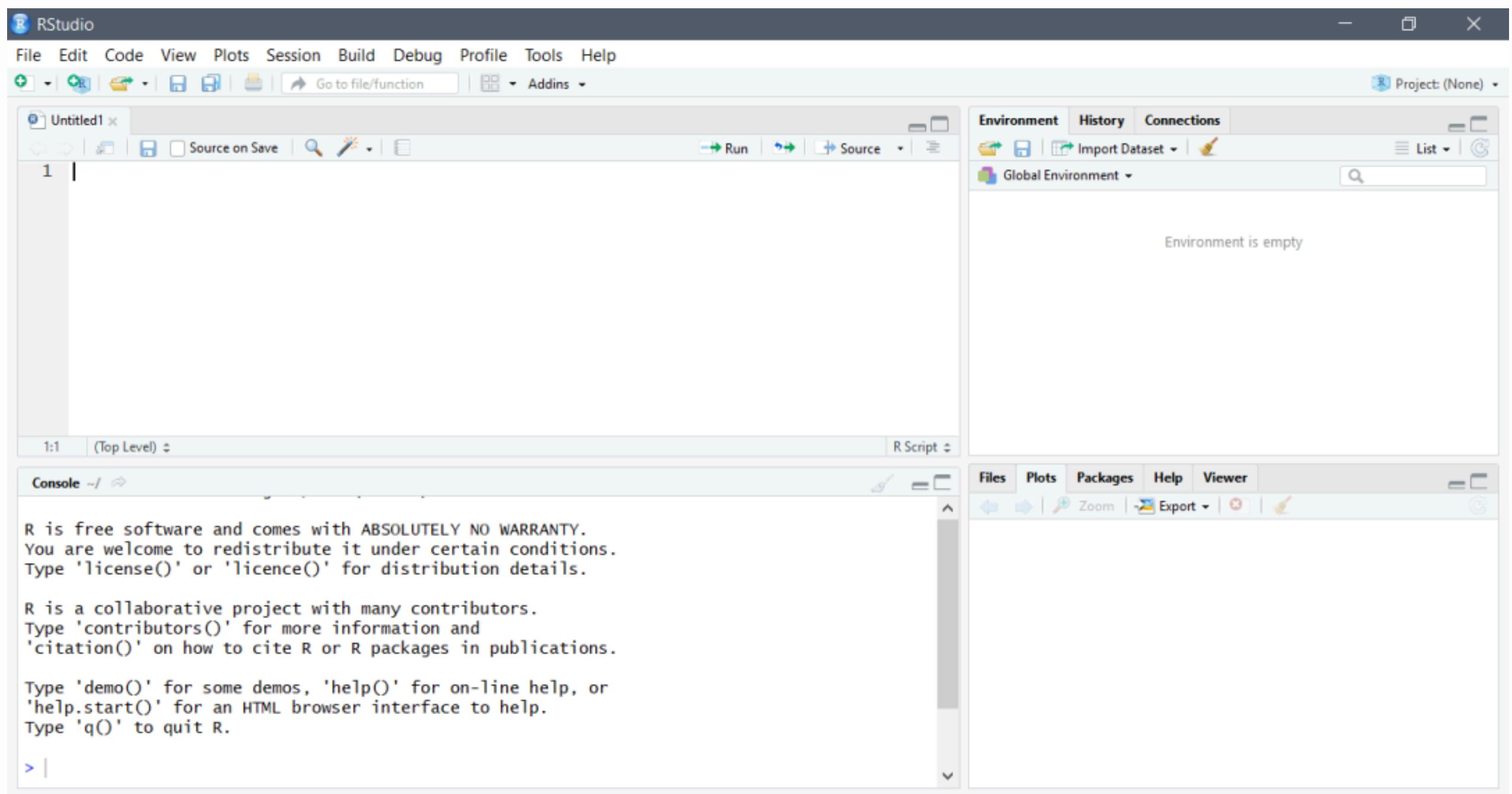
R LO PODEMOS UTILIZAR PARA MUCHOS CASOS



- Modelos estadísticos, desde regresión lineal hasta redes neuronales
 - Series de tiempo
 - Visualización de datos
 - Páginas web
 - Informes con código R

