

The background is a solid red color. It features several decorative elements: a large, faint white circle on the left side; a smaller white circle in the upper center; and a cluster of white dots in the lower right corner. There are also a few small white dots scattered elsewhere.

**EANT**

# **Introducción a Data Analytics con R**

# Quienes somos?



**Karina Bartolome**  
 @karbartolome

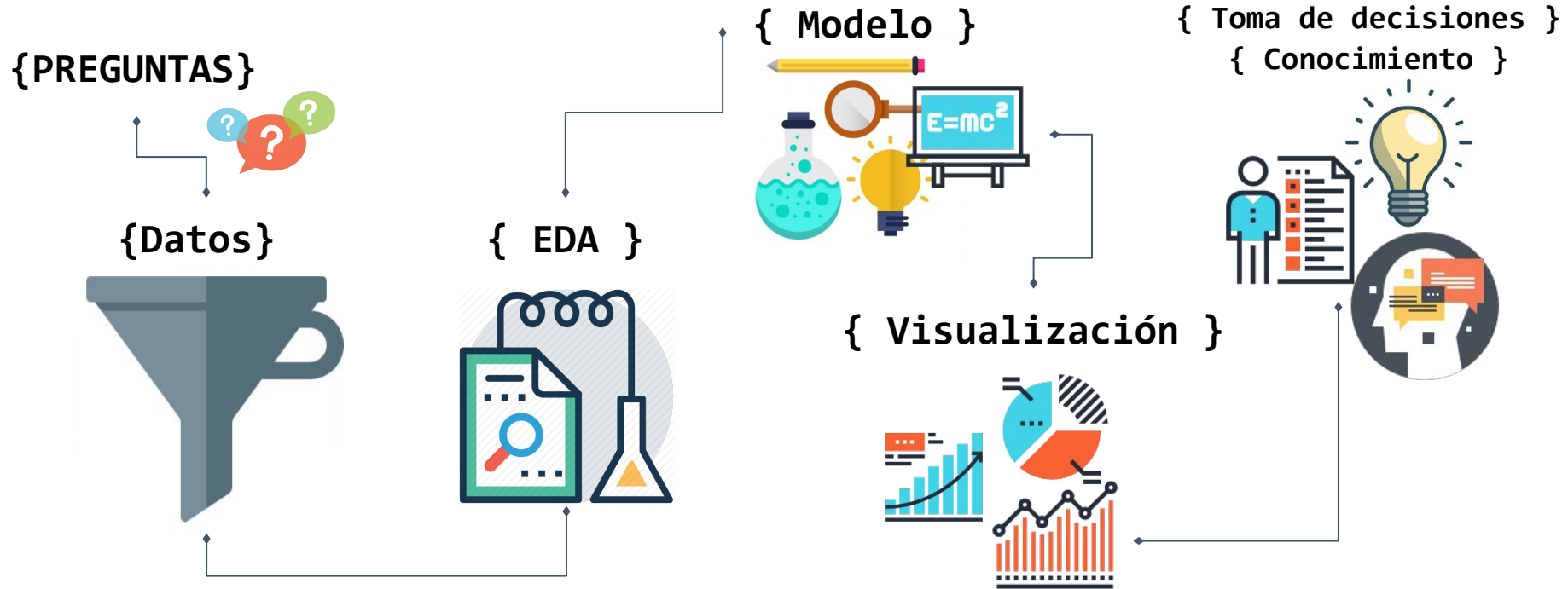


**Juan Gago**  
 @Juan\_MGago

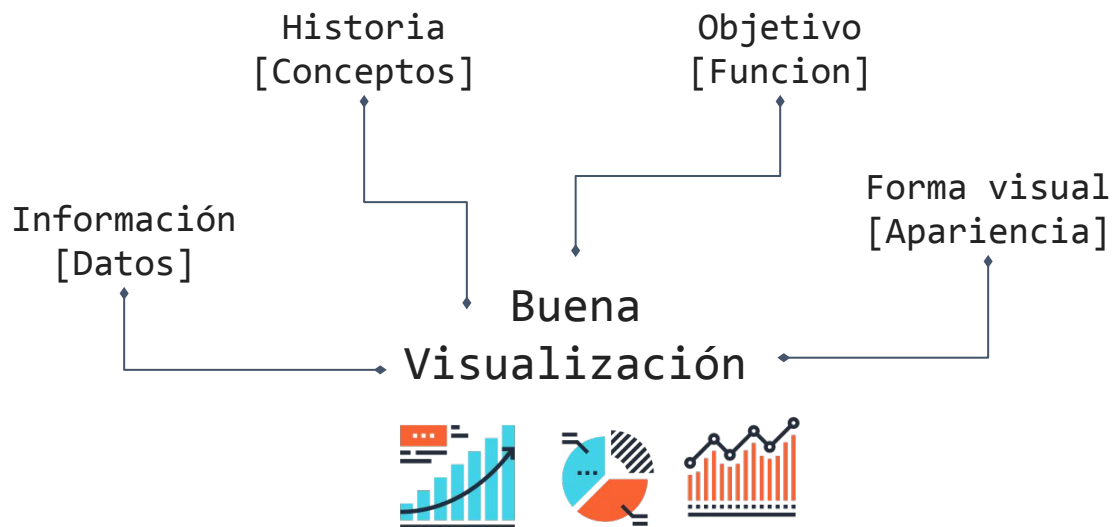


**Romina Mendez**  
 @r0mymendez

