



```
library(dplyr)
rladies_global %>%
filter(city=="Barranquilla" | city== "Galápagos")
```



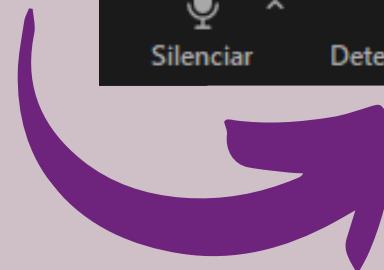
Manipulación de datos con dplyr y visualización con ggplot2

¡Conoce **Zoom**!



R-Ladies

¡Conoce zoom!



R-Ladies

GitHub

The screenshot shows a GitHub repository page for the user 'isavasquez' with the repository name 'Meetup-presentaciones_barranquilla'. The page includes a navigation bar with links like 'Solicitudes de extracción', 'Cuestiones', 'Mercado', and 'Explorar'. Below the navigation bar, there are statistics: 'Reloj 0 0', 'Estrella 0 0', and 'Tenedor 1'. The main content area shows a commit from 'isavasquez' and files like 'README.md'. A dropdown menu titled 'Código' is open, showing options for cloning with SSH or HTTPS, and download links for 'Abrir con el escritorio de GitHub' and 'Descargar ZIP'. A large green circle with the number '1' points to the 'Código' button. A large green circle with the number '2' points to the 'Descargar ZIP' link. A red arrow points from the text 'Click en "Descargar ZIP"' to the 'Descargar ZIP' link. Another red arrow points from the text 'Seleccionar la opción de descargar código' to the 'Código' button.

Buscar o saltar a ... / Solicitudes de extracción Cuestiones Mercado Explorar

isavasquez / Meetup-presentaciones_barranquilla

bifurcado de rladies / meetup-presentaciones_barranquilla

Código Solicitudes de extracción Comportamiento Proyectos Wiki Seguridad Percepciones

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](#)

Use una clave SSH protegida por contraseña.

git@github.com:isavasquez/meetup-prese

Abrir con el escritorio de GitHub Descargar ZIP

1

2

Meetup-presentaciones_barranquilla

2 commits

isavasquez cometió 4fc4d3e hace 16 horas

Agregar archivos mediante carga

README.md

Compromiso inicial

README.md

Lanzamientos

No hay publicaciones publicadas

Idiomas

R 100.0%

Selección de descarga

www.meetup.com/r ladies-barranquilla

Léame

Click en "Descargar ZIP"

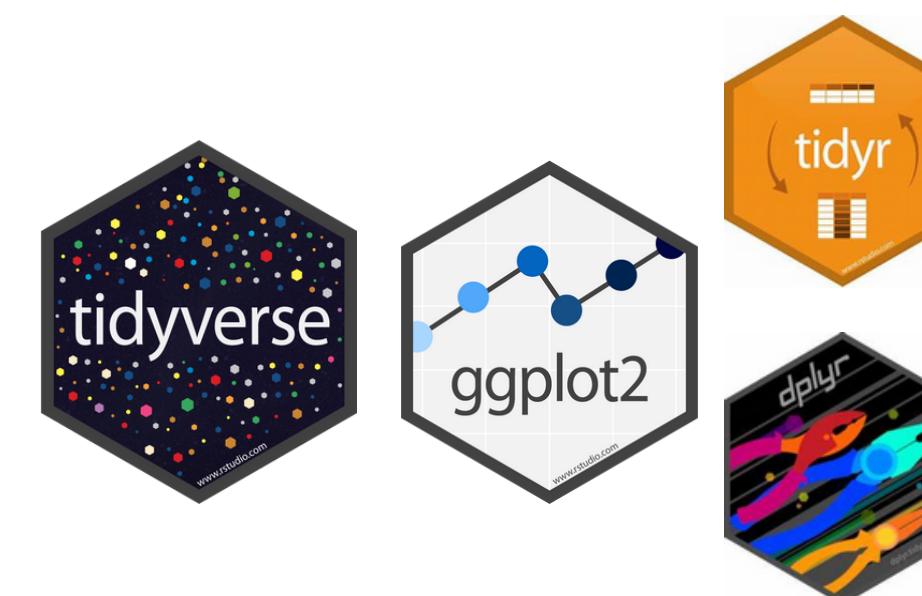
</> Paquetes </>



dplyr: es un paquete con distintas funciones que permiten 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

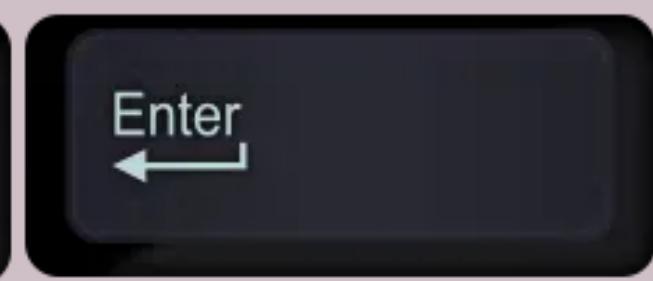
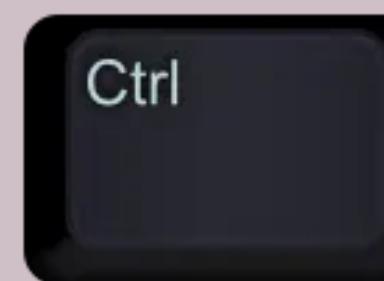
tidyverse: facilita la manipulación y organización de un dataset



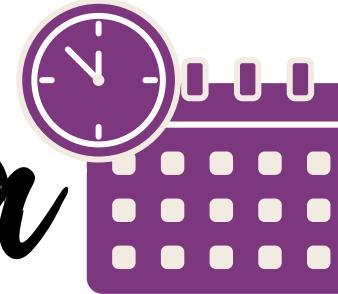
¿CÓMO INSTALAR UN PAQUETE EN R Studio® ?



```
File Edit Code View Plots Session Build Debug Profile Tools Help
+ R Go to file/function Addins
Untitled1* Run Source
Source on Save
1 -----
2 Encuentro R-Ladies Galápagos y Barranquilla
3 Sábado, Octubre 10 2020
4 Hagamos aRte en RStudio
5 -----
6 #Instalando los paquetes necesarios
7 #Forma 1:
8 install.packages(c("dplyr","tidyverse"))
9 #Forma 2:
10 install.packages("tidyverse")
11
```



Orden del día



- ✓ R-Ladies.
- ✓ ¿Cómo iniciar sesión en RStudio Cloud?
- ✓ ¿Cómo subir archivos a RStudio Cloud?
- ✓ Funciones en dplyr, tidyr y ggplot2
- ✓ Ejercicio 1 dplyr
- ✓ Ejercicio 2 dplyr y ggplot2



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, conferenciantes, participantes de la conferencia, educadores o usuarios (ver estadísticas recientes).

Gabriela de Queiroz



Fundadora R-Ladies Global.

