

R como minero de datos

Sesión 1

Natalie Julian - www.nataliejulian.com

Estadística UC y Data Scientist en Zippedi Inc.

The New Era of Information Abundance: What Does It Mean for Higher Education?

An automated health care system that understands when to step in

Digital learning statistics reveal how Highland schools have adapted to coronavirus crisis

How India is using Artificial Intelligence to combat COVID-19

Big Data Apps Wasting Billions in the Cloud

Digital transformation has become 'unstoppable' and essential for retailers, study finds

El flujo explosivo de información ha conllevado un replanteamiento del panorama tecnológico.

¿Cómo darnos cuenta de esta *explosión de la información* en nuestro día a día? Solo pensemos:

- ¿Cuántas plataformas tecnológicas uso a diario?
- ¿Cuántas *stories* de Instagram publican mis amig@s por semana?
- ¿Cuánta información tuya aportas a internet?

El flujo explosivo de información ha conllevado un replanteamiento del panorama tecnológico.

¿Cómo darnos cuenta de esta *explosión de la información* en nuestro día a día? Solo pensemos:

- ¿Cuántas plataformas tecnológicas uso a diario?
Instagram, Facebook, Tinder, Google, LinkedIn, Zoom, Tik Tok, Netflix, Amazon Prime Video, Photoshop, Spotify, Youtube, Gmail...
- ¿Cuántas *stories* de Instagram publican mis amig@s por semana?
No sé, ¿500?
- ¿Cuánta información tuya aportas a internet?
No sé y prefiero no saber!

Tanto dato necesita **minarse**

¿Se imaginan cuántos datos se generan cada segundo? ¡Muchísimos!. Pero no sirve de nada si sólo está ahí en el ciberespacio: Además de almacenarla, se necesitan realizar inferencias y así descubrir *lo verdadero, entender el comportamientos de los datos*.

La Ciencia: En búsqueda de la verdad

La ciencia, está hecha de errores, pero de errores útiles de cometer, pues poco a poco, conducen a la verdad.

Julio Verne

¿Cómo determinar científicamente *lo verdadero*?

- Realizar una transcripción de los fenómenos y sucesos del mundo de manera compacta: Es necesario cuantificar la información y almacenarla en **Datos**.
- Aplicar la lógica en los datos: Desarrollar análisis, usar metodología y procedimientos para obtener **Inferencias** basadas en **Evidencia** presente en los datos.

¿Cómo recopilar evidencia? Usando metodología, análisis, usando **estadística**.

Lenguaje de programación **R**

Aprenderemos a utilizar una de las herramientas más fuertes e imprescindibles a la hora de analizar datos: **manejo de software**.

El software que aprenderemos será **RStudio**, una plataforma amigable para aprender el lenguaje de programación **R**. Este lenguaje es eficaz para la manipulación de bases de datos y desarrollo de análisis estadísticos.

Vista general

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Console Terminal x Jobs x

~/

R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |

Environment History Connections

Import Dataset

Global Environment

Environment is empty

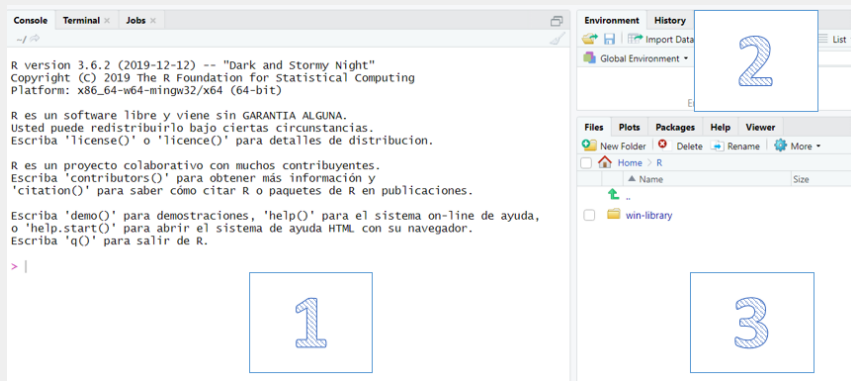
Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > R