# PROCESO DE ANÁLISIS



# Visualización



# Por qué es importante?

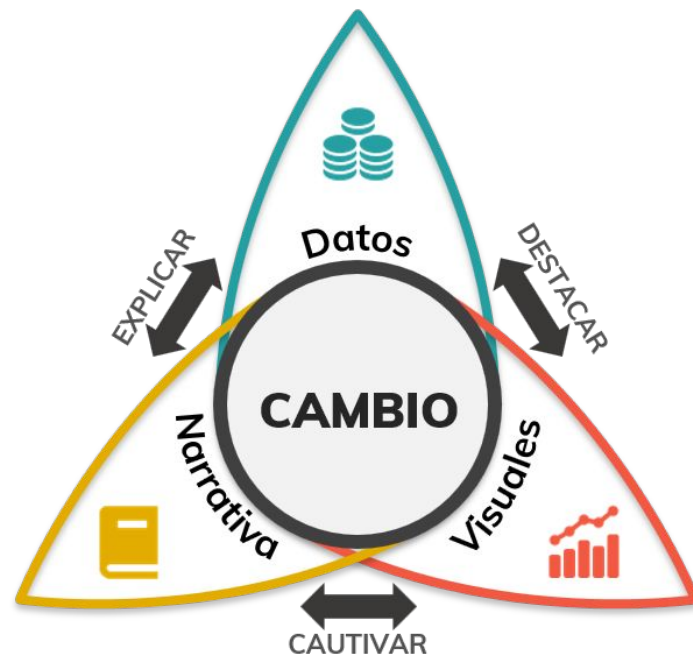
La visualización genera la posibilidad de comprender una gran cantidad de información muy rápidamente. Por lo cual la visualización...

- Permite identificar patrones y comunicar relaciones y significado.
- Puede inspirar a generar nuevas preguntas y una mayor exploración.
- Ayuda a identificar subproblemas.
- Es realmente buena para identificar tendencias y valores atípicos, descubrir o buscando puntos de datos interesantes o específicos en un campo más amplio.