Código de conducta

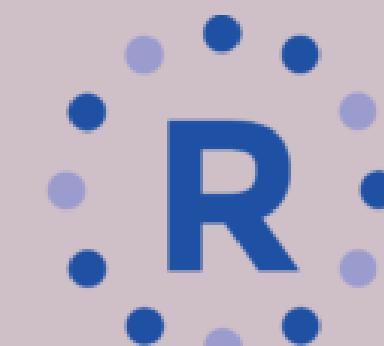
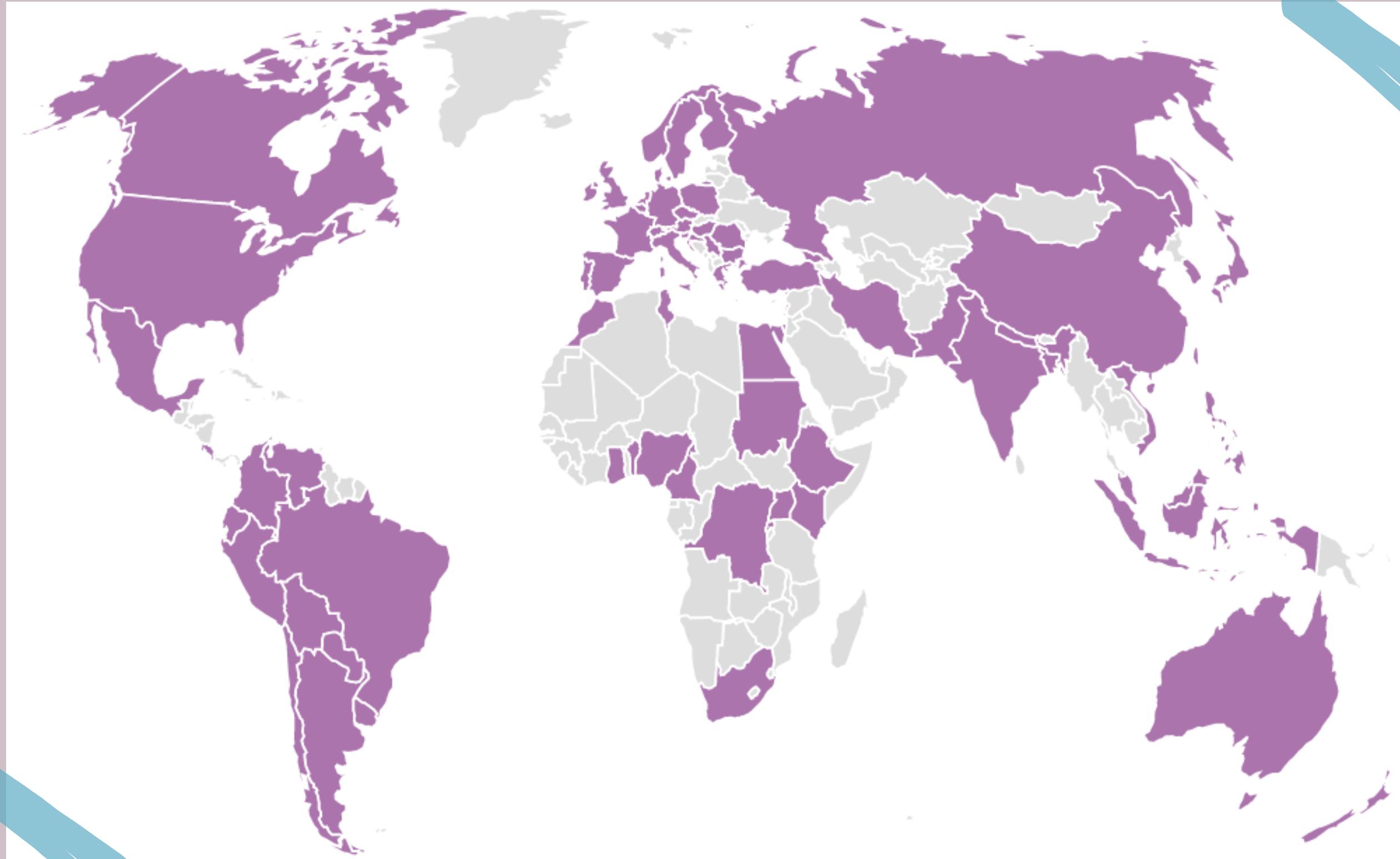
Espacio libre de acoso

Todas(os)
debemos ser
tratados con
respeto

Alertar cuando
se perciba
alguna situación
peligrosa

Participantes
infractores
serán
sancionados

R-Ladies en el mundo



consortium

193

Chapters

75333

Members

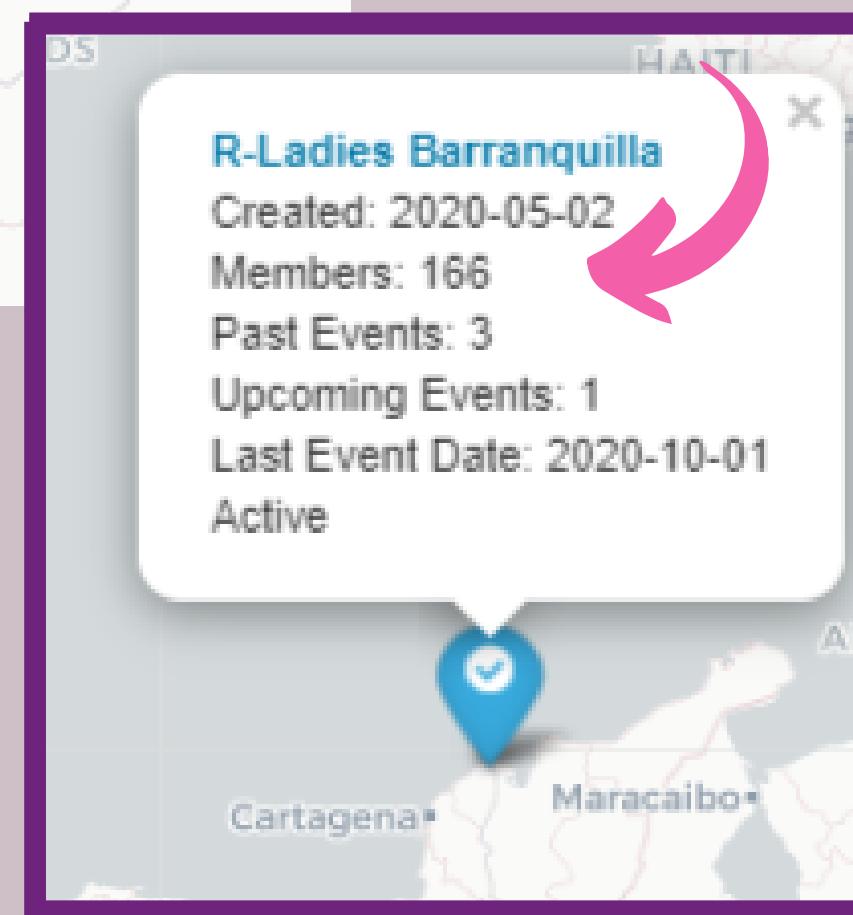
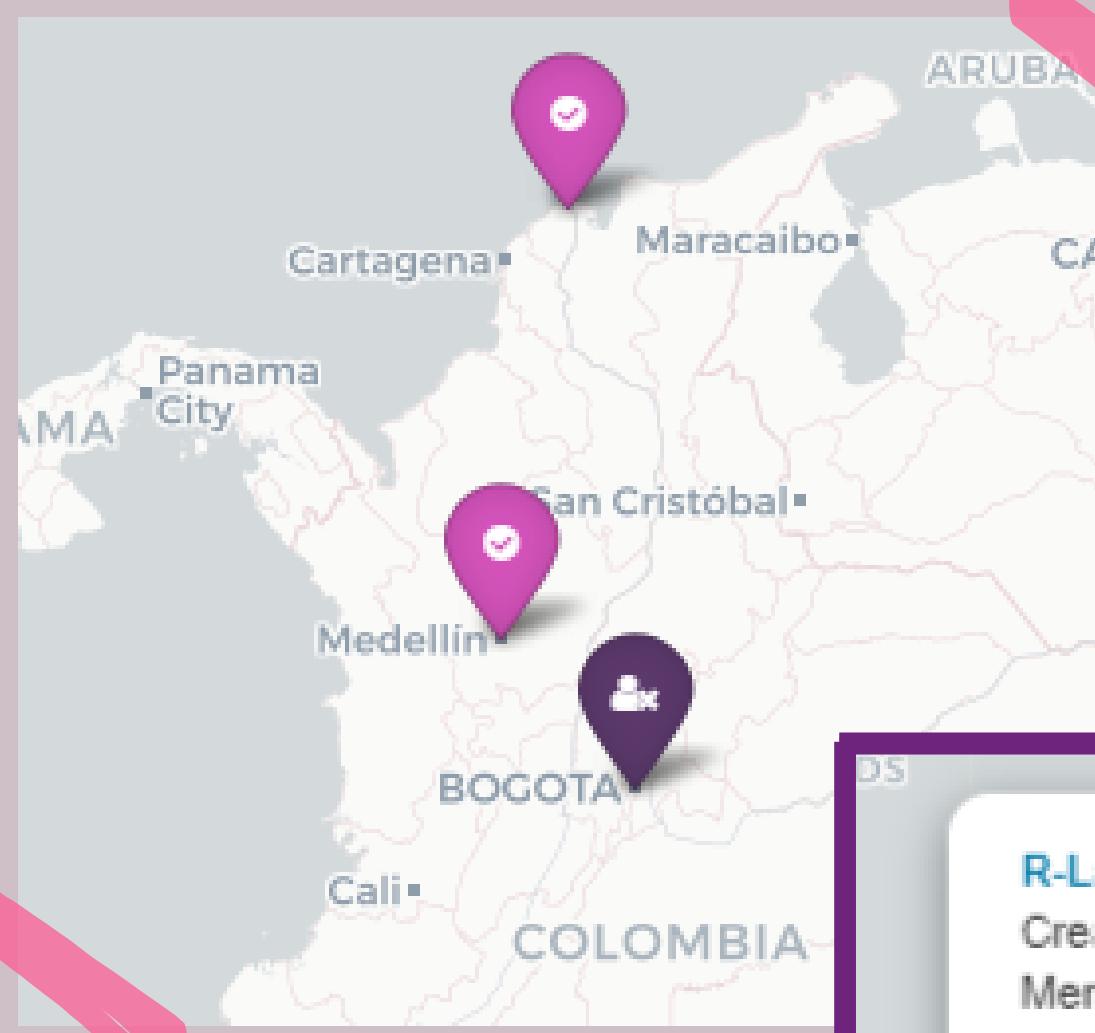
54