The word cloud is centered around the word "data" in a large, dark gray font. Surrounding "data" are various other words in different colors and sizes, representing related concepts. The words include "scientist" (red), "learn" (green), "science" (pink), "machine" (blue), "python" (orange), "explain" (teal), "introduction" (light blue), "regression" (purple), "deep learning" (yellow), "new" (light green), "network" (light purple), "problem" (light pink), "book" (light orange), "algorithm" (light teal), "use" (light yellow), "simplification" (light pink), "whilst" (light purple), "the" (light orange), "analysis" (light pink), "time" (light purple), "tutorial" (light pink), "article" (light purple), "top" (light pink), "you" (light purple), "language" (light pink), "your" (light purple), "cluster" (light teal), "how" (light orange), "neural network" (light pink), "statistician" (light purple), "free" (light green), "great question" (light pink), "predict" (light purple), "guidelines" (light green), "difference" (light green), "analytical" (light green), "what" (light green), and "know" (light green).

Interfaz de RStudio



RStudio es una IDE o entorno de desarrollo integrado. En breves palabras, es un programa para manejar R y utilizarlo de manera más comoda en algunos aspectos.

¿Qué son los datos Tidy?

¿ QUÉ SON LOS DATOS TIDY?



Los datos Tidy (ordenados) tienen como mínimo las siguientes propiedades:

- **Cada variable tiene su propia columna**
- **Cada observación de esa variable debe estar en su propia fila**

country	year	cases	population
Afghanistan	1990	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	31737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	21666	1280426583

variables

country	year	cases	population
Afghanistan	1990	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	31737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216706	1280426583

observations

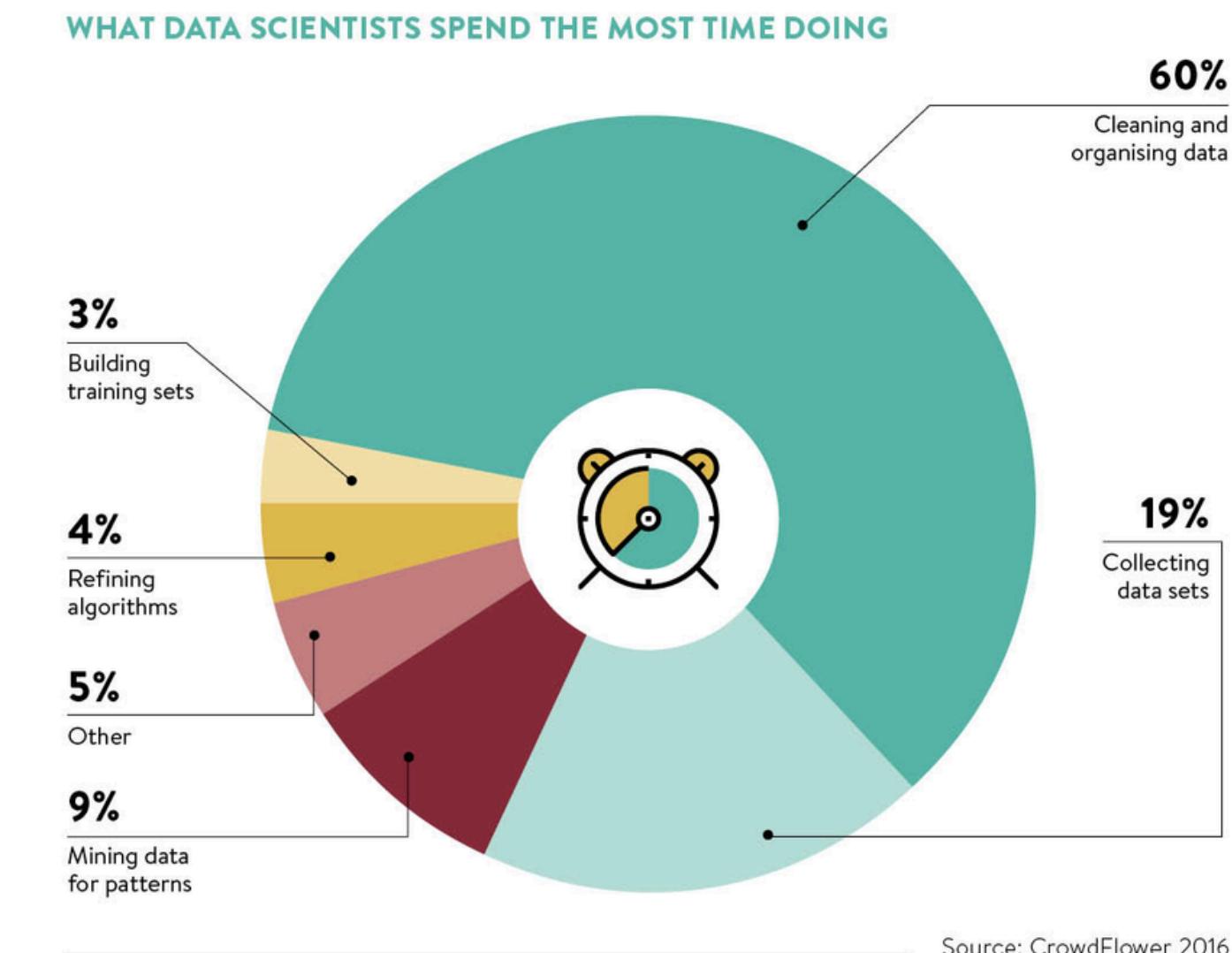
country	year	cases	population
Afghanistan	1990	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	31737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216706	1280426583

