

Unidad I. Introducción a R y RStudio

Del Curso introductorio al lenguaje de programación R orientado al análisis cuantitativo en Ciencias Sociales por Sarahí Aguilar González

Objetivo: Que el estudiante se familiarice con R y el entorno de desarrollo de RStudio, y sea capaz de crear nuevos scripts y Proyectos de R.

Agenda



Definiendo R

La importancia de R

Scripts de R

Entorno de desarrollo de RStudio

Directorios de trabajo

Proyectos de R



Definiendo R

La importancia de R

Scripts de R

Entorno de desarrollo de RStudio

Directorios de trabajo

Proyectos de R

R es un lenguaje de programación y entorno enfocado al cómputo estadístico y graficación.



R es un **lenguaje de programación** y **entorno** enfocado al **cómputo estadístico** y **graficación**.



R es un **lenguaje de programación** y entorno enfocado al cómputo estadístico y graficación.

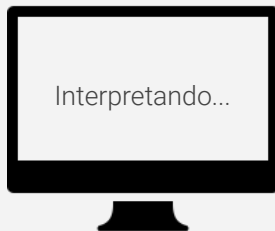
Un lenguaje de programación es un **lenguaje formal** que comprende un **conjunto de cadenas de caracteres** que producen varios tipos de salida de **código de máquina**.

Código de R

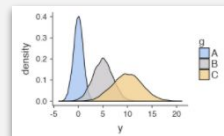
Texto escrito empleando las reglas del lenguaje de programación R.

```
N <- 10000
df <- data.frame(x=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
                y=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
                g=rep(c('A','B','C'), each=N))
ggplot(df, aes(y, fill=g)) +
  geom_density(alpha=0.8) +
  jmvcore::theme_default()
```

Acción de ejecución



Resultado



Un lenguaje de programación no es un programa.

R es un lenguaje de programación y **entorno** enfocado al cómputo estadístico y graficación.

El entorno de R es un **conjunto integrado de instalaciones de software** para la **procesamiento**, el **análisis** y la **visualización de datos**.

El entorno de R incluye:



Una instalación eficaz de manejo y almacenamiento de datos



Un conjunto de operadores para cálculos en matrices



Una colección amplia, coherente e integrada de herramientas para el análisis de datos



Facilidades gráficas para el análisis de datos y visualización en pantalla o en papel

R es un lenguaje de programación y entorno enfocado al **cómputo estadístico** y graficación.

R fue creado inicialmente con fines de investigación estadística
y **hoy proporciona una amplia variedad de técnicas estadísticas mediante implementaciones accesibles y simples.**

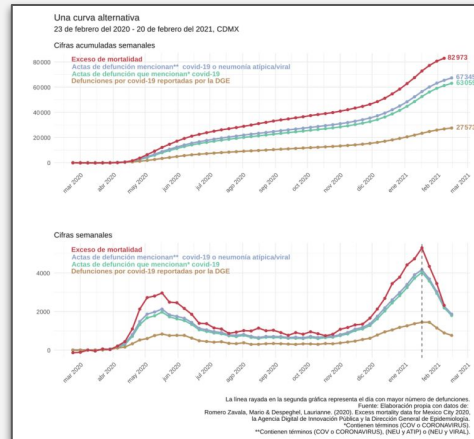
Por ejemplo, modelado lineal y no lineal, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, clustering, entre otras.



R es un lenguaje de programación y entorno enfocado al cómputo estadístico y **graficación**.

Facilidad con la que se pueden producir gráficos de alta calidad; una de las fortalezas de R.

En R, es fácil obtener un *buen* diseño en las visualizaciones de datos pues el usuario conserva el control total de todos los elementos gráficos.



Ejemplo de gráfico producido 100% en R.

Agenda



Definiendo R

La importancia de R

Scripts de R

Entorno de desarrollo de RStudio

Directorios de trabajo

Proyectos de R

La importancia de R



R es **gratis**.



La **documentación** de R es accesible para todas las áreas de estudio.



La **comunidad** de programadores de R (también conocidos como “useRs”) es diversa e incluyente y abierta a ayudar y colaborar.



R es constantemente **actualizado**.