Countries

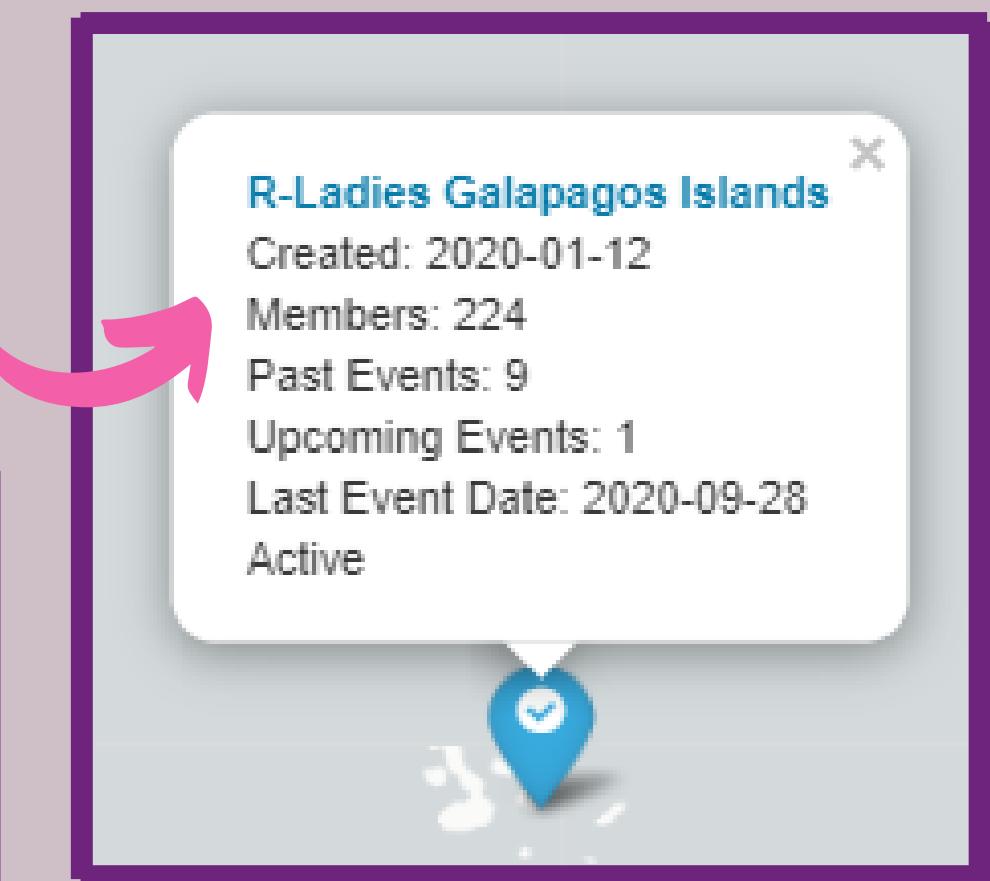
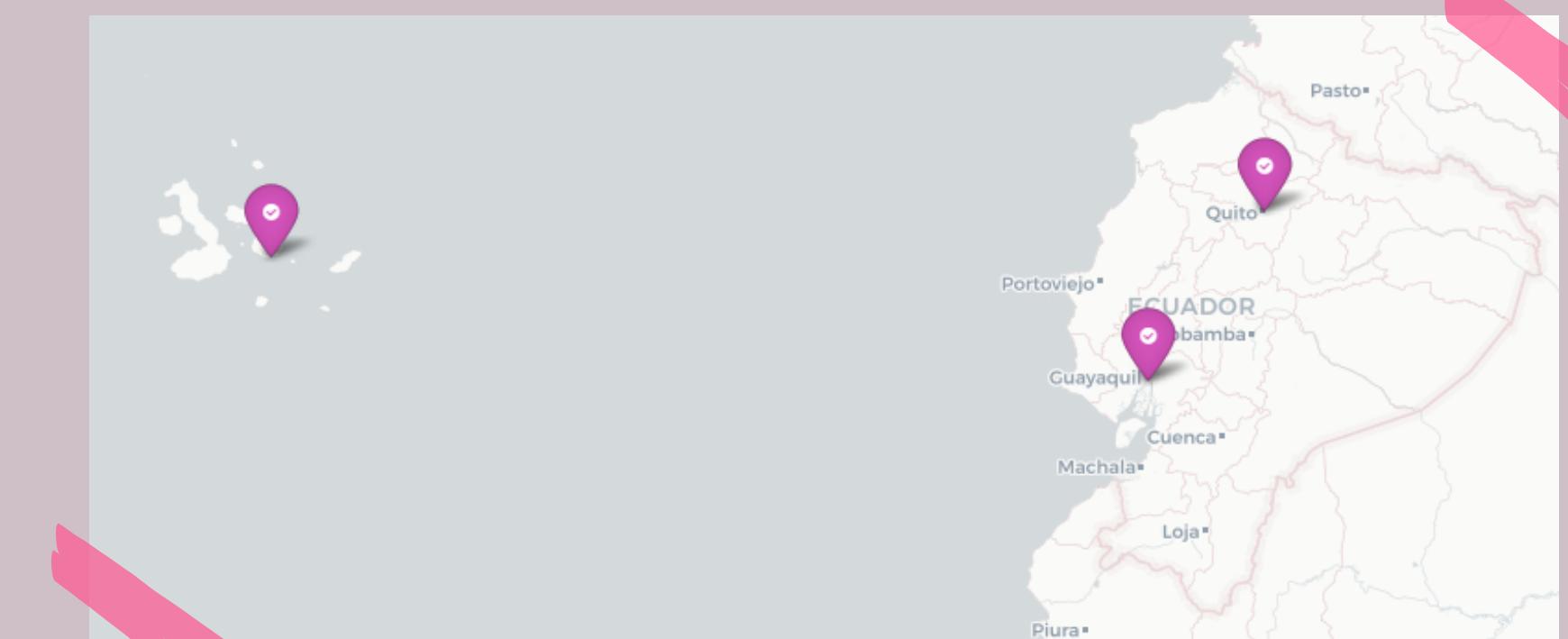