Name Size Modified

..
win-library



- 1 Aquí se encuentra la *consola*, aquí se entregan resultados y también puede utilizarse como una calculadora. ¡Pero **nada** queda guardado en la consola!
- 2 Aquí se mostrarán todos los objetos que definamos, los objetos del *workspace*.
- 3 Aquí veremos los gráficos que realicemos, entre otras cosas que comentaremos más adelante.

R como calculadora

Escribamos de una en una las siguientes operaciones en la consola:

■ $2 + 2$

■ $3 - \frac{1}{5} \cdot 100$

■ $2 \cdot 2 + \frac{3}{6} - \frac{6}{8+1}$

Para que se muestre el resultado, debemos apretar Enter.

R como calculadora

```
Console Terminal x Jobs x
~/
R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> 2+2
[1] 4
> 3-1/5*100
[1] -17
> 2*2+3/6-6/(8+1)
[1] 3.833333
> |
```

Muchas operaciones

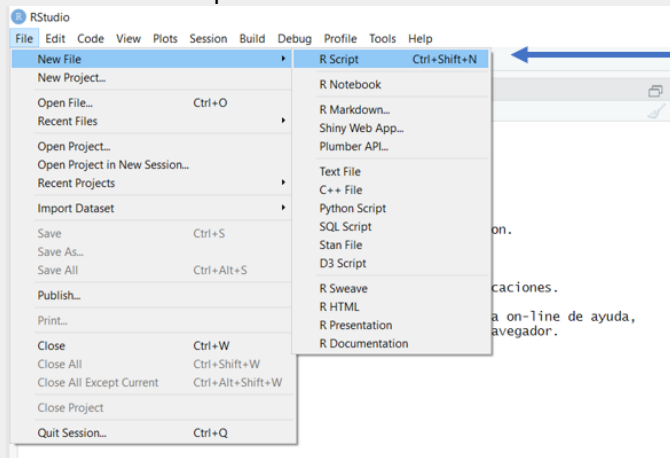
Algunas operaciones matemáticas que podemos realizar:

Función	Ejemplo 1	Ejemplo 2
Suma	$6+1$	<code>sum(2, 5)</code>
Multiplicación	$90*11$	<code>prod(4, 3)</code>
Valor absoluto	<code>abs(-2)</code>	<code>abs(-0.005)</code>
Potencia	$4**8$	8^10
Raíz cuadrada	<code>sqrt(4)</code>	<code>sqrt(0.05)</code>
División	$8/9$	$5/3$
Exponencial	<code>exp(-1/2)</code>	<code>exp(1)</code>
Logaritmo	<code>log(5, exp(2))</code>	<code>log(5, 3)</code>
Seno	<code>sin(10)</code>	<code>sin(pi)</code>

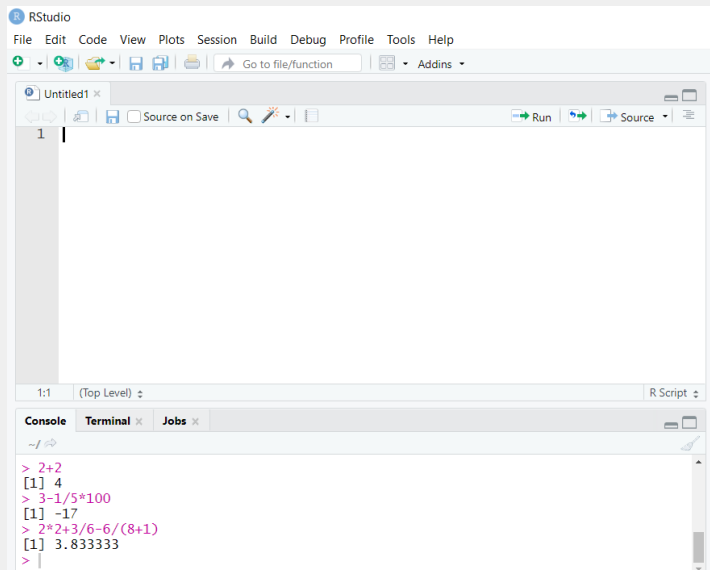
¿Cómo guardar mis códigos?

