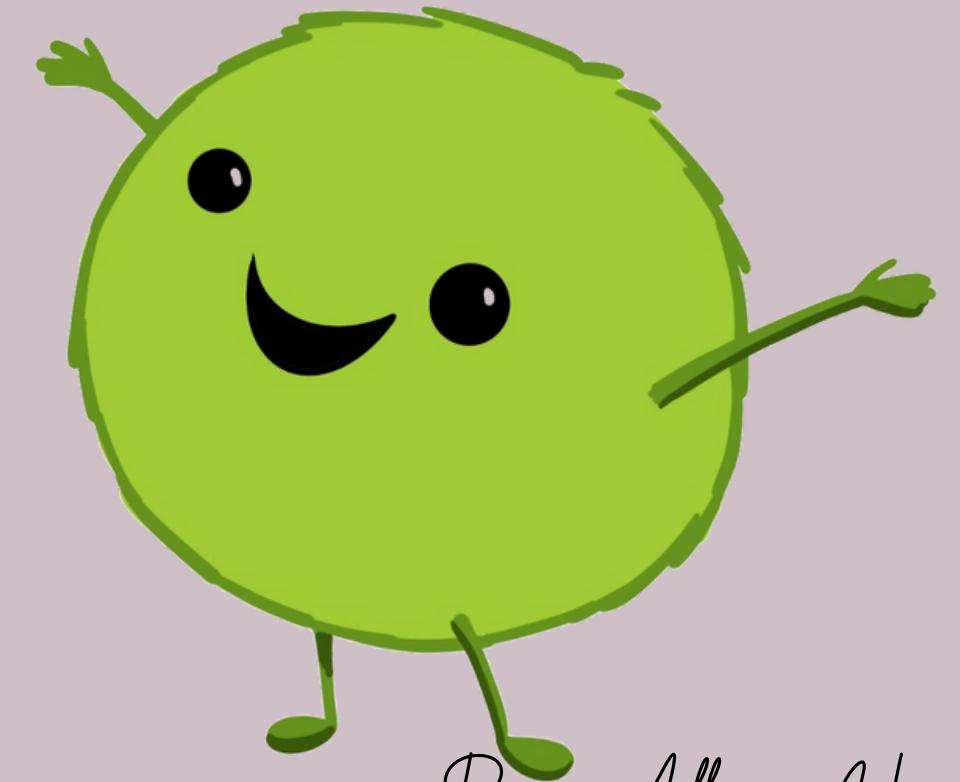
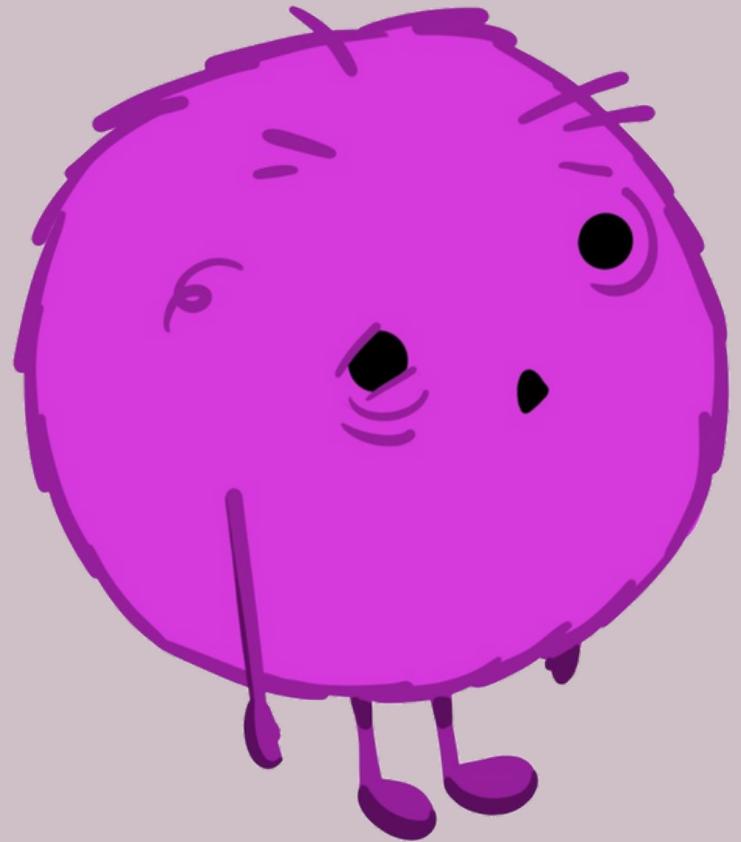


# Club de lectura: R para ciencia de datos.



and there's  
something new  
**EVERY DAY!**

By: Allison Horst  
Via: @allison\_horst

# ¡Conoce **zoom!**



R-Ladies

# GitHub

The screenshot shows a GitHub repository page for the user 'isavasquez' named 'Meetup-presentaciones\_barranquilla'. The repository is a fork of 'rladies / meetup-presentaciones\_barranquilla'. The main navigation bar includes links for 'Solicitudes de extracción', 'Cuestiones', 'Mercado', and 'Explorar'. On the right side, there are metrics for 'Reloj' (0 0), 'Estrella' (0 0), and 'Tenedor' (1). Below the navigation, there are tabs for 'Código', 'Solicitudes de extracción', 'Comportamiento', 'Proyectos', 'Wiki', 'Seguridad', and 'Percepciones'. A green circle with the number '1' points to the 'Código' tab. A purple arrow points from the text 'Seleccionar la opción de descargar código' to the 'Código' tab. A red arrow points from the text 'Click en "Descargar ZIP"' to the 'Descargar ZIP' button in the code download section. The code download section also features a green circle with the number '2'.

isavasquez / Meetup-presentaciones\_barranquilla

bifurcado de rladies / meetup-presentaciones\_barranquilla

Código

Rama: Maestro

Ir al archivo Agregar archivo Código

Clonar con SSH Use HTTPS

No tiene ninguna clave SSH pública en su cuenta de GitHub. Puede [agregar una nueva clave pública](#) o intentar clonar este repositorio mediante [HTTPS](#)

git@github.com:isavasquez/meetup-prese

Abrir con el escritorio de GitHub Descargar ZIP

1

2

Selecciónar la opción de descargar código

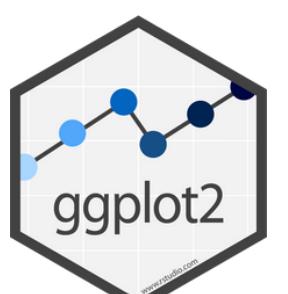
Click en "Descargar ZIP"



# </> Paquetes </>

**tidyverse:** Tidyverse es un conjunto de paquetes en R diseñados para ciencia de datos. Esto lo que significa es que ayuda en todo el proceso de importar, transformar, visualizar, modelar y comunicar toda la información que normalmente utilizamos en procesos de ciencia de datos.