R-Ladies

R-Ladies en



R-Ladies en



R-Ladies



R-Ladies Barranquilla



@rladiesbquilla



https://github.com/rladies/meetup-presentations_barranquilla



barranquilla@rladies.org

R-Ladies Barranquilla



R-Ladies Barranquilla



R-Ladies Galápagos



https://github.com/rladies/meetup-presentations_galapagos-islands



galapagos@rladies.org



@rladiesgps



R-Ladies Galápagos



R-Ladies Ecuador

Equipo organizador



Maria Isabel Arrieta



Denisse Fierro



Isabel Vasquez



Viriana Florez



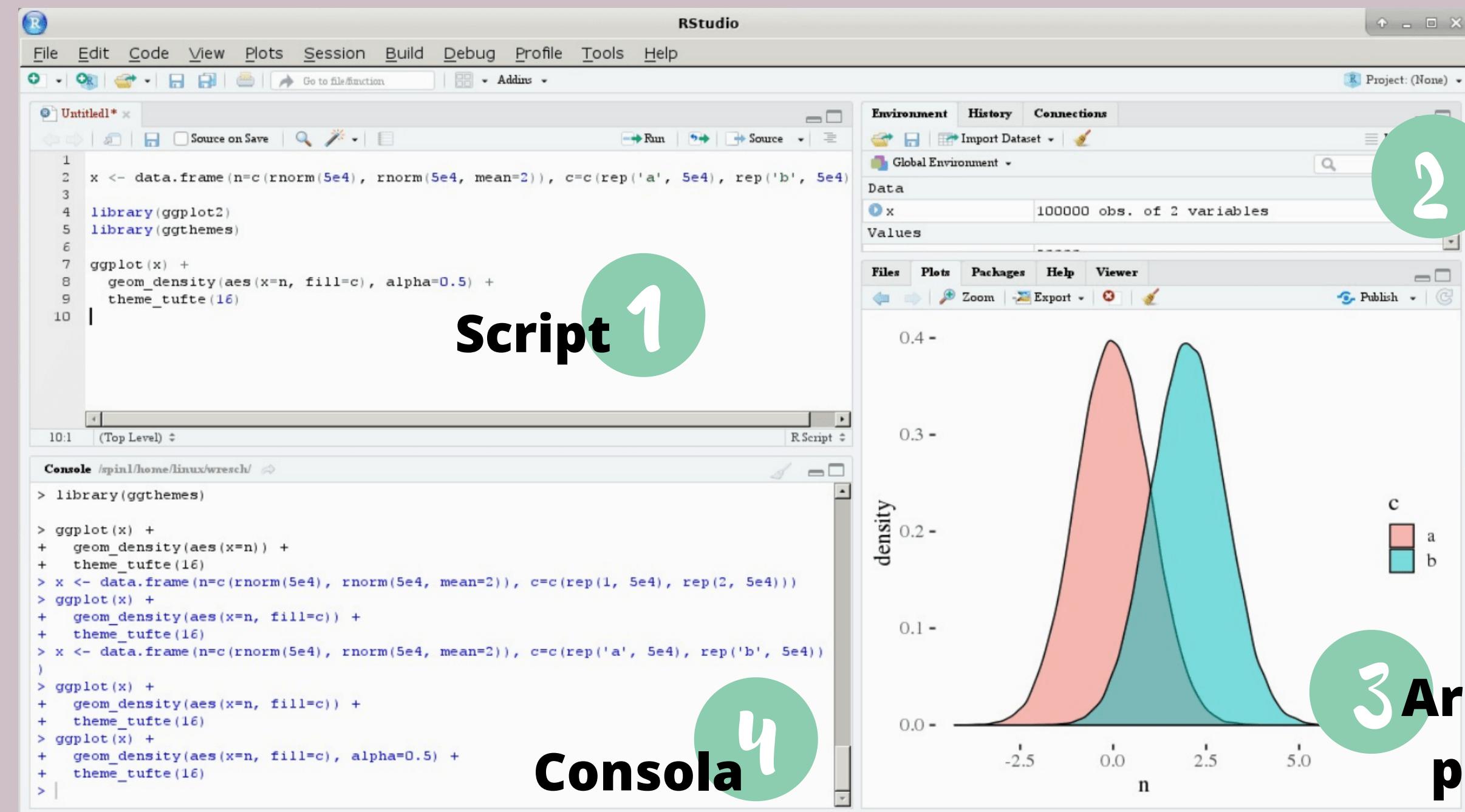
Danisse Carrascal



Mary Jane Rivero

El entorno de R Studio®

La pantalla se divide principalmente en 4 paneles o ventanas.