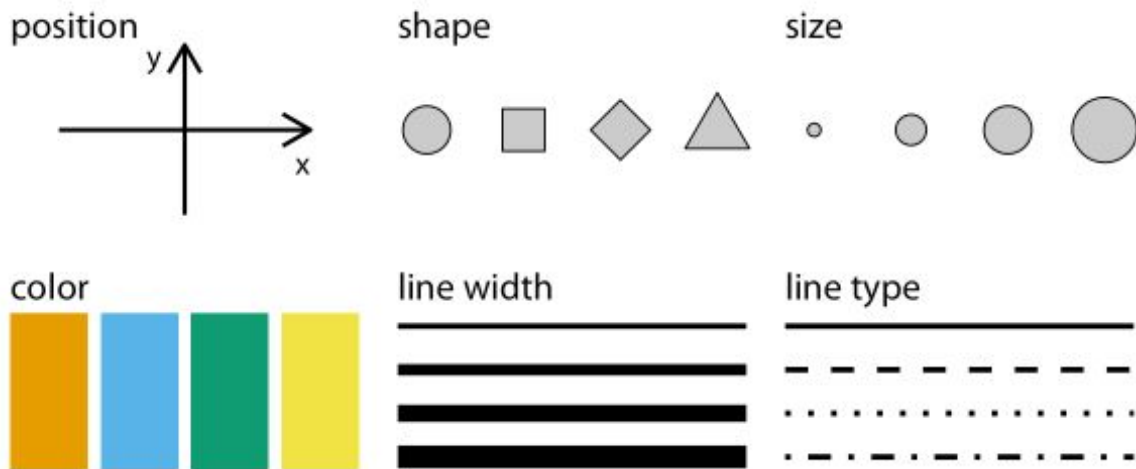
# Contar una historia

La mayor parte de la visualización de datos se realiza con fines de comunicación. Tenemos una idea sobre un conjunto de datos, y tenemos una audiencia potencial, y nos gustaría transmitir nuestra información a nuestra audiencia.



# Estéticas

Todos los elementos gráficos tienen una *forma* , un *tamaño*, posición y un *color*.



# QUE ES R?

R es un entorno y lenguaje de programación con un enfoque al análisis estadístico.

Se trata de uno de los lenguajes de programación más utilizados en investigación científica.





# HISTORIA DE R

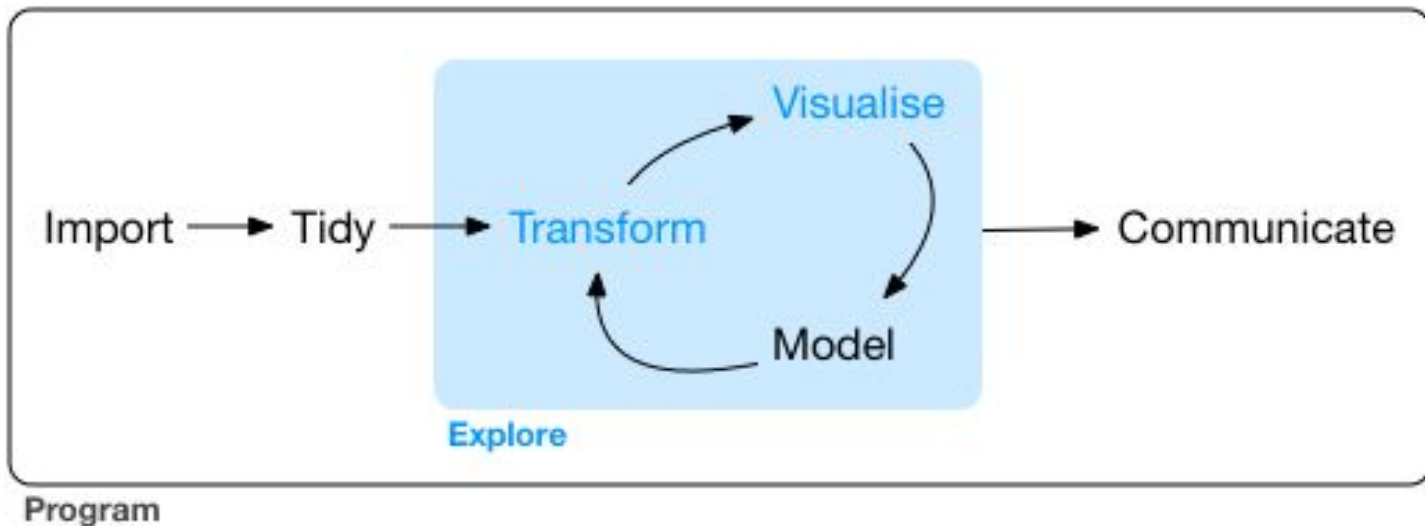
Nació como una reimplementación de software libre del lenguaje S

Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993.