**modelr:** proporciona funciones que ayudan a crear diversos tipos de modelos, apoya la enseñanza de los conceptos básicos del modelado dentro de tidyverse.



+

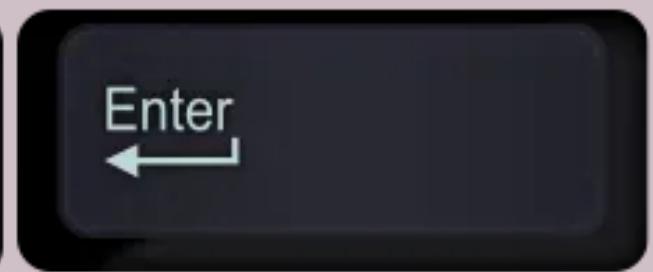
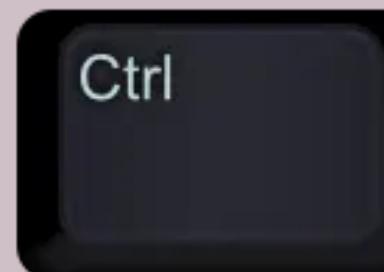


# ¿Cómo instalar un paquete en R Studio ?



```
Club de lectura: R para ciencia de datos
Jueves, 06 de mayo 2021
Hagamos arte en RStudio

#Instalar el paquete tidyverse
install.packages("tidyverse")
#Cargar a la librería el paquete
library(tidyverse)
```



# Orden del día



- ✓ R-Ladies.
- ✓ Retroalimentación capítulo 23.
- ✓ Ejercicios en RStudio.

R-Ladies



# R-Ladies Global

La comunidad R sufre de una representación insuficiente de los géneros minoritarios (incluidas, entre otras, mujeres cis / trans, hombres trans, no binarios, queer, a-género) en todos los roles y áreas de participación, ya sea como líderes, desarrolladores de paquetes, conferencistas, participantes de la conferencia, educadores o usuarios (ver estadísticas recientes).

# Gabriela de Queiroz



Fundadora R-Ladies Global.

# *Código de conducta*

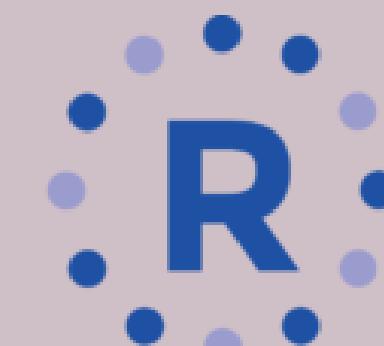
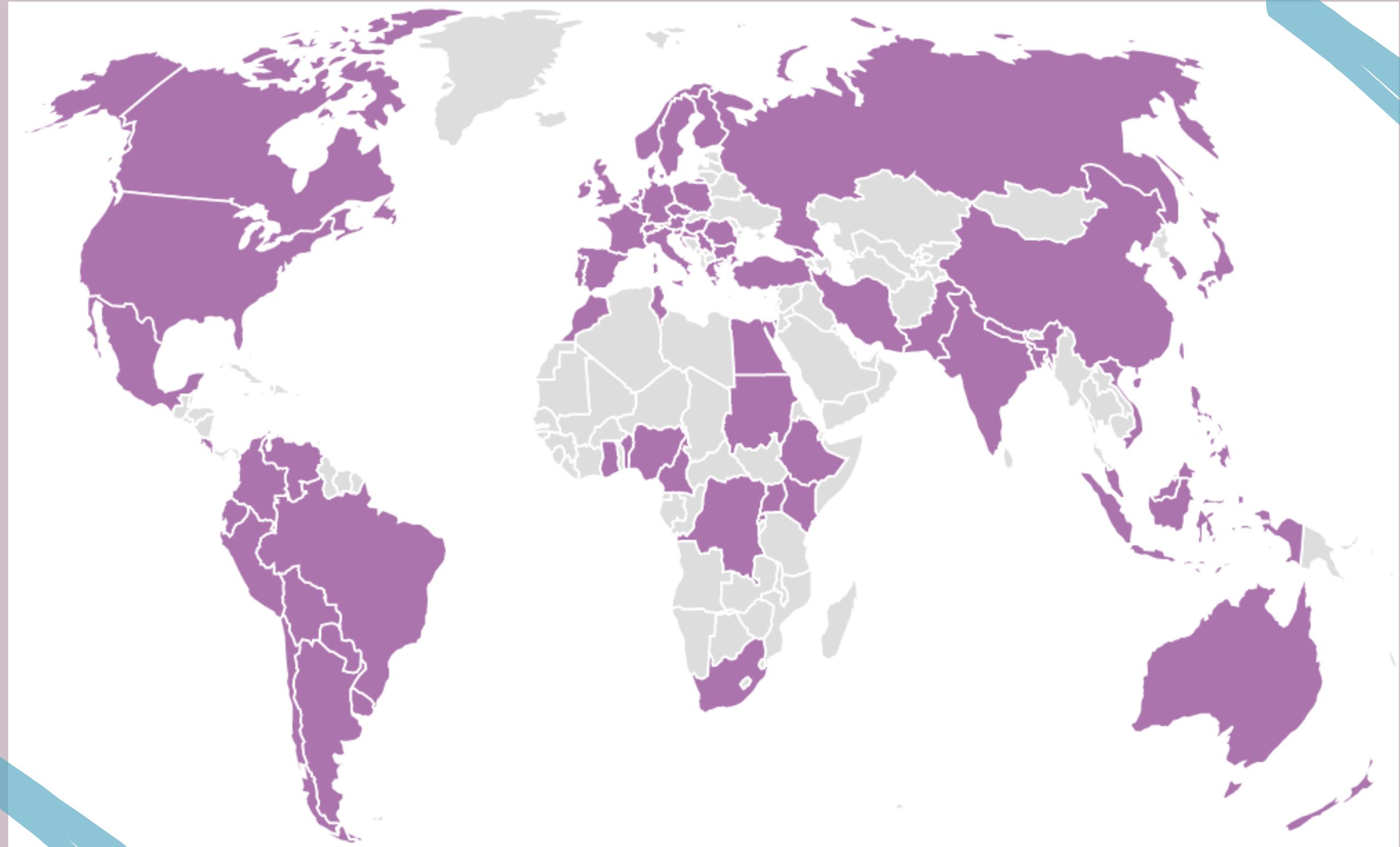
Espacio libre de acoso