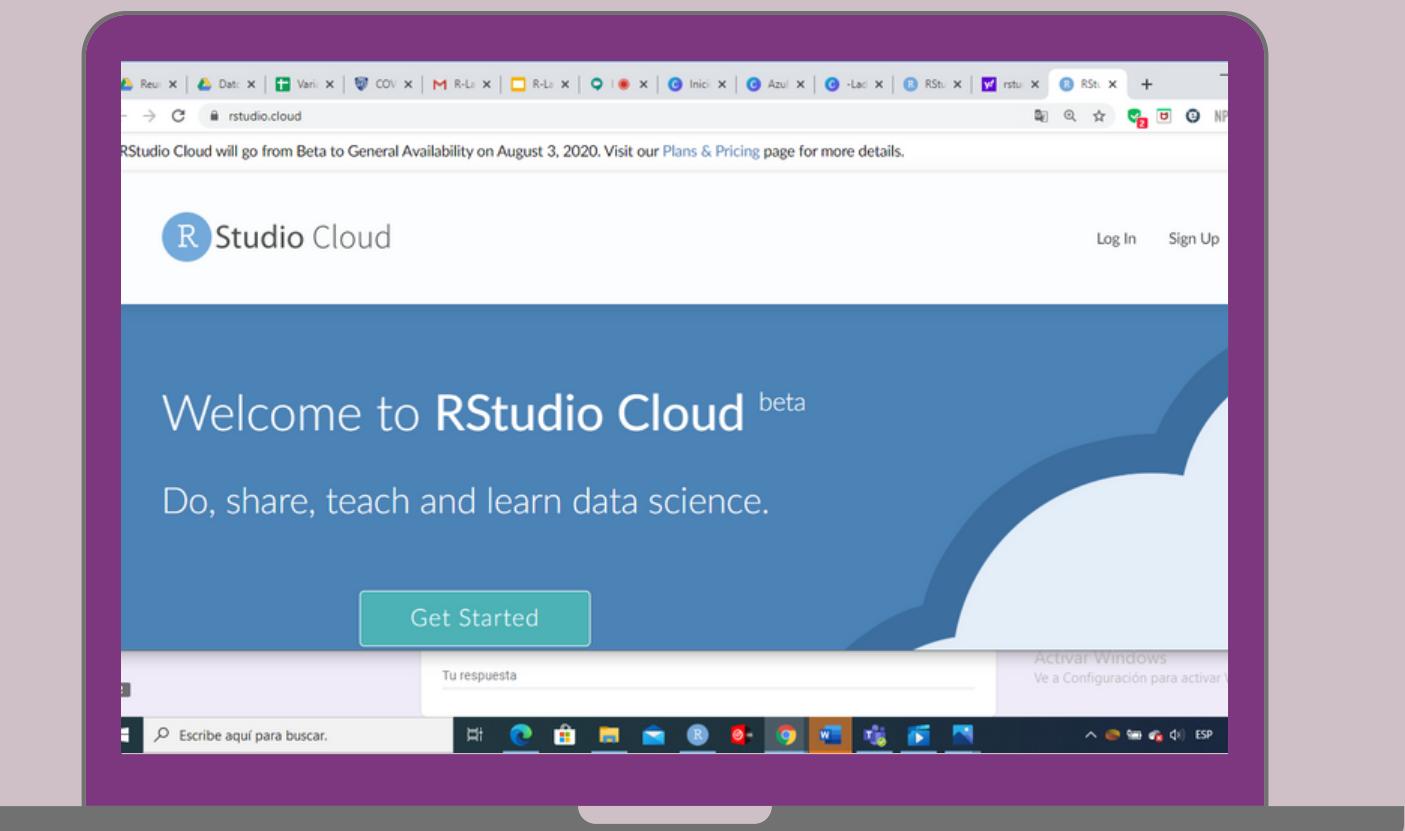


¿Cómo iniciar sesión



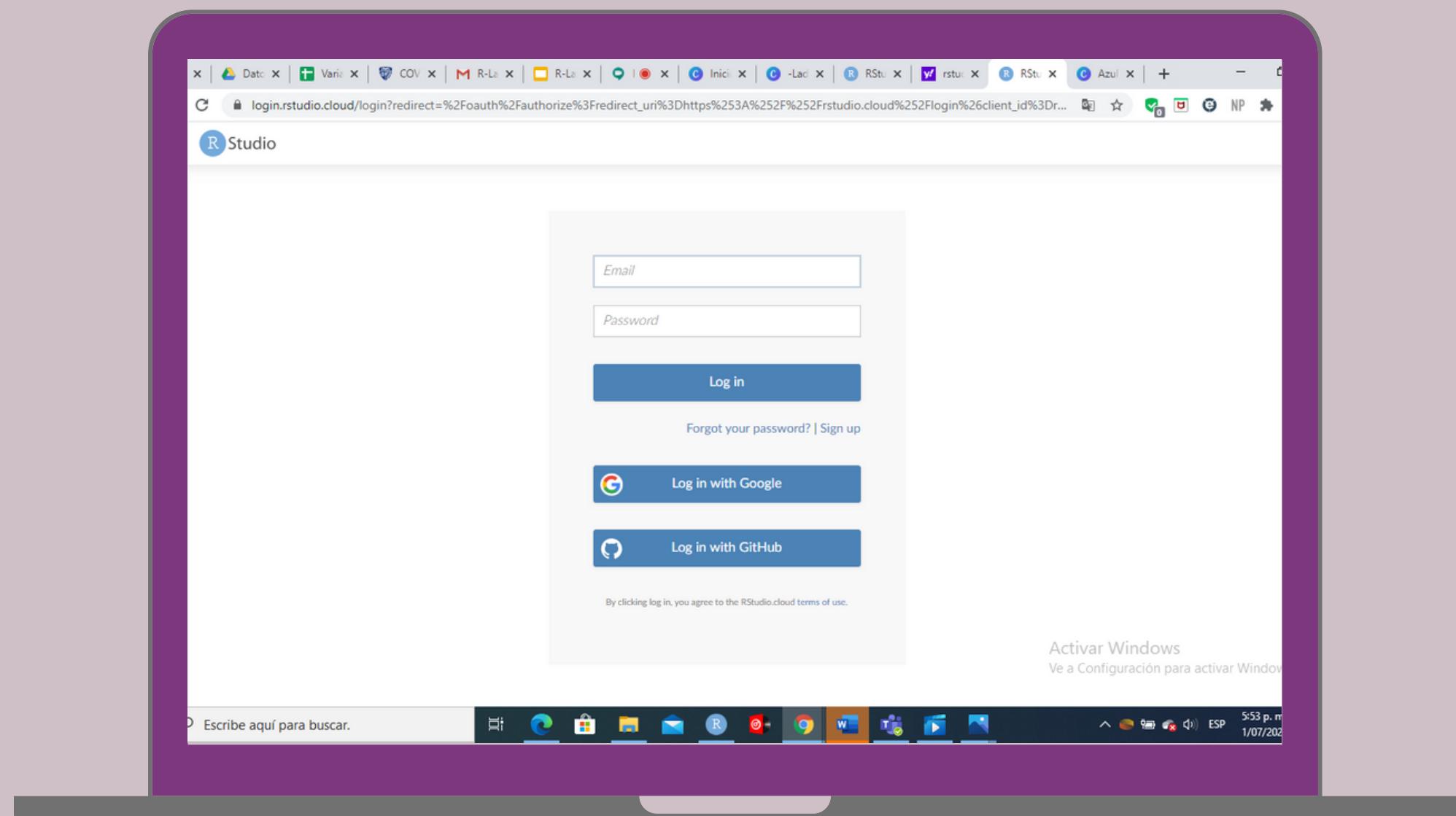
Paso 1:

Ingresá al link
<https://rstudio.cloud/>



Paso 2:

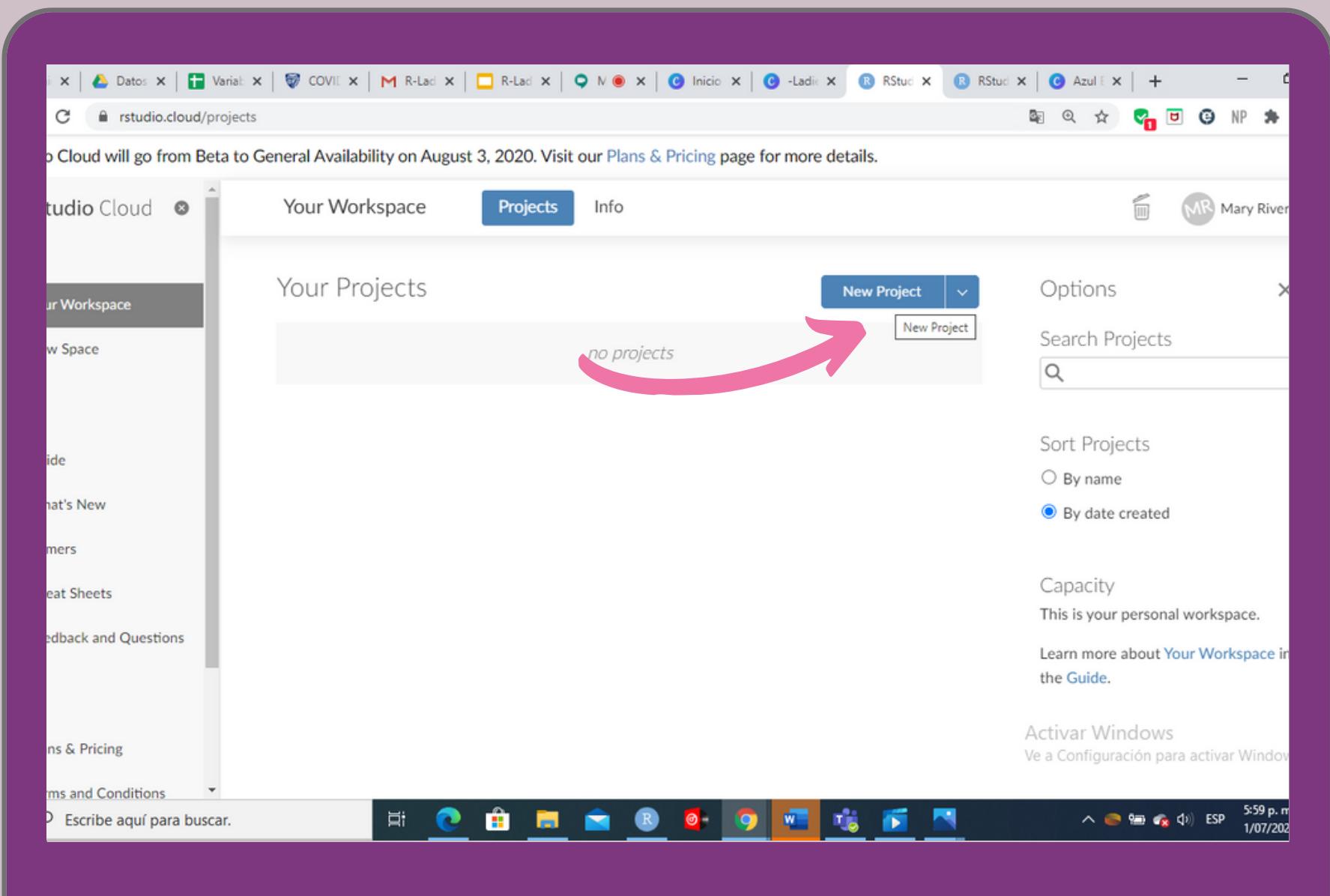
Regístrate o inicia
sesión con una cuenta
ya existente.