En julio 2021, R ocupa el puesto #12 en el índice TIOBE, una medida de **popularidad** de los lenguajes de programación.



R está disponible como **Software Libre** bajo los términos de la Licencia Pública General GNU de The Free Software Foundation.



R es **escalable** a través de sus paquetes.

Agenda



Definiendo R

La importancia de R

Scripts de R

Entorno de desarrollo de RStudio

Directorios de trabajo

Proyectos de R

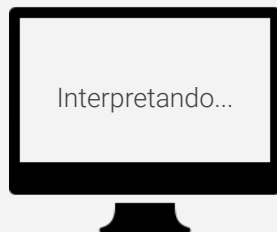
¿Qué significa *programar*?

Código

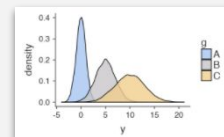
Texto escrito empleando las reglas de un lenguaje de programación.

```
N <- 10000
df <- data.frame(x=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
                 y=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
                 g=rep(c('A','B','C'), each=N))
ggplot(df, aes(y, fill=g)) +
  geom_density(alpha=0.8) +
  jvcore::theme_default()
```

Acción de ejecución

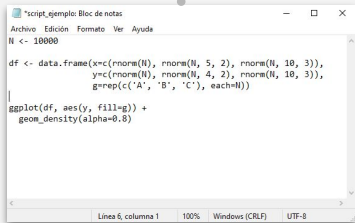


Resultado



```
N <- 10000
df <- data.frame(
  g=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
  y=c(rnorm(N), rnorm(N, 4, 2), rnorm(N, 10, 3)),
  g=rep(c('A', 'B', 'C'), each=N))
ggplot(df, aes(y, fill=g)) +
  geom_density(alpha=0.8)
```

Líneas de código.



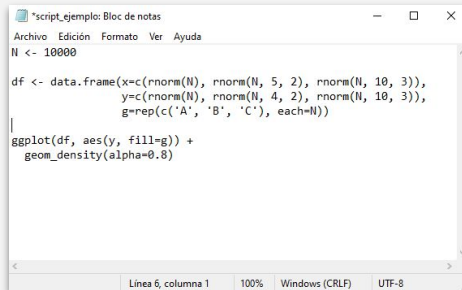
Script
Archivo de texto que contiene líneas de código.

Un script de R es un **archivo de texto que contiene líneas de código escritas en el lenguaje de programación de R.**

Tiene la terminación **.R**

Las líneas de código pueden ser **comandos** o **comentarios**.

El script se puede **editar** y **guardar** para ser ejecutado después.



```
*script_gjemplo: Bloc de notas
Archivo Edición Formato Ver Ayuda
N <- 10000

df <- data.frame(x=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
                 y=c(rnorm(N), rnorm(N, 4, 2), rnorm(N, 10, 3)),
                 g=rep(c('A', 'B', 'C'), each=N))

ggplot(df, aes(y, fill=g)) +
  geom_density(alpha=0.8)
```

Linea 6, columna 1 100% Windows (CRLF) UTF-8

Agenda



Definiendo R

La importancia de R

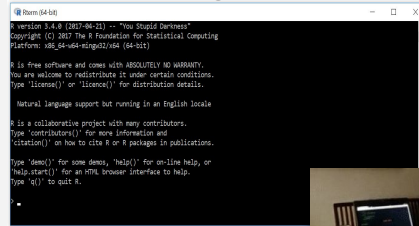
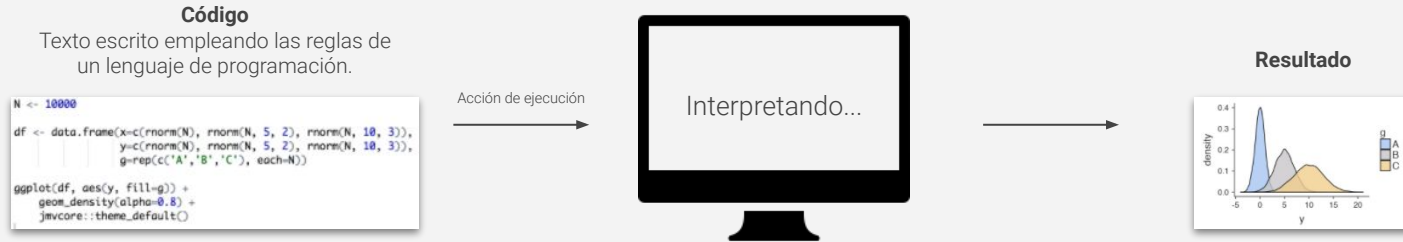
Scripts de R