Todas(os)  
debemos ser  
tratados con  
respeto

Alertar cuando  
se perciba  
alguna situación  
peligrosa

Participantes  
infractores  
serán  
sancionados

# R-Ladies en el mundo



## consortium

**206**

Chapters

**88280**

Members

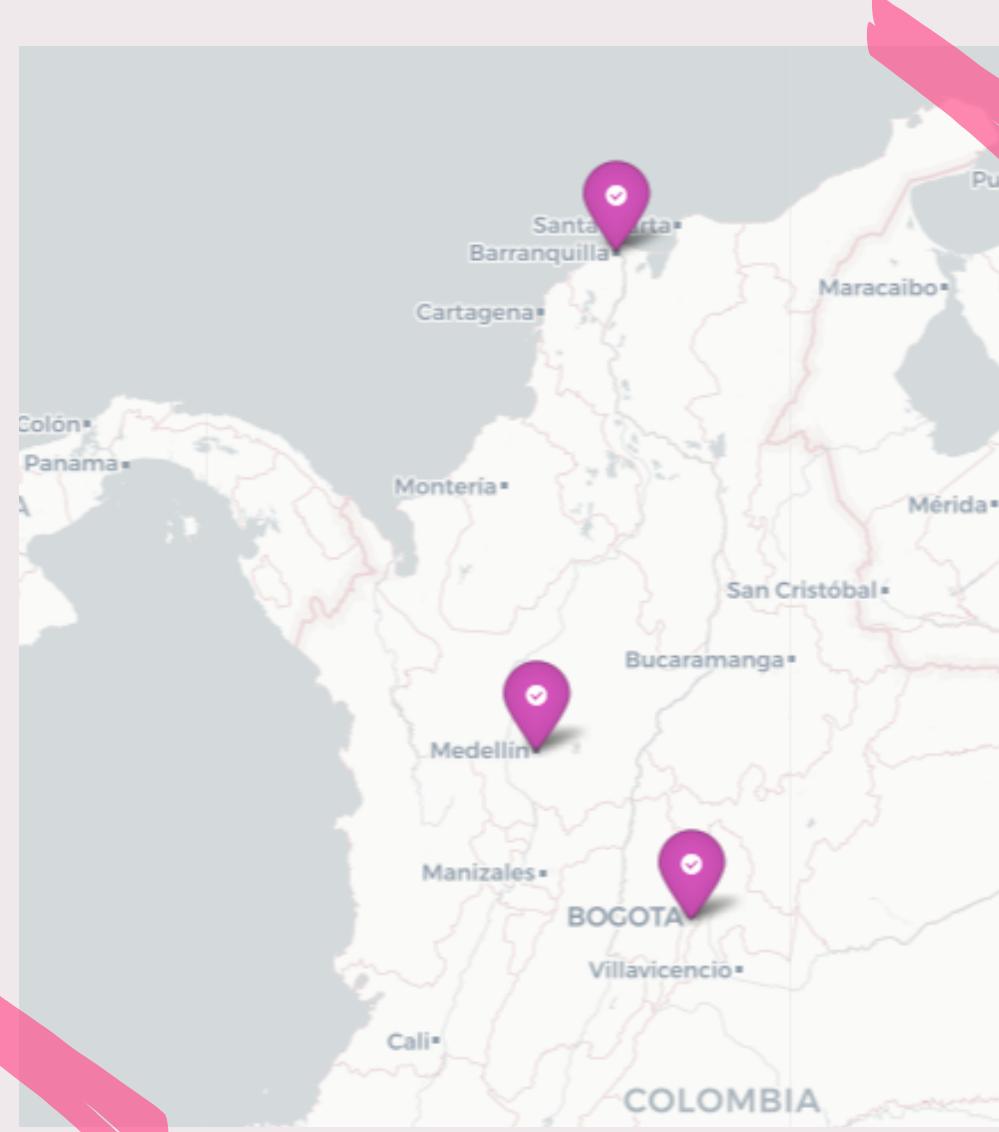