¿Cómo iniciar sesión

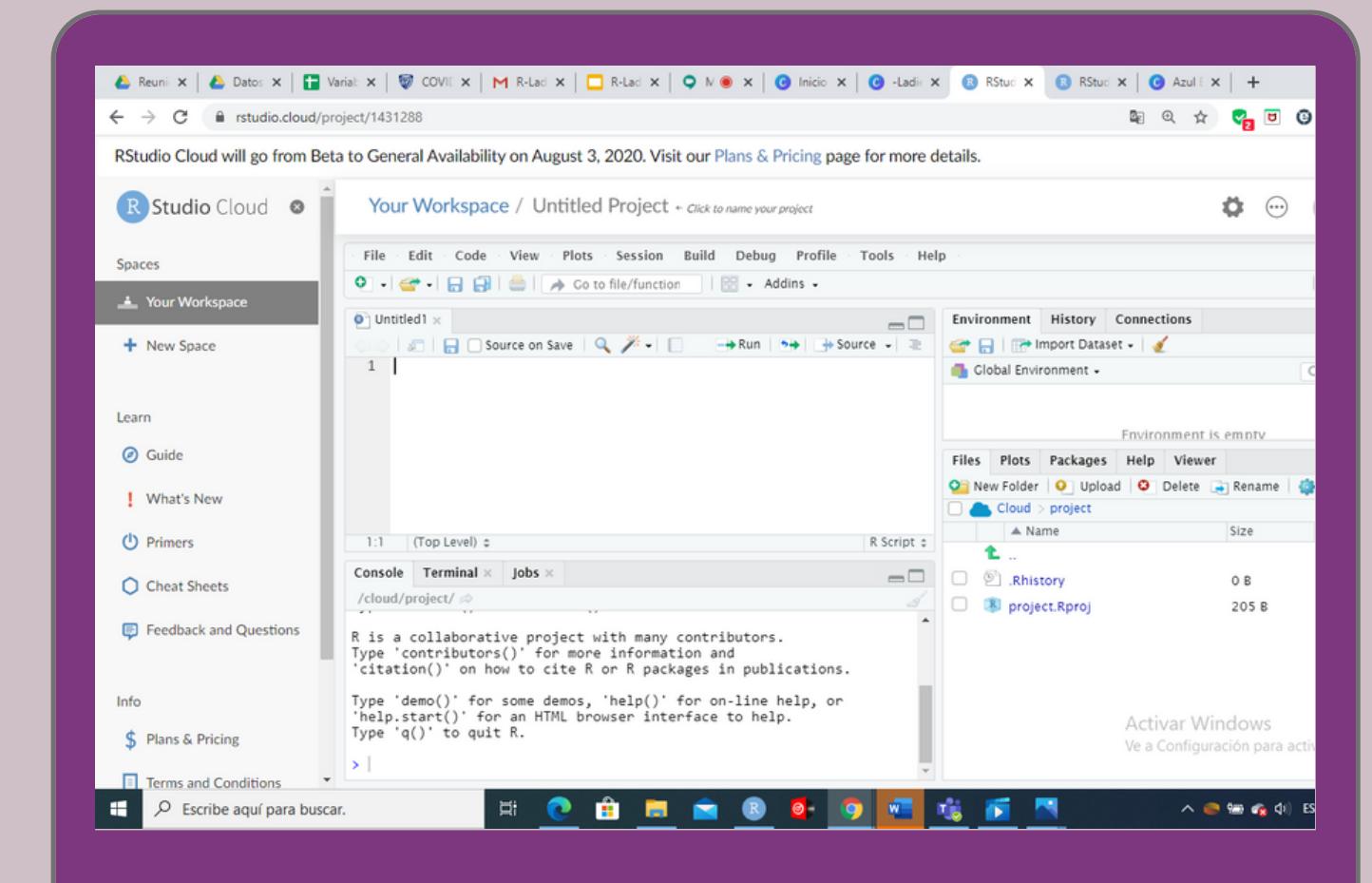


Paso 3: Crea un nuevo proyecto.

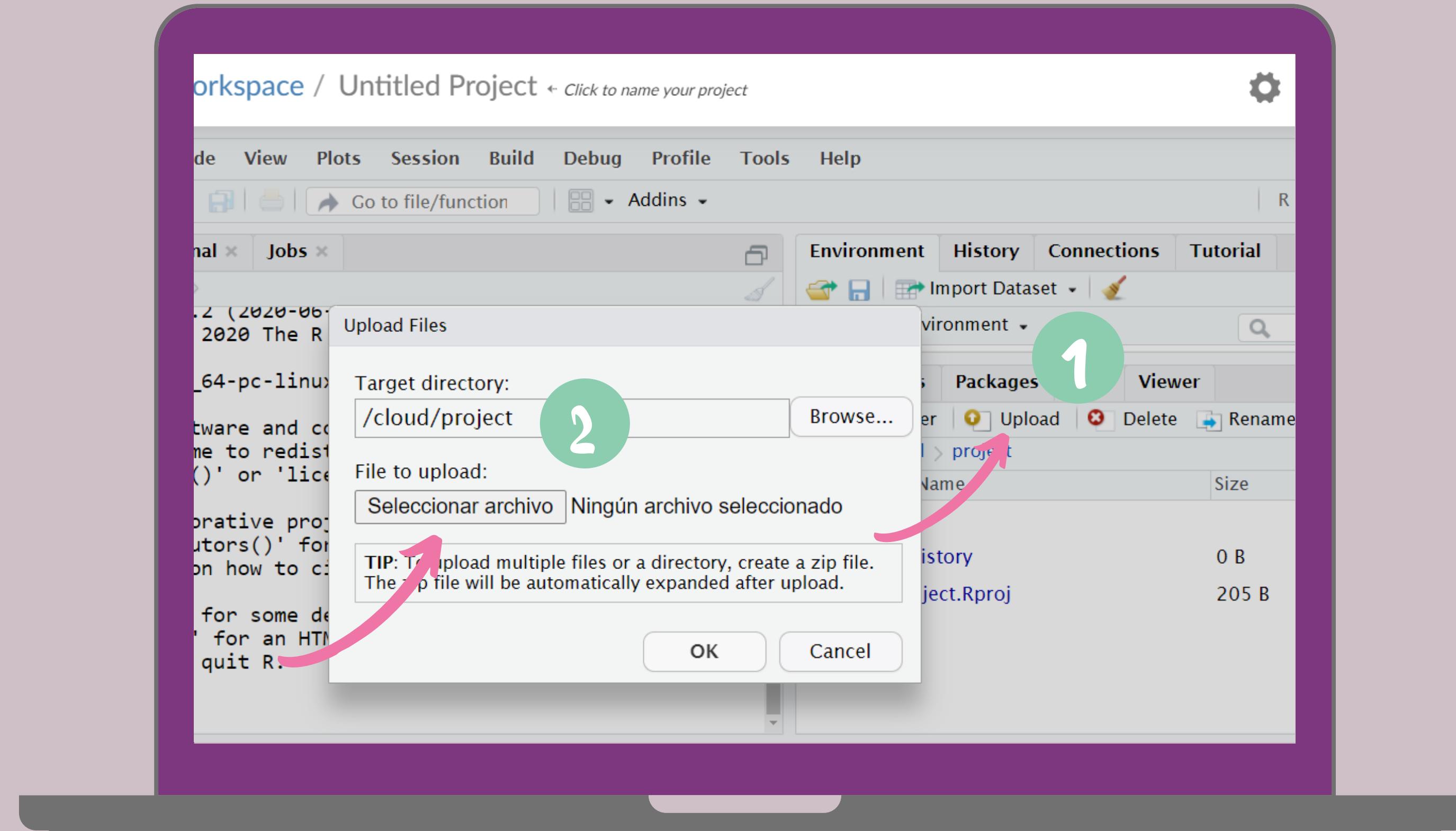


Paso 4:

Empieza a disfrutar de
RStudio Cloud.