values

¿Por qué necesitamos datos ordenados?

Se nos hará más sencillo:

- **Manipular los datos y crear nuestros resúmenes**
- **Unir bases de datos ordenados entre sí**
- **Visualizar nuestros datos**
- **Modelar**



A tener en cuenta: Mucho del trabajo que hagamos al analizar nuestros datos se tratará de limpieza.

¿Qué es tidyverse?





Tidyverse es un conjunto de packages en R, enfocados en el procesamiento y análisis de datos. Su objetivo principal es proporcionar herramientas coherentes y eficientes para trabajar con datos a través de R.

Presenta:

- Sintaxis coherente
- Compatibilidad
- Visualización
- Manipulación de datos eficiente
- Documentación y comunidad

PACKAGES INCLUÍDOS



PACKAGES INCLUÍDOS



- **ggplot 2:** crea visualizaciones y gráficos de alta calidad
- **dplyr:** proporciona funciones para manipular y transformar datos de manera intuitiva
- **tidyverse:** organiza y estructura datos para que sean más fáciles de analizar
- **readr:** permite leer datos en R desde diferentes formatos de archivos
- **purrr:** ofrece herramientas para trabajar con funciones y vectores
- **tibble:** proporciona una estructura de datos eficiente y amigable para trabajar con datos rectangulares
- **stringr:** ayuda en la manipulación y análisis de cadenas de texto
- **forcats:** se utiliza para trabajar con variables categóricas

VENTAJAS



- **Sintaxis coherente:** están diseñados con una sintaxis coherente, lo que facilita el aprendizaje y escritura del código
- **Compatibilidad:** Los packages se complementan entre sí, lo que permite un flujo de trabajo fluido y eficiente
- **Visualización:** ggplot2, proporciona herramientas para crear visualizaciones personalizadas y atractivas
- **Manipulación de datos eficiente:** dplyr y tidyr ofrecen funciones optimizadas para manipular y transformar datos de manera rápida y eficiente
- **Documentación y comunidad:** tidyverse cuenta con información y una comunidad activa, lo que facilita la resolución de problemas



LADIES,
Concepción

¿Qué es ?