**57**

Countries



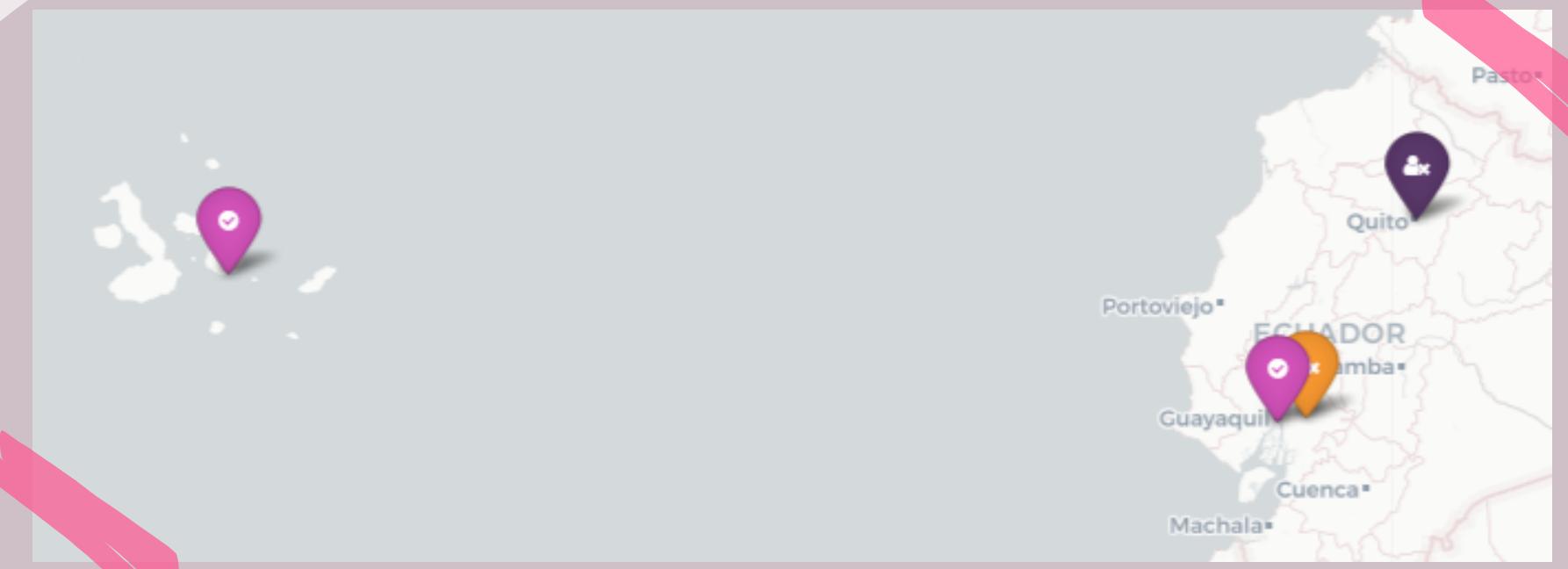
*R-Ladies*

# R-Ladies en

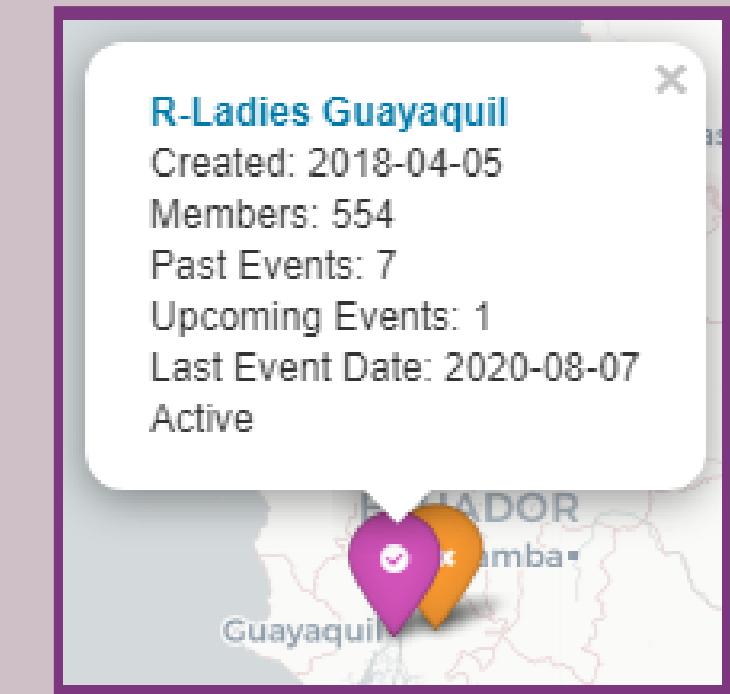
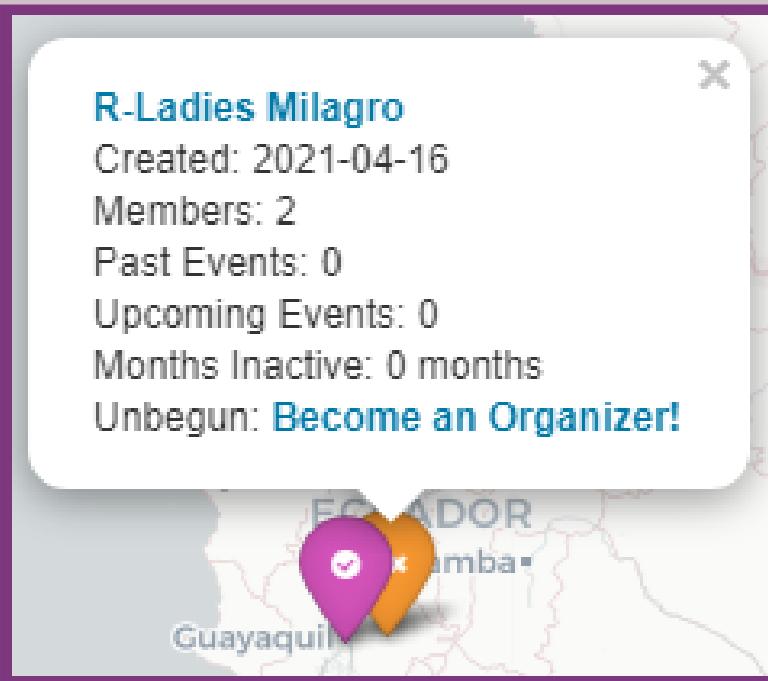
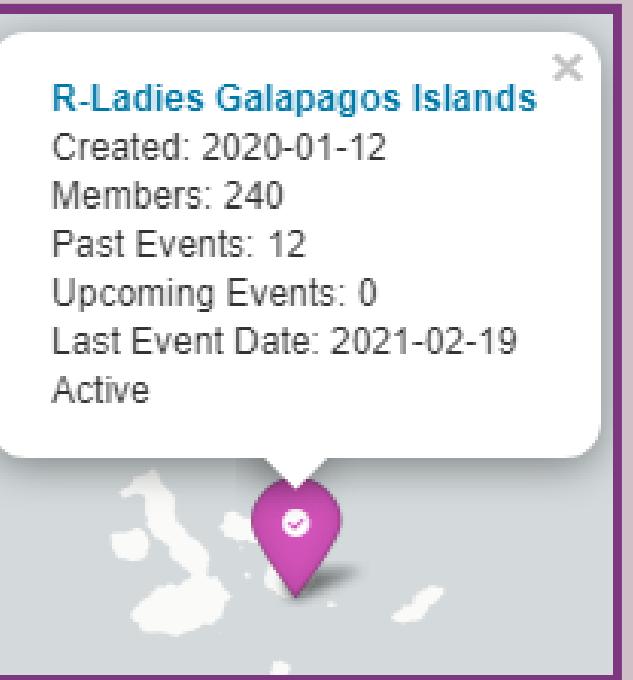
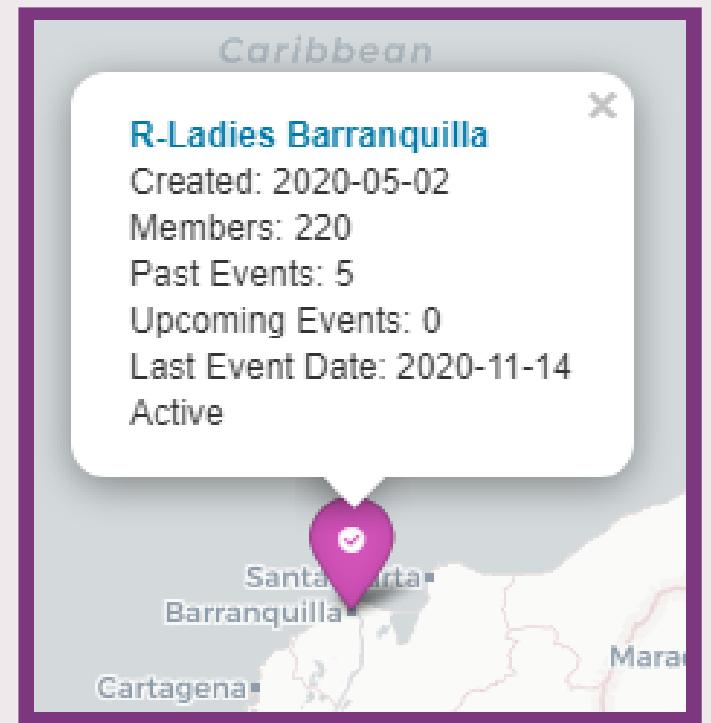