Entorno de desarrollo de RStudio

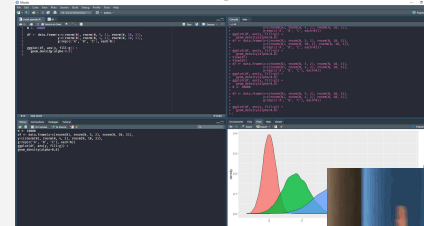
Directorios de trabajo

Proyectos de R

¿Qué significa *programar*?



Consola con intérprete integrado.



Entorno de desarrollo integrado (IDE) con intérprete integrado y otras utilidades para facilitar la programación.



RStudio es un entorno de desarrollo integrado (IDE) para R.

Incluye una consola, un editor de código con resaltado de sintaxis que admite la ejecución directa de código, herramientas para la graficación, control del historial, la depuración de código y la gestión del espacio de trabajo.

RStudio está disponible en ediciones comerciales y de código abierto y se ejecuta en el escritorio o en un navegador conectado a RStudio Server o RStudio Workbench.

Entorno de desarrollo de RStudio



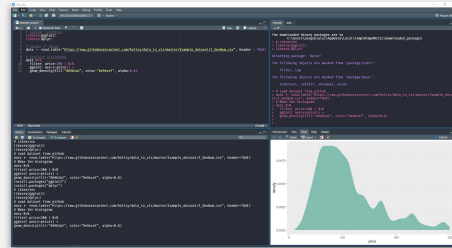
Lenguaje de programación y entorno enfocado al cómputo estadístico y graficación.

corre en



Entorno de desarrollo integrado (IDE) para R.

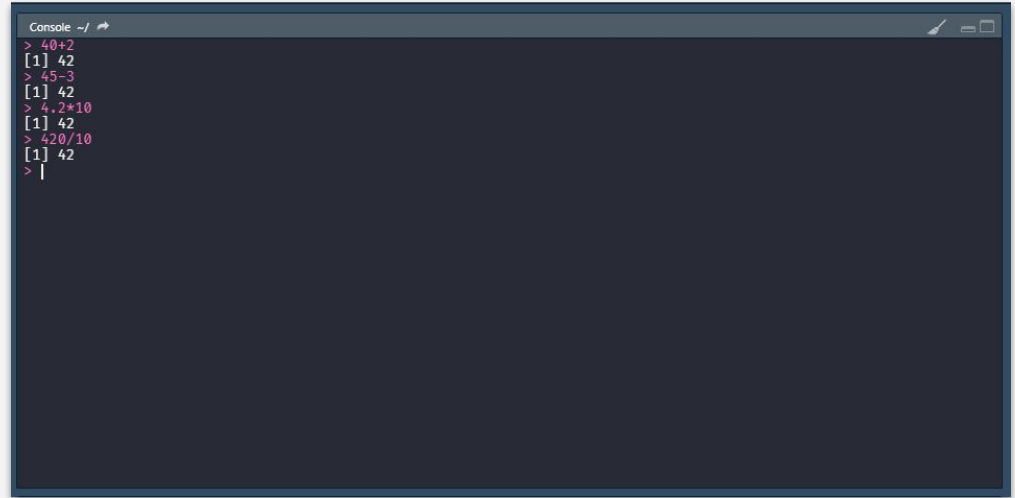
```
N <- 10000
df <- data.frame(x=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
                 y=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
                 g=rep(c("A", "B", "C"), each=N))
ggplot(df, aes(y, fill=g)) +
  geom_density(alpha=0.8) +
  jmvcore::theme_default()
```




Entorno de desarrollo de RStudio

RStudio está compuesto de múltiples ventanas.
Algunas de las más relevantes son:

- **Consola**
Permite ejecutar líneas de código de R.