¿Cómo subir archivos en R Studio Cloud?





*Hadley
Wickham*

- Proporciona una "gramática" para la manipulación de datos y para operar en los marcos de datos.
- Opera de forma rápida



¿Base de datos o informe?

Fecha	Producto	Precio	Cantidad
20/02/2017	Manzana	10	1.0
20/02/2017	Pera	15	2.0
21/03/2017	Ciruela	13	0.5
22/03/2017	Naranja	9	3.0

Mes	Ventas
Febrero	40.0
Marzo	33.5

Datos rectangulares

Ventas				
Frutería “La desprolja”	NA	NA	NA	NA
Primer Trimestre	NA	NA	NA	NA
Manzanas	100	NA	NA	NA
Peras	120	NA	Ventas de fruta de carozo:	614
Ciruelas	85	NA	NA	NA
Segundo Trimestre	NA	NA	Ventas de frutas de pepita:	166
Manzanas	110	NA	NA	NA
Peras	40	NA	NA	NA
Ciruelas	25	NA	NA	NA
Tercer Trimestre	NA	NA	NA	NA
Manzanas	60	NA	NA	NA
Peras	45	NA	NA	NA
Ciruelas	22	NA	NA	NA

Fruta	Primer trimestre	Segundo Trimestre	Tercer Trimestre
Manzanas	100	110	60
Peras	120	40	45
Ciruelas	85	25	22

Ordenar los datos

Cada **columna** es una **variable** y
cada **fila** una **observación**.

¿Los datos están ordenados o
desordenados?

Raza	Indicador	2018	2019
Beagle	Precio	300.000	350.000
Pastor Alemán	Precio	400.000	340.000
Bulldog	Precio	350.000	400.000

¿Y ahora?

Raza	Año	Precio
Beagle	2018	300.000
Beagle	2019	350.000
Pastor Alemán	2018	400.000
Pastor Alemán	2019	340.000
Bulldog	2018	350.000
Bulldog	2019	400.000

Operador pipe %>%

Útil para concatenar
múltiples
operaciones de dplyr

El siguiente ejemplo muestra que cada vez que queremos aplicar mas de una función, la instrucción es una secuencia de llamadas a funciones de forma anidada y que resulta ilegible:

```
third(second(first(x)))
```

Este anidamiento no es una forma natural de expresar una secuencia de operaciones. El operador %>% nos permite escribir una secuencia de operaciones de izquierda a derecha:

```
first(x) %>% second(x) %>% third(x)
```

Pipe o pipeline.

Argumento.

Función.

dplyr::select

Selecciona las variables que
satisface.n
tu interés

```
library(dplyr)  
select(dataset, name, mass)
```

Seleccióna las variables.

Conjunto de datos

No olvides separar los argumentos con coma.

Variables de interés sin comillas.

name	age	mass
Mary Jane	21	68
Danisse	19	60
Michael	25	70

dplyr::glimpse

Conoce y explora las
c.a.r.a.c.t.e.rí.s.t.i.c.a.s
de tus variables

```
library(dplyr)  
glimpse(x = dataset)
```



Explora la estructura
de tus datos.

name	age	mass
Mary Jane	21	68
Danisse	19	60
Michael	25	70

```
Console Terminal × Jobs ×  
/cloud/project/ ↗  
>glimpse(x = dataset, width = NULL)  
Rows: 3  
Columns: 2  
$ name <chr> "Mary Jane", "Danisse", "Michael"  
$ mass <dbl> 68, 60, 70
```

dplyr::arrange

```
library(dplyr)  
arrange(x = dataset, mass)
```

↑
Ordena las filas
ascendentemente por
defecto.

↑
Conjunto de datos.
¡la coma!

↑
Variable a ordenar

name	mass
Danisse	60
Mary Jane	68
Michael	70

name	mass
Mary Jane	68
Danisse	60
Michael	70

name	mass
Michael	70
Mary Jane	68
Danisse	60

name	mass
Mary Jane	68
Danisse	60
Michael	70

Ordena las
f.i.l.a.s
de tus variables

dplyr::rename

```
library(dplyr)  
rename(Storms, tormenta =  
      storm, viento = wind, presion  
      = pressure, fecha = date)
```

Conjunto de datos

Cambia el nombre de tus columnas

¡la coma!

Nuevo nombre

Antiguo nombre

Renombra las variables de tu dataframe

	tormenta	viento	presion	fecha
	(chr)	(int)	(int)	(date)
1	Alberto	110	1007	2000-08-03
2	Alex	45	1009	1998-07-27
3	Allison	65	1005	1995-06-03
4	Ana	40	1013	1997-06-30
5	Arlene	50	1010	1999-06-11
6	Arthur	45	1010	1996-06-17

dplyr::mutate

Crea una nueva variable
a partir de variables existentes

```
library(dplyr)  
mutate(Storms, ratio = pressure/wind)
```

Genera una nueva
columna o
variable.

Conjunto de datos.

Nueva variable.

```
.after = "wind" | .before = "wind"
```

storm	wind	pressure	date
Alberto	110	1007	2000-08-12
Alex	45	1009	1998-07-30
Allison	65	1005	1995-06-04
Ana	40	1013	1997-07-01
Arlene	50	1010	1999-06-13
Arthur	45	1010	1996-06-21



storm	wind	pressure	date	ratio
Alberto	110	1007	2000-08-12	9.15
Alex	45	1009	1998-07-30	22.42
Allison	65	1005	1995-06-04	15.46
Ana	40	1013	1997-07-01	25.32
Arlene	50	1010	1999-06-13	20.20
Arthur	45	1010	1996-06-21	22.44

dplyr::filter

```
library(dplyr)  
filter(Storms, wind >= 50 &  
      pressure <= 1010)
```

Seleccióna tus filas.

Conjunto de datos

Condición 1

Condición 2

Operador lógico

Storms

storm	wind	pressure	date
Alberto	110	1007	2000-08-12
Alex	45	1009	1998-07-30
Allison	65	1005	1995-06-04
Ana	40	1013	1997-07-01
Arlene	50	1010	1999-06-13
Arthur	45	1010	1996-06-21



storm	wind	pressure	date
Alberto	110	1007	2000-08-12
Allison	65	1005	1995-06-04
Arlene	50	1010	1999-06-13

✓ ✗ ✓ ✗ ✓ ✗

Conserva las filas que
satisface.n
tu interés

Las condiciones pueden ser expresiones lógicas construidas mediante los operadores relacionales y lógicos:

Relacionales

<	Menor que
>	Mayor que
==	Igual que
<=	Menor igual que
>=	Mayor igual que
!=	Diferente que
%in%	Pertenece al conjunto
is.na	Es NA
!is.na	No es NA

Lógicos

&	Booleano y
\	Booleano o
xor	O inclusivo
!	No
any	Cualquier verdadero
all	Todos los verdaderos

dplyr::summarise()

```
library(dplyr)  
summarise(pollution, mediana =  
median(amount), varianza = var(amount))
```

Realiza operaciones y
crea un nuevo data
frame.

Conjunto de datos.