3 Chapters

4 Chapters



# R-Ladies en



# R-Ladies



@rladiesbquilla



[https://github.com/rladies/meetup-presentations\\_barranquilla](https://github.com/rladies/meetup-presentations_barranquilla)



@RLadiesBquilla



R-Ladies Barranquilla



barranquilla@rladies.org



R-Ladies Barranquilla

R-Ladies  
Barranquilla



R-Ladies Barranquilla



[https://github.com/rladies/meetup-presentations\\_galapagos-islands](https://github.com/rladies/meetup-presentations_galapagos-islands)



R-Ladies Ecuador



galapagos@rladies.org



R-Ladies  
Galápagos



@rladiesgps



R-Ladies Galápagos



@Rladies\_Milagro



[https://github.com/rjladies/meetup-presentations\\_milagro](https://github.com/rjladies/meetup-presentations_milagro)



*R-Ladies  
Milagro*



R-Ladies Ecuador



[milagro@rladies.org](mailto:milagro@rladies.org)



R-Ladies Milagro



[https://github.com/rjladies/meetup-presentations\\_guayaquil](https://github.com/rjladies/meetup-presentations_guayaquil)



R-Ladies Ecuador



[guayaquil@rladies.org](mailto:guayaquil@rladies.org)



*R-Ladies  
Guayaquil*



@rladiesgye



R-Ladies Guayaquil

# Expositoras



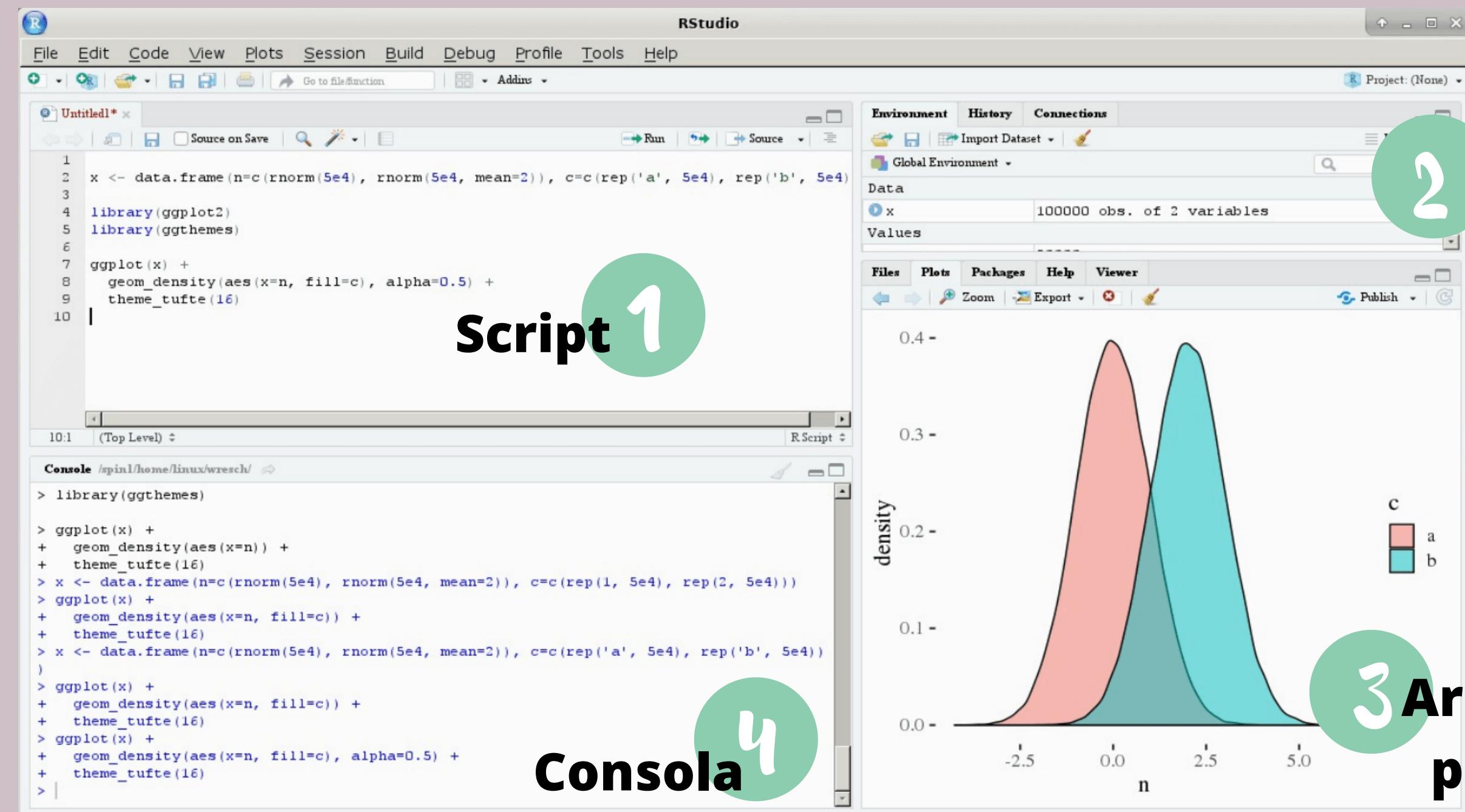
Danisse Carrascal



Viviana Flórez

# El entorno de R Studio®

La pantalla se divide principalmente en 4 paneles o ventanas.