Sus bases iniciales se remontan en los Bell Laboratories de AT&T y ahora Alcatel-Lucent en Nueva Jersey con el lenguaje S.



# QUE PUEDO HACER CON R?



# R STUDIO

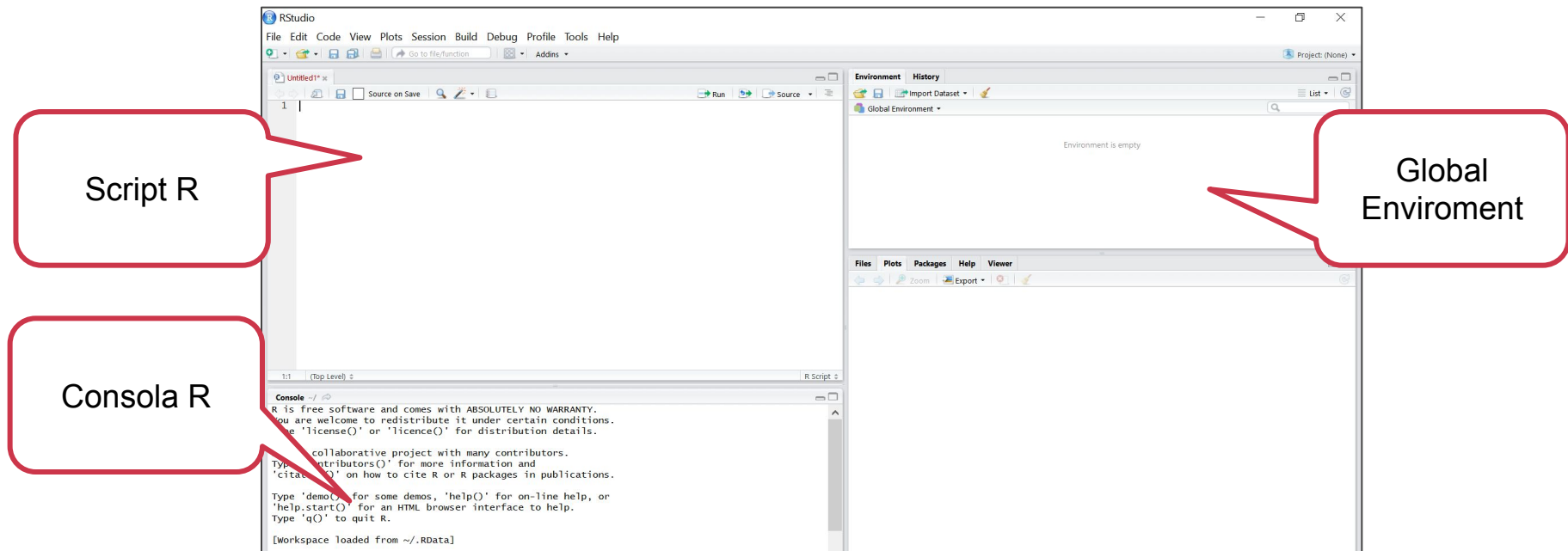
IDE: Su siglas hacen referencia a entorno de desarrollo integrado.

Es un marco de desarrollo amigable, el cual permite:

1. Autocompletado
2. Reconocimiento de sintaxis de programación
3. Depurador de errores
4. Manual de usuarios y ayuda en línea.



# R STUDIO



# Comentarios en R

Para escribir comentarios en R solo debes colocar el símbolo # al comenzar una línea.

Pero también puedes combinar tu código y tus comentarios con una R Notebook.