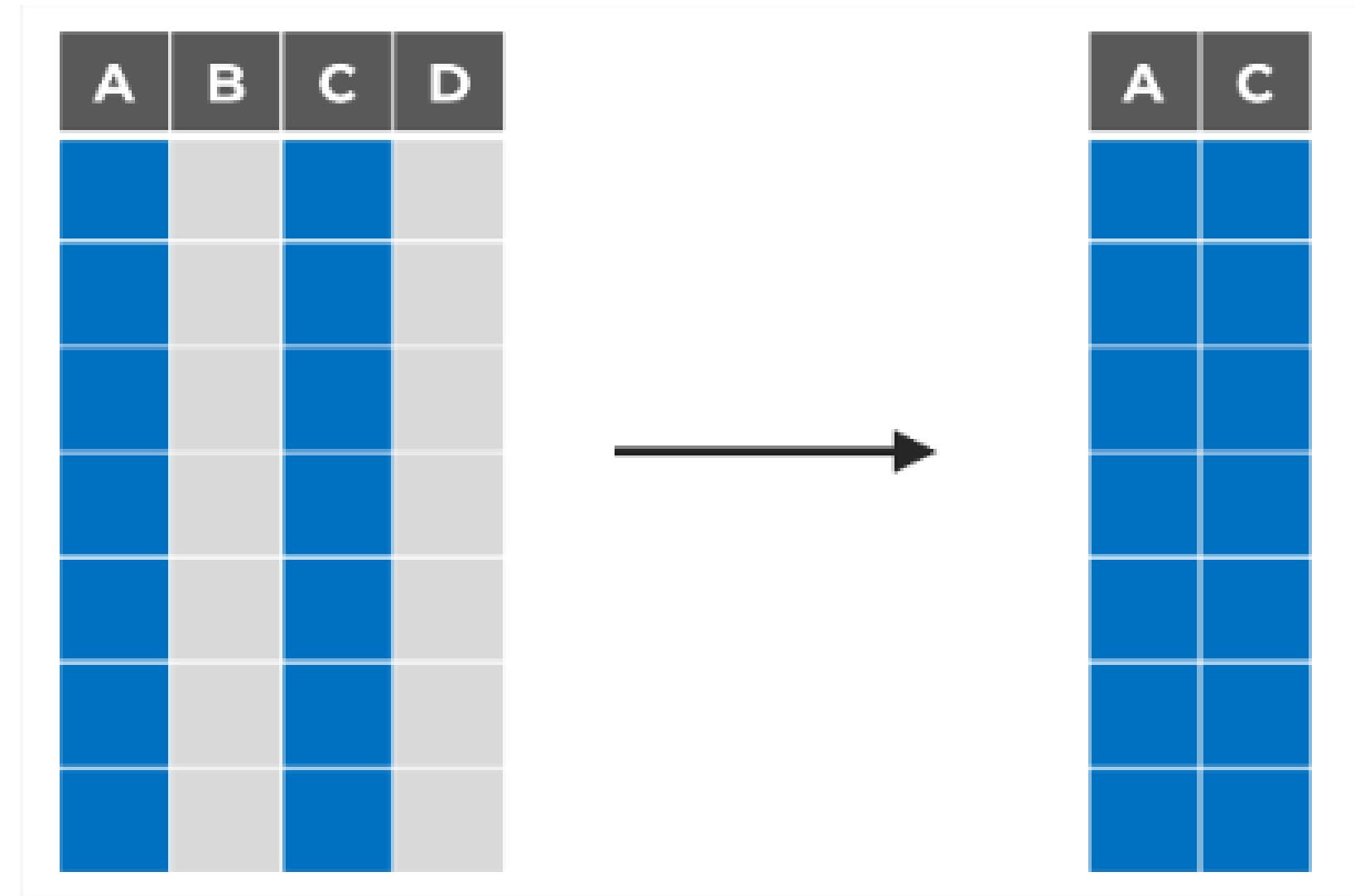


El paquete dplyr facilita la manipulación de tablas, datos o listas mediante una gramática intuitiva. Permite comunicar acciones a través del nombre de la función, simplificando así la manipulación de datos.

- select
- filter
- mutate
- summarise/summarize
- group_by
- %>% (pipe)
- case_when