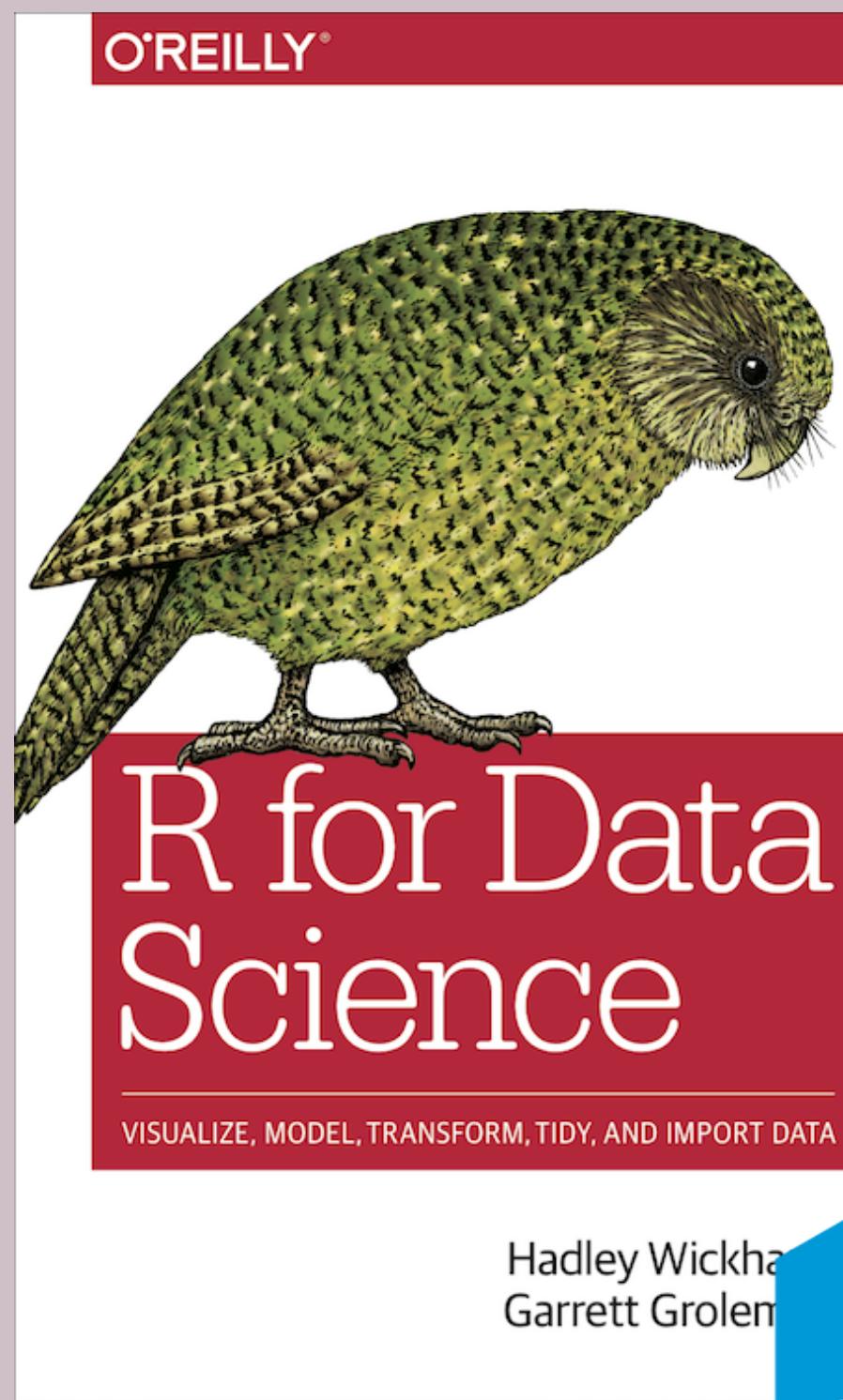


Script 1

Consola

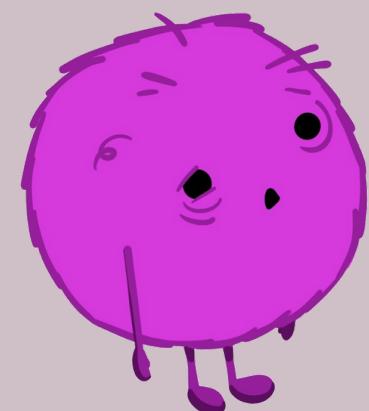
2 Entorno de trabajo

3 Archivos/gráficos/  
paquetes/ayuda

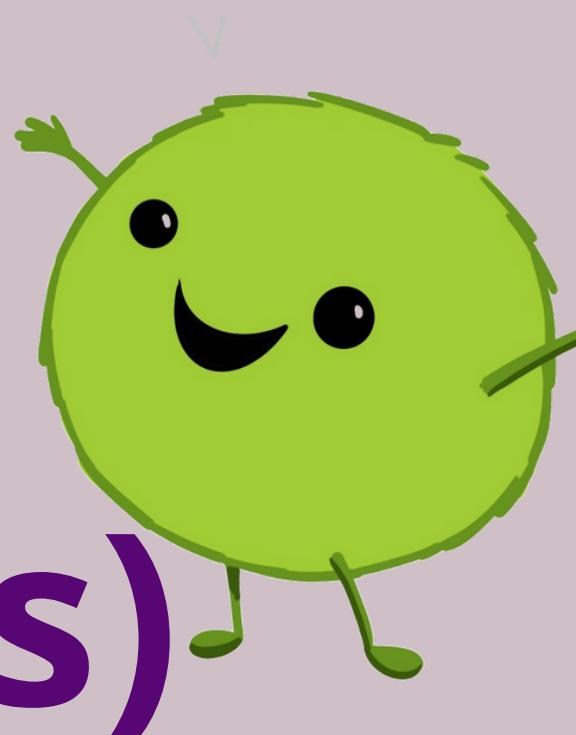


# Bienvenidas (os)

¡Esperamos que aprendamos mucho juntos!



R para ciencia de datos  
en español



## Capítulo 23. Modelos: conceptos básicos

Los modelos nos ayudan a capturar ciertos patrones en nuestros datos. El objetivo es conocer el valor de una variable a partir de otras variables explicativas.

El modelo que mejor se ajuste a nuestros datos no siempre significa que en realidad tengamos un buen modelo, o que ese modelo es verdadero.



## Hipótesis de una RLS

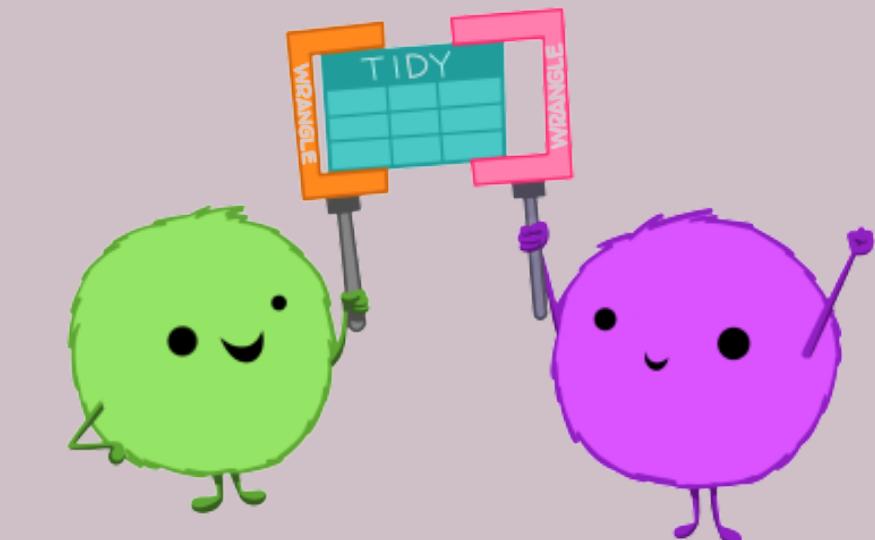
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon.$$