Nombre de la variable en el nuevo data frame.

pollution

city	particle size	amount ($\mu\text{g}/\text{m}^3$)
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56



median	varianza
22.5	1731.6

Realiza operaciones con variables existentes y crea un nuevo data frame

A continuación presentamos las funciones que trabajan conjuntamente con la función summarise(). Todas ellas toman como argumento un vector y devuelven un único resultado.

Paquete base

min(), max()

Valores mínimo
y máximo.

mean()

Media.

median()

Mediana.

sum()

Suma de los
valores.

var(), sd()

Varianza y
desviación
típica.

Paquete dplyr

first()

Primer valor
en un vector.

last()

El último
valor en un
vector.

n()

El número de
valores en un
vector.

n_distinct()

El número de
valores
distintos en
un vector.

nth()

Extrae el valor que ocupa
la posición n en un vector

dplyr::group_by()

Toma una tbl existente y crea una tbl nueva donde las operaciones se realizan por grupo

```
library(dplyr)  
group_by(pollution, city)
```

Agrupa por grupos.

Conjunto de datos.

Variable para agrupar.

pollution

city	particle size	amount ($\mu\text{g}/\text{m}^3$)
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

city	particle size	amount ($\mu\text{g}/\text{m}^3$)
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56

3

G
R
U
P
O
S

La función **group_by()** es extremadamente útil trabajando en conjunto con la función **summarise()**:

```
library(dplyr)  
pollution %>% group_by(city) %>%  
  summarise(mean = mean(amount), sum =  
    sum(amount), n = n())
```

	city	size	amount
1	New York	large	23
2	New York	small	14
3	London	large	22
4	London	small	16
5	Beijing	large	121
6	Beijing	small	56

	city	mean	sum	n
1	(chr)	(dbl)	(dbl)	(int)
2	Beijing	88.5	177	2
3	London	19.0	38	2
4	New York	18.5	37	2

tidy::gather()

Toma un dataset existente y convierte las columnas en valores u observaciones

```
library(dplyr)  
messy %>%  
  gather(drug, heartrate, a:b)
```

Base de datos.

Re-ordena el dataset

Nombre nuevas columnas

Columnas originales

```
messy  
#>   name  a  b  
#> 1 Wilbur 67 56  
#> 2 Petunia 80 90  
#> 3 Gregory 64 50
```

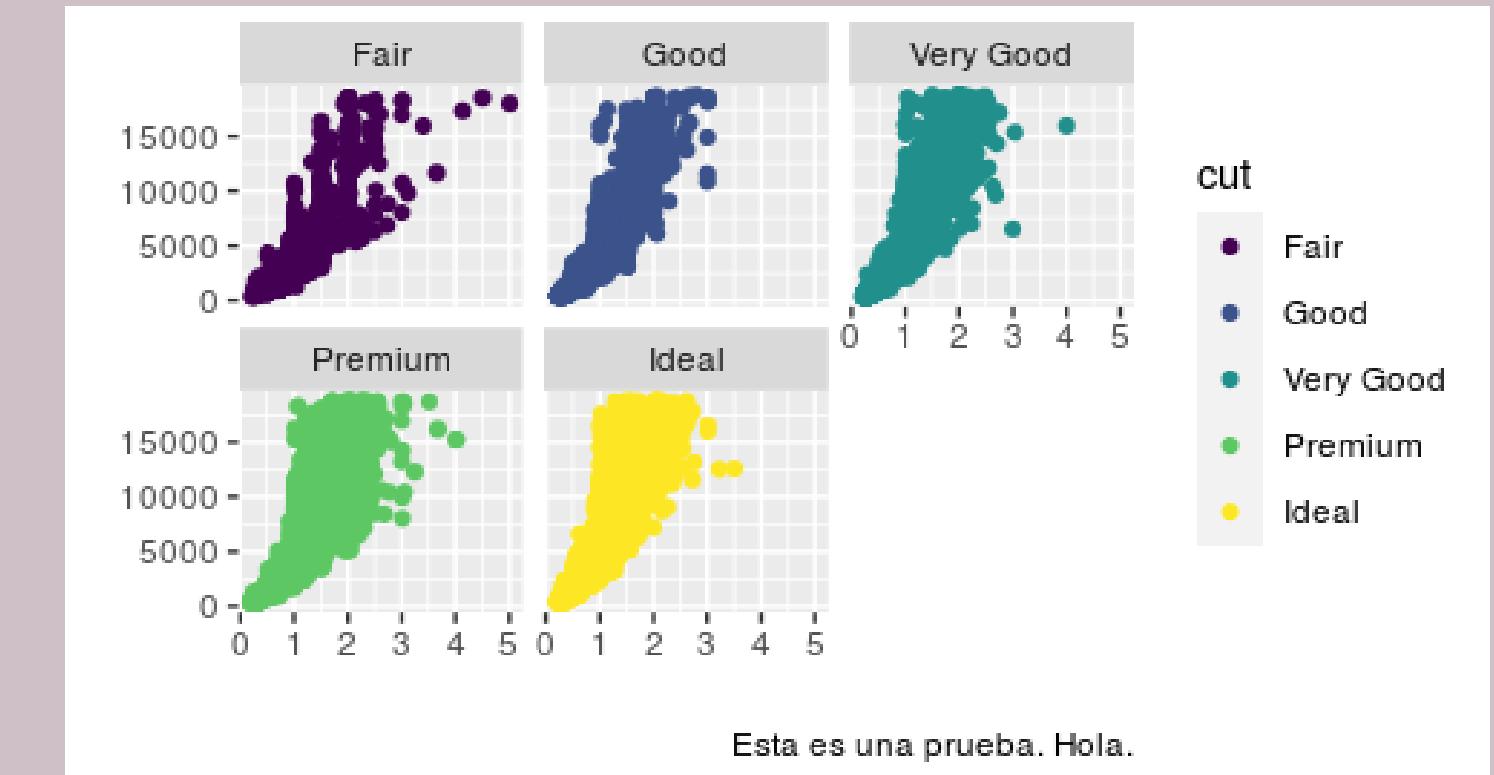
```
messy %>%  
  gather(drug, heartrate, a:b)  
#>   name drug heartrate  
#> 1 Wilbur a 67  
#> 2 Petunia a 80  
#> 3 Gregory a 64  
#> 4 Wilbur b 56  
#> 5 Petunia b 90  
#> 6 Gregory b 50
```

ggplot2

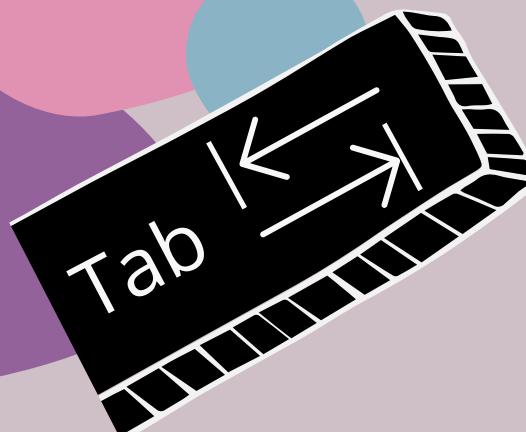


Construye distintos
gráficos
a partir de un conjunto de datos

```
ggplot(diamonds, mapping = aes(x=carat, y=price, color=cut)) + # es la base, no aparecerá nada
  geom_point() + #IMPORTANTE EL SIGNO "+" ANTES DE EMPEZAR OTRA CAPA
  facet_wrap(~cut)+ #Para que aparezcan separadas
  labs(title = "", subtitle = "",
       y="", x="", caption = "Esta es una prueba. Hola.")
```



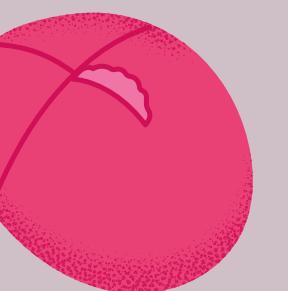
install.packages()



??function

help()

*Hagamos
un poco de
a R te*



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

Ctrl + Enter