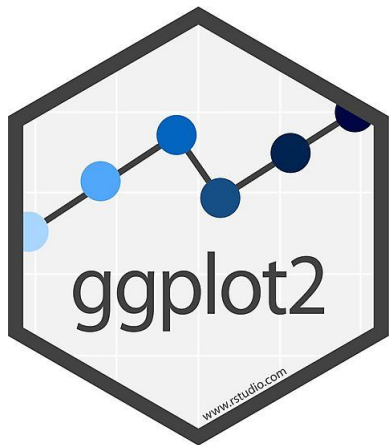
# CÓMO REALIZAR ASIGNACIÓN?

Si queremos crear un objeto debemos utilizar el operador de asignación `<-`.

La sintaxis para crear un objeto es:

```
ObjetoNombre <- valor
```

# { GG PLOT 2 }



```
library(ggplot2)
```

El paquete ggplot2 utiliza un enfoque altamente modular para los gráficos, que le permite construir y personalizar las gráficas más fácilmente.

El paquete fue creado por Hadley Wickham, quien también escribió ggplot2: Elegant Graphics for Data Analysis (Springer, 2009). Dicho libro explica el paradigma detrás de ggplot2 y cómo usar las funciones del paquete.



# { GG PLOT 2: CAPAS }

Las capas de una gráfica ggplot pueden ser las siguientes:







## { GGLOT2: CAPAS }

Cada capa tiene una funcionalidad diferentes que se van adicionando para poder realizar la gráfica:

- **DATA:** Es la primer capa y se define el set de datos que se va utilizar
- **AESTHETICS:** En la siguiente capa nos permite especificar las características, las columnas (es decir, la dimensión) que queremos trazar.

*Observación: estas dos capas no grafican nada, solo realizan la selección de los datos y los ejes.*

- **GEOMETRICS:** En esta capa definimos las formas que pretendemos usar para presentar los datos usando ggplot. Después de agregar esta capa, el ggplot sabe cómo mostrar los datos.



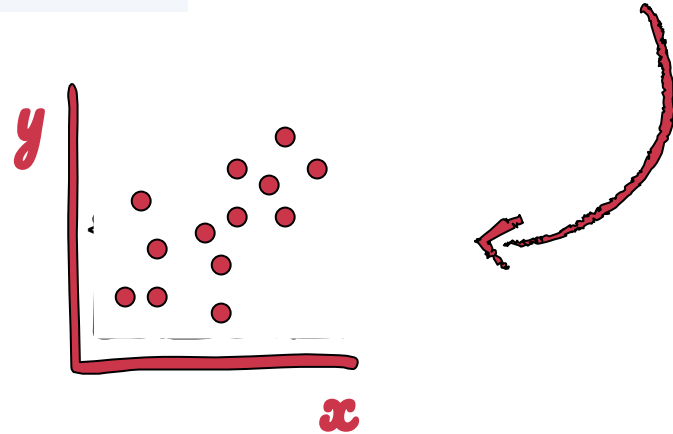
## { GGPLOT2: CAPAS }

Cada capa tiene una funcionalidad diferentes que se van adicionando para poder realizar la gráfica:

- **FACETS:** En esta capa se permite poder realizar varios gráficos generando una agrupación en base a una variable categórica.
- **STATISTICS:** En esta capa se pueden agregar alguna medida estadísticas a un gráfico.
- **COORDINATES:** A menudo se usa para aplicar el límite en el eje x o el eje y para jugar con la relación x vs y, por lo tanto, personalizar la imagen según sea necesario.
- **THEME:** En esta capa se permite aplicar diferentes estilos, los mismos se pueden personalizar pero ggplot2 tiene algunos definidos en funciones.



```
{ Ahora vamos a crear nuestro  
primer gráfico con  
ggplot2 }
```