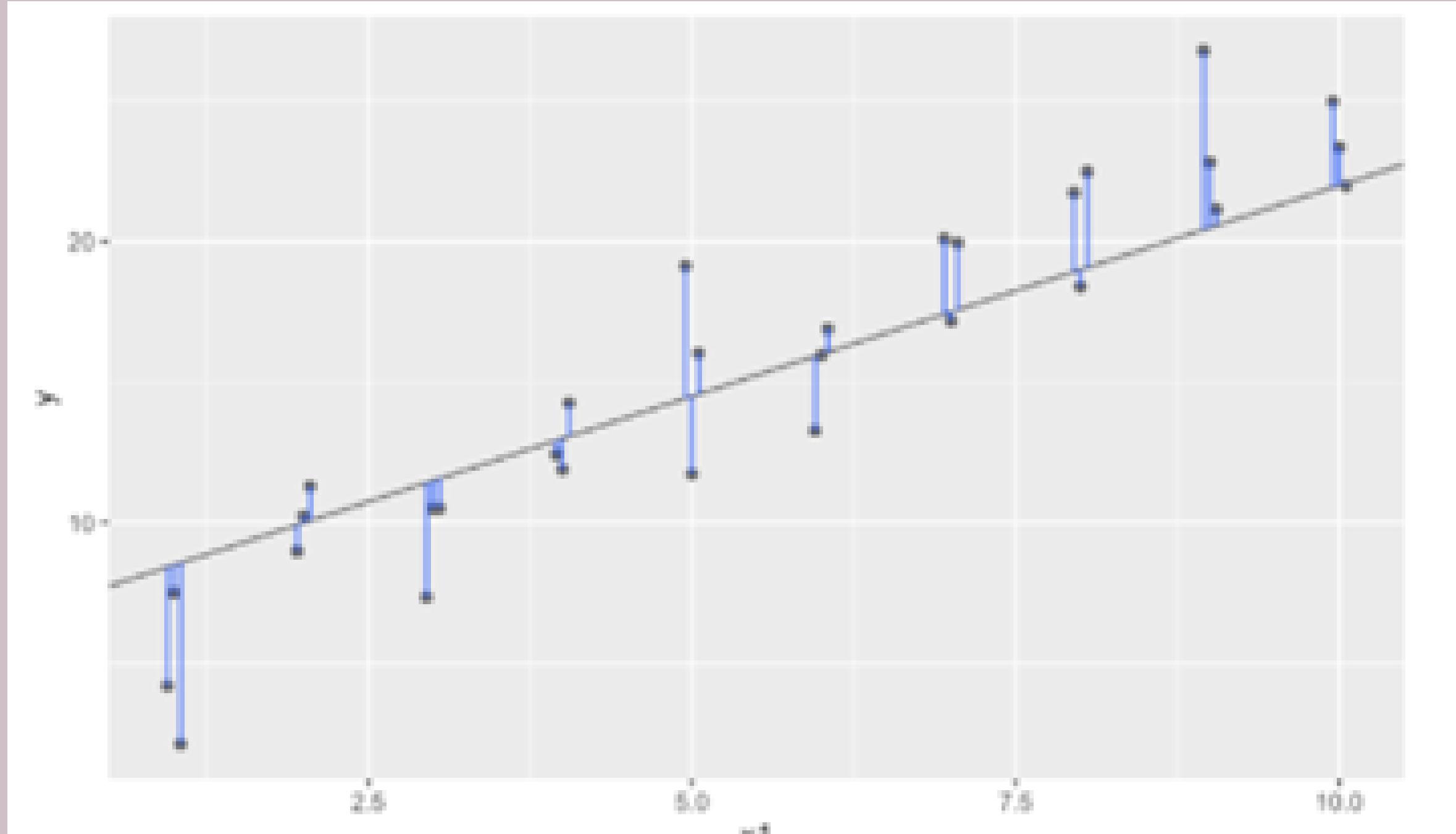
**H0:** todos los coeficientes del modelo son iguales a cero

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \epsilon_i.$$

**H1:** al menos uno de los coeficientes es distinto de cero

También se quiere que los residuos sean aleatorios y que se ajusten a la distribución normal.





## Hipótesis de una RLS

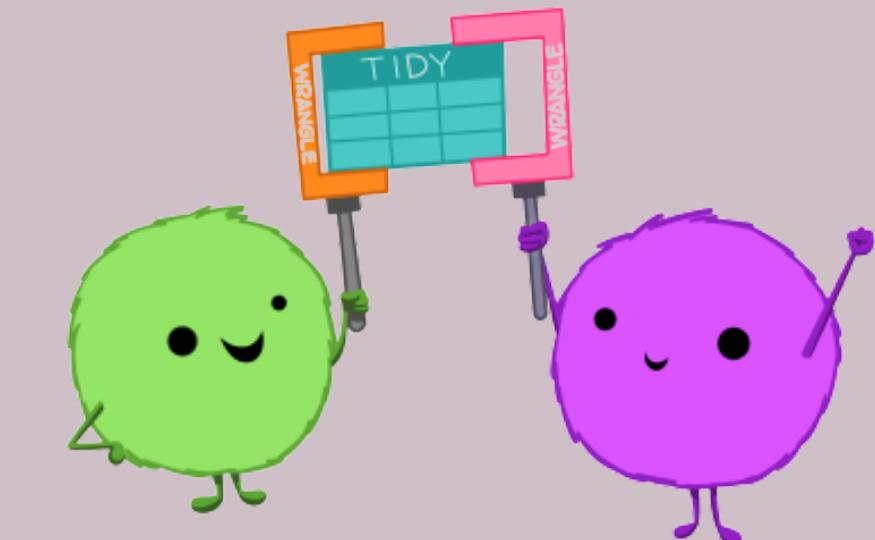
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon.$$

**H0:** todos los coeficientes del modelo son iguales a cero

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \epsilon_i.$$

**H1:** al menos uno de los coeficientes es distinto de cero

También se quiere que los residuos sean aleatorios y que se ajusten a la distribución normal.



# *Transformaciones*

Es posible realizar transformaciones dentro de la fórmula del modelo. Estas transformaciones son útiles porque podemos aproximar funciones no lineales.

## *Transformación logarítmica*

Podemos usar este tipo de transformación cuando vemos que nuestros datos no se ajustan a un modelo lineal. También podemos intentar, una vez que realizamos el modelo y observamos que no se cumplen los supuestos de los residuos.