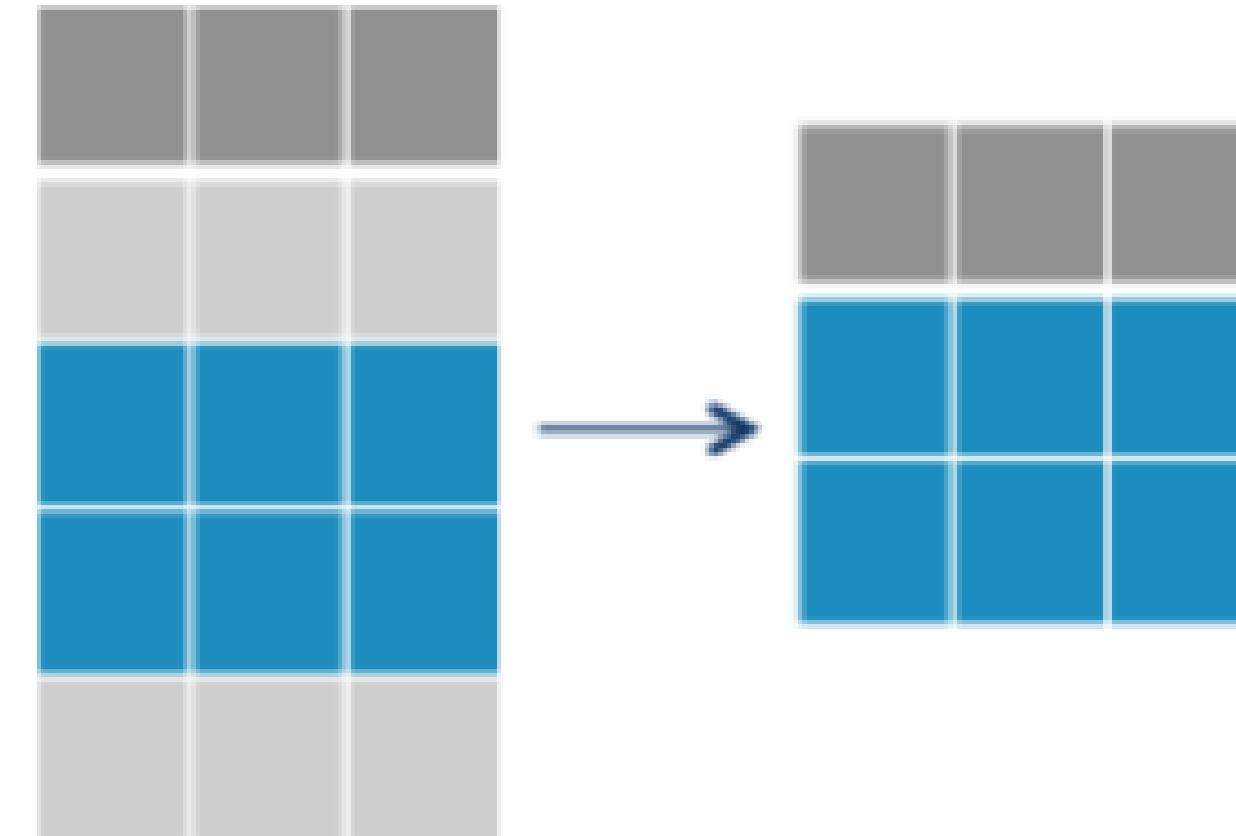
- **dplyr::select**

Devuelve un conjunto de columnas



- **dplyr::filter**

Devuelve un conjunto de filas según una o varias condiciones lógicas



- **dplyr::mutate**

Crea nuevas variables a partir de las existentes



- **dplyr::case_when**



Permite especificar condiciones y asignar valores según esas condiciones. Suele acompañarse de un mutate.

```
variable_creada = case_when( variable_madre == 1 ~ "uno",  
                            variable_madre == 2 ~ "dos")
```



variable_madre	variable_creada
1	uno
2	dos
1	uno

- **dplyr::summarize**



Genera resúmenes de los datos (sumas, conteos, media, rango, la mediana, la sd)