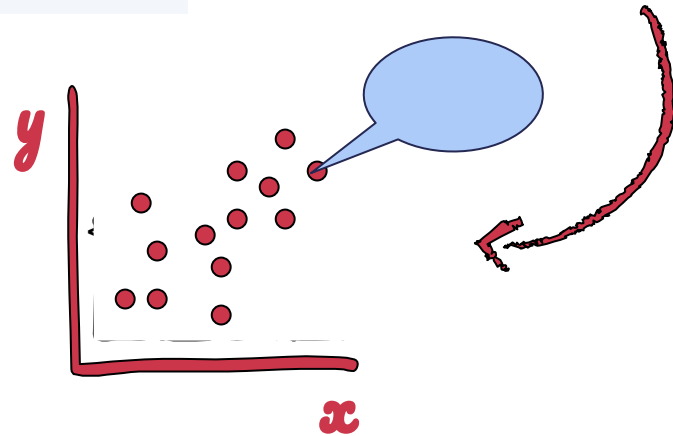
**Plotly** proporciona herramientas de gráficos, análisis y estadísticas, así como paquetes de código abierto con implementación de gráficos interactivos para Python , R , MATLAB , Perl , Julia, Arduino y REST .

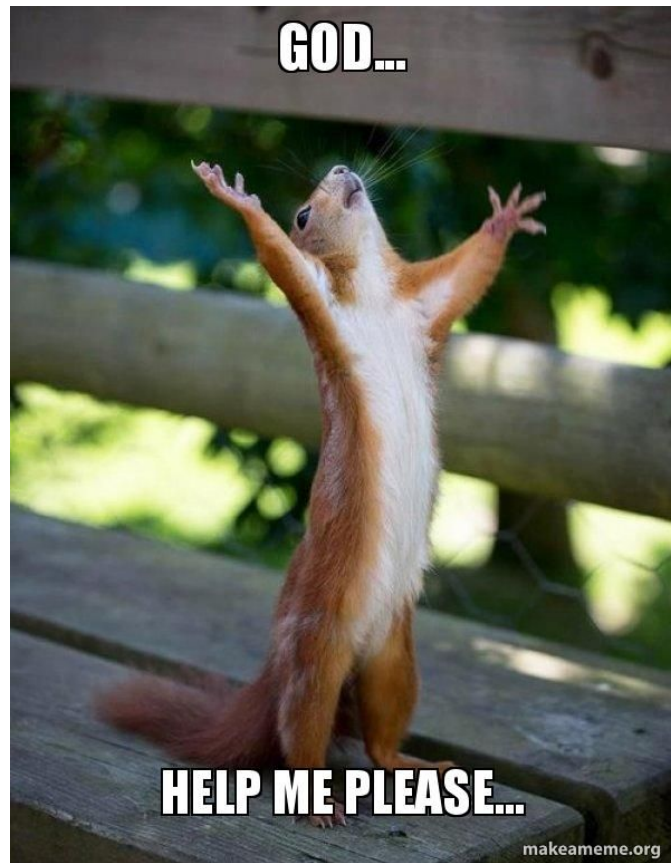
Para conocer más de la implementación de Ploty en R: <https://plot.ly/r/>



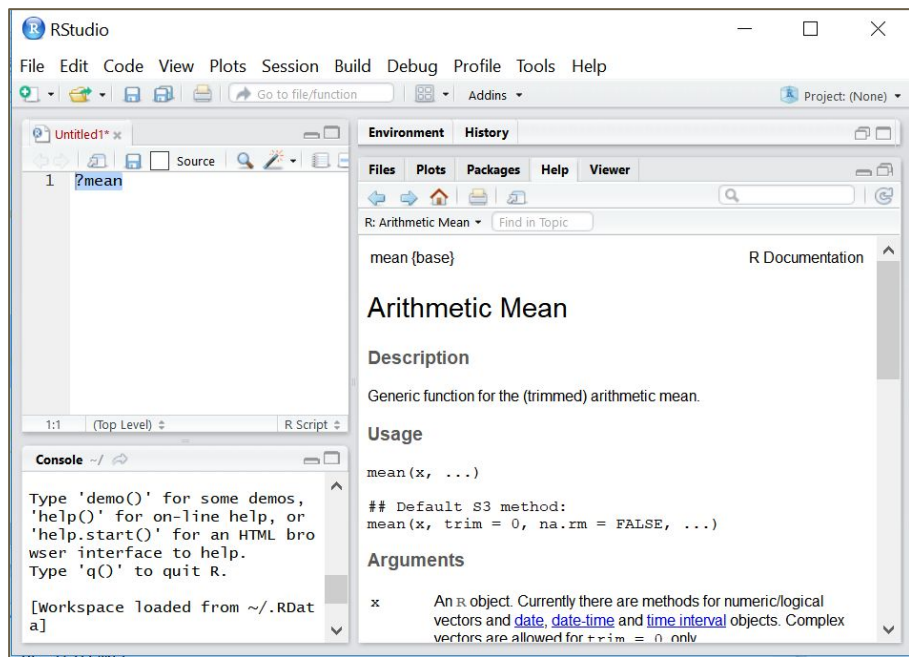


```
{ Ahora vamos a crear nuestro  
primer gráfico con  
plotly }
```