# Familias de modelos

## Modelos lineales generalizados

`stats::glm()`

Extienden los modelos lineales para incluir respuestas no continuas.

## Modelos lineales penalizados

`glmnet::glmnet()`

Incorporan un término de penalización a la distancia y así penalizan modelos complejos

## Modelos generalizados aditivos

`mgcv::gam()`

Extienden los modelos lineales generalizados para incorporar funciones suaves arbitrarias



# Familias de modelos

## Árboles

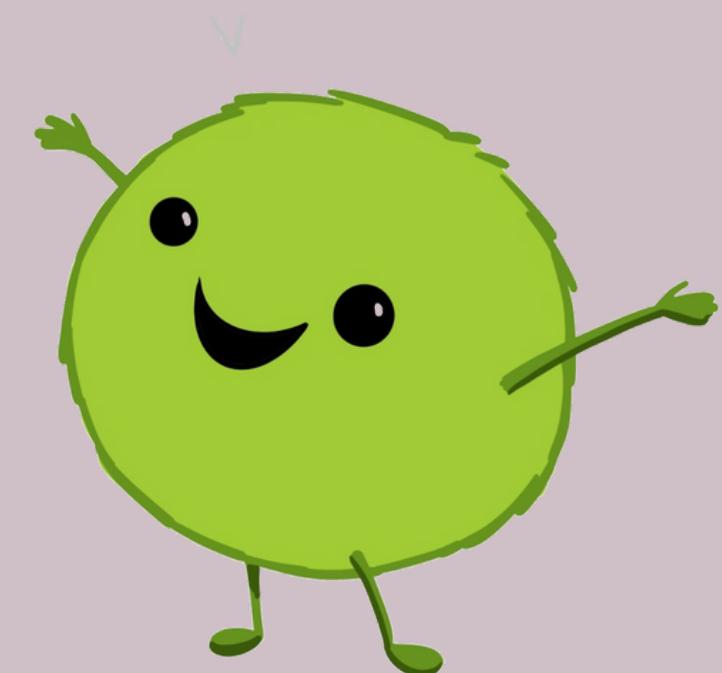
**rpart::rpart()**

Ajustan un modelo constante por partes, dividiendo los datos en partes progresivamente más y más pequeñas. Bosques aleatorios (del inglés random forests) o máquinas aceleradoras de gradiente, del inglés gradient boosting machines (es decir, xgboost::xgboost.).

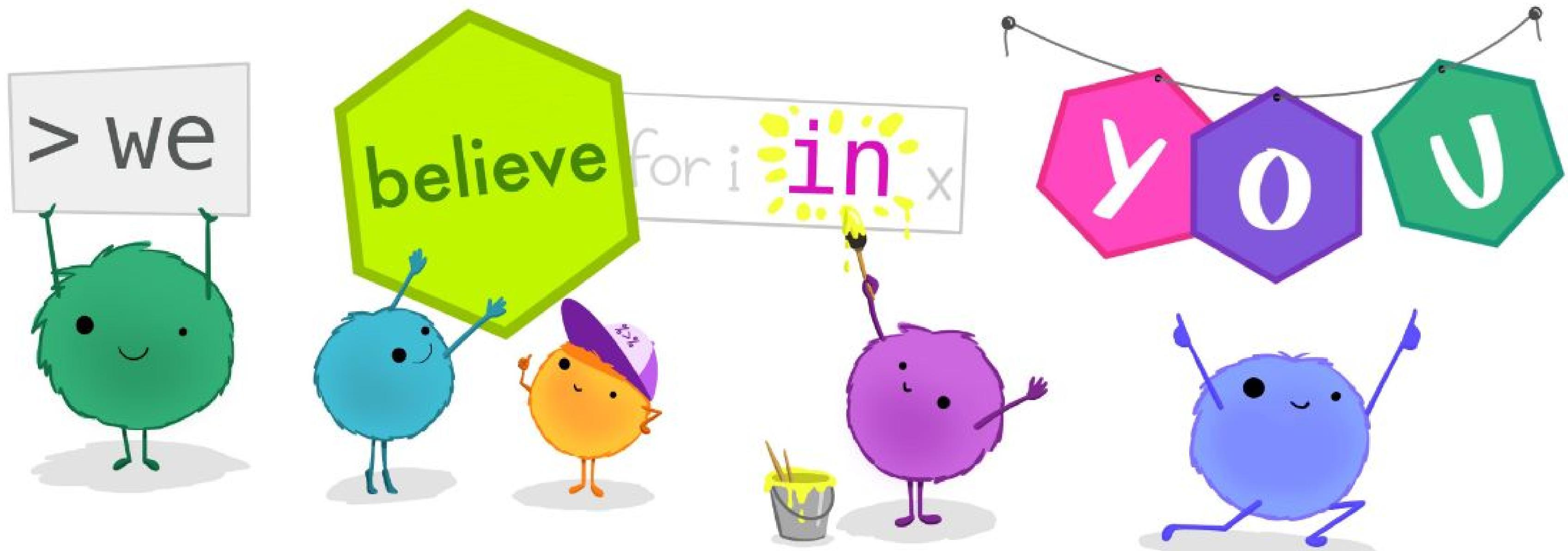
## Modelos lineales robustos

**MASS::rlm()**

Modifican la distancia para restar importancia a los puntos que quedan muy alejados



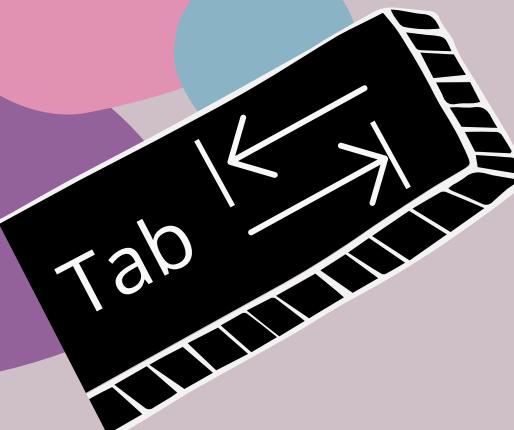
# remote R learners,



@allison\_horst

By: Allison Horst  
Via: @allison\_horst

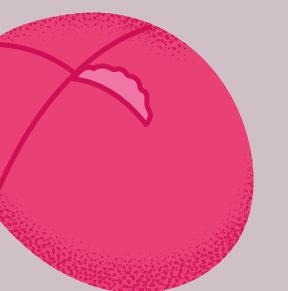
install.packages()



??function

help()

*Hagamos  
un poco de  
a R te*



```
datos <- read.csv2("RLadies.csv")  
View(datos)
```

Ctrl + Enter