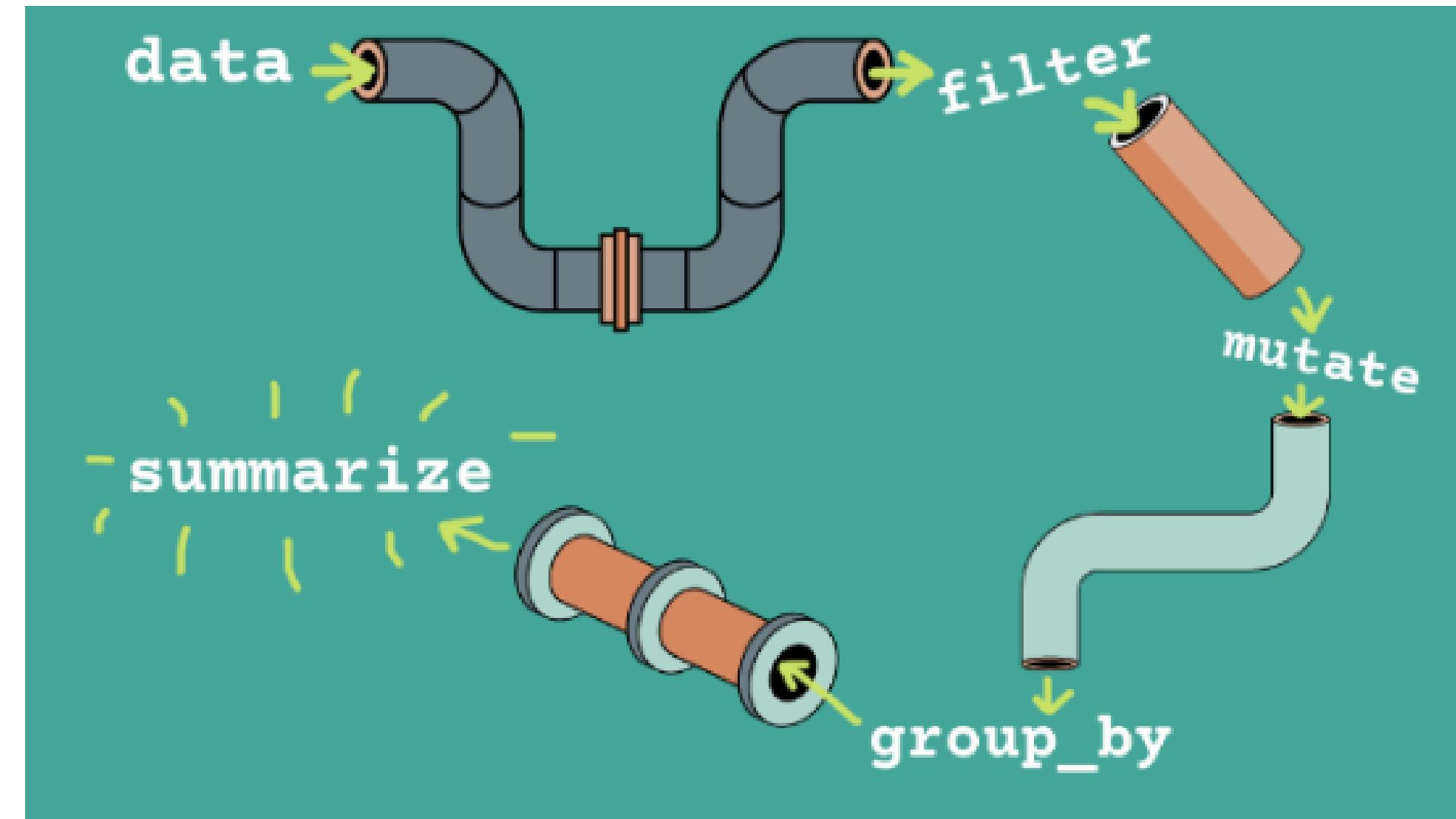
- **dplyr::group_by**

Agrupa la data según una característica determinada. Suele ser acompañada de un resumen de los grupos (summarize).



- `dplyr::pipe(%>%)`

- El operador “pipe” es usado para conectar múltiples acciones en una única “pipeline” (tubería).





¡AHORA, MANOS A LA OBRA!

[HTTPS://POSIT.CLOUD/CONTENT/5960904 \(CAMBIAR\)](https://posit.cloud/content/5960904)

+ INFORMACIÓN



- **Sitio web oficial de tidyverse:** <https://www.tidyverse.org/>
- **Libro "R for Data Science" escrito por Hadley Wickham y Garrett Grolemund:** <https://r4ds.had.co.nz/>
- **Documentación de los paquetes individuales en tidyverse, disponibles en el sitio web de CRAN (Comprehensive R Archive Network).**



Introducción al



Sábado 08 junio - 10:30 UTC-4



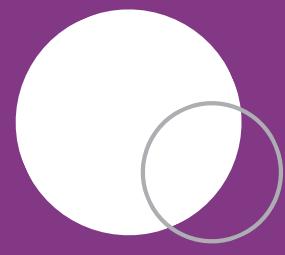
Una actividad organizada por el capítulo de R-Ladies Concepción





PREGUNTAS FRECUENTES





PRÓXIMO TALLER