# R STUDIO: Help!



La ayuda de R es muy buena para encontrar información sobre una función sobre sus argumentos y hasta ejemplos!

Solo se debe colocar **?xxx**, siendo “xxx” la función que quiero encontrar información.

Posteriormente se activará la ventana de Help en el margen inferior derecho.

Por ejemplo: **?mean**

# HELP! Cheat Sheet

## Base R Cheat Sheet

### Getting Help

#### Accessing the help files

##### ?mean

Get help of a particular function.

**help.search('weighted mean')**

Search the help files for a word or phrase.

**help(package = 'dplyr')**

Find help for a package.

#### More about an object

##### str(iris)

Get a summary of an object's structure.

##### class(iris)

Find the class an object belongs to.

### Using Libraries

#### install.packages('dplyr')

Download and install a package from CRAN.

#### library(dplyr)

Load the package into the session, making all its functions available to use.

#### dplyr::select

Use a particular function from a package.

#### data(iris)

Load a built-in dataset into the environment.

### Vectors

#### Creating Vectors

c(2, 4, 6)	2 4 6	Join elements into a vector
2:6	2 3 4 5 6	An integer sequence
seq(2, 3, by=0.5)	2.0 2.5 3.0	A complex sequence
rep(1:2, times=3)	1 2 1 2 1 2	Repeat a vector
rep(1:2, each=3)	1 1 1 2 2 2	Repeat elements of a vector

#### Vector Functions

<b>sort(x)</b>	<b>rev(x)</b>
Return x sorted.	Return x reversed.
<b>table(x)</b>	<b>unique(x)</b>
See counts of values.	See unique values.

#### Selecting Vector Elements

##### By Position

<b>x[4]</b>	The fourth element.
<b>x[-4]</b>	All but the fourth.
<b>x[2:4]</b>	Elements two to four.
<b>x[-(2:4)]</b>	All elements except two to four.
<b>x[c(1, 5)]</b>	Elements one and five.