Debemos abrir un *script*, un script corresponde al código que iremos escribiendo para obtener los resultados.

Para abrir un script:



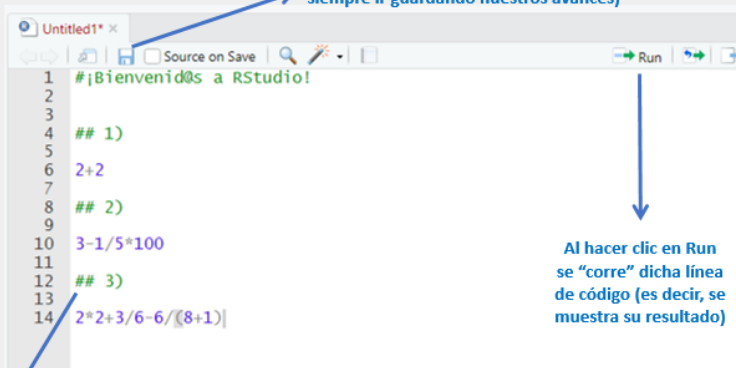
Vista del script



Escribiendo en el script

Podemos escribir comentarios anteponiendo el símbolo # y añadir las operaciones que antes realizamos:

Para guardar un archivo con nuestro código hacemos clic aquí (es importante siempre ir guardando nuestros avances)



```
1 #¡Bienvenid@s a RStudio!  
2  
3  
4 ## 1)  
5  
6 2+2  
7  
8 ## 2)  
9  
10 3-1/5*100  
11  
12 ## 3)  
13  
14 2*2+3/6-6/(8+1)
```

Al hacer clic en Run se “corre” dicha línea de código (es decir, se muestra su resultado)

Para añadir un comentario se añade el símbolo #

Definiendo objetos en RStudio

Supongamos que queremos conocer el resultado de la siguiente operación:

$$\frac{10000000000000 \cdot 20}{\exp(20)} + \frac{\log(20)}{10000000000000} + \text{seno}\left(\frac{10000000000000}{20}\right)$$

Escribir esa expresión en R resulta un poco larga... ¿Qué hacer?

Podemos crear un objeto llamado *a* que sea equivalente al valor 10000000000000 y un objeto *b* equivalente a 20.

Siempre que queramos definir un objeto debemos utilizar la flecha de asignación "<-" con la siguiente estructura:

```
nombreobjeto<-valor del objeto
```


Calculando expresiones con objetos

Creando los objetos a y b podemos calcular el valor de la expresión en R de forma sencilla:

```
7  
8 ## 2)  
9  
10 3-1/5*100  
11  
12 ## 3)  
13  
14 2*2+3/6-6/(8+1)  
15  
16 ## Creamos el objeto a:  
17  
18 a<-1000000000000  
19  
20 ## Creamos el objeto b:  
21  
22 b<-20  
23  
24 (a*b)/exp(b)+log(b)/a+sin(a/b)  
25
```

Al objeto a se le asigna el valor 1000000000000 y al objeto b se le asigna el valor 20

Al realizar la asignación, R comprende como a y b a los números 1000000000000 y 20 respectivamente

Al hacer clic en Run, mostramos el resultado de la operación

Console

```
> a<-1000000000000  
> b<-20  
> (a*b)/exp(b)+log(b)/a+sin(a/b)  
[1] 41222.51  
> |
```