### Programming

#### For Loop

```
for (variable in sequence){  
  Do something  
}
```

##### Example

```
for (i in 1:4){  
  j <- i + 10  
  print(j)  
}
```

#### While Loop

```
while (condition){  
  Do something  
}
```

##### Example

```
while (i < 5){  
  print(i)  
  i <- i + 1  
}
```

#### If Statements

```
if (condition){  
  Do something  
} else {  
  Do something different  
}
```

##### Example

```
if (i > 3){  
  print('Yes')  
} else {  
  print('No')  
}
```

#### Functions

```
function_name <- function(var){  
  Do something  
  return(new_variable)  
}
```

##### Example

```
square <- function(x){  
  squared <- x*x  
  return(squared)  
}
```

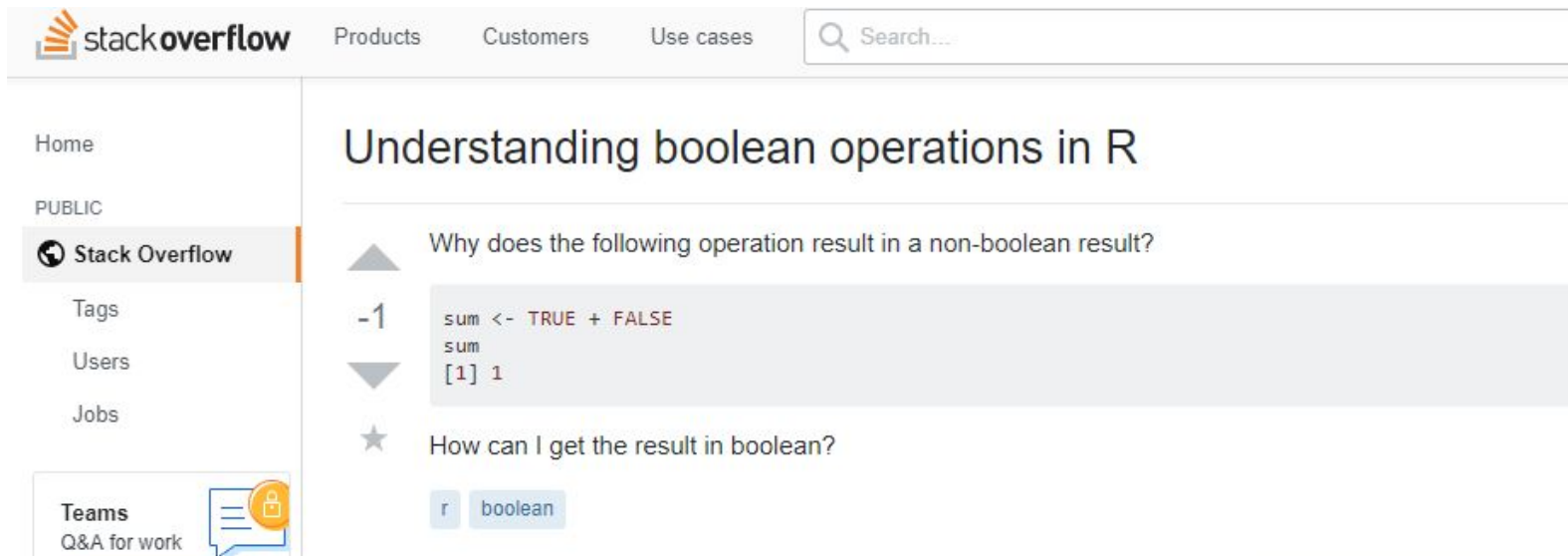
### Reading and Writing Data

Input	Output	Description
df <- read.table('file.txt')	write.table(df, 'file.txt')	Read and write a delimited text file.



[illegible]

# HELP! StackOverflow



The screenshot shows the Stack Overflow website interface. The top navigation bar includes the Stack Overflow logo, links for Products, Customers, and Use cases, and a search bar. The left sidebar contains links for Home, PUBLIC, Stack Overflow (selected), Tags, Users, Jobs, and Teams. The main content area displays a question titled "Understanding boolean operations in R". The question text is "Why does the following operation result in a non-boolean result?". Below the text is a code block showing R code: 

```
sum <- TRUE + FALSE
sum
[1] 1
```

. The code block is preceded by a "-1" icon. Below the code block is a star icon and the text "How can I get the result in boolean?". At the bottom of the question, there are two tags: "r" and "boolean".

stackoverflow Products Customers Use cases Search...

Home

PUBLIC

Stack Overflow

Tags

Users

Jobs

Teams  
Q&A for work

## Understanding boolean operations in R

Why does the following operation result in a non-boolean result?

```
sum <- TRUE + FALSE
sum
[1] 1
```

How can I get the result in boolean?

r boolean



{ ¿PREGUNTAS? }





{ GRACIAS }



# EANT

ESCUELA ARGENTINA  
DE NUEVAS TECNOLOGIAS

Estate atento a más eventos!



EANT



eanttech



@eanttech