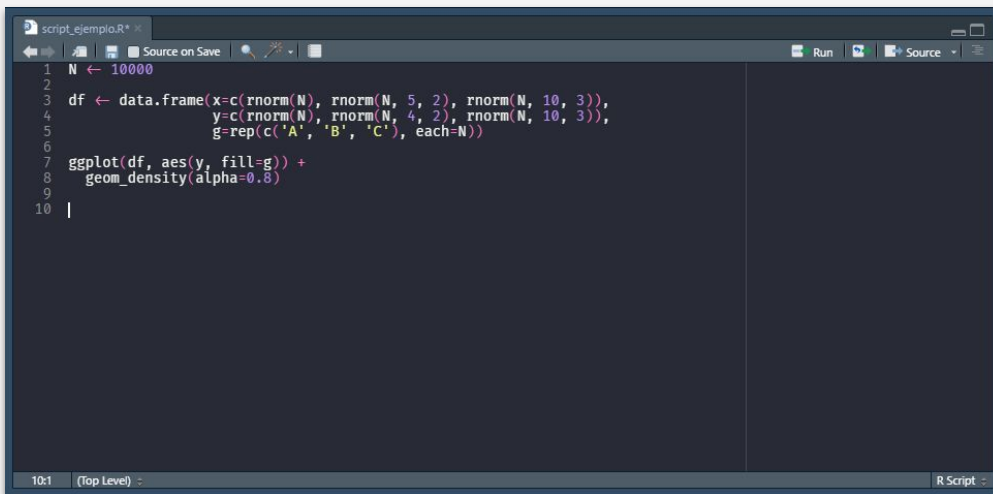
```
Console ~/   
> 4+2  
[1] 42  
> 45-3  
[1] 42  
> 4-2*10  
[1] 42  
> 420/10  
[1] 42  
> |
```

Entorno de desarrollo de RStudio

RStudio está compuesto de múltiples ventanas. Algunas de las más relevantes son:

- **Editor de código**

Permite abrir, editar, guardar y ejecutar scripts de R.



```
script_ejemplo.R*
Source on Save
Run
Source

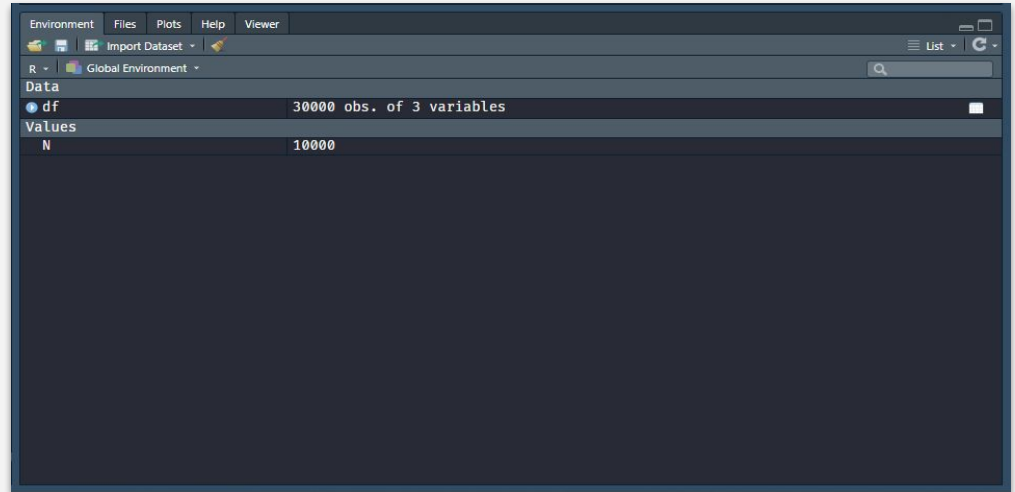
1 N <- 10000
2
3 df <- data.frame(x=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
4                 y=c(rnorm(N), rnorm(N, 4, 2), rnorm(N, 10, 3)),
5                 g=rep(c('A', 'B', 'C'), each=N))
6
7 ggplot(df, aes(y, fill=g)) +
8   geom_density(alpha=0.8)
9
10 |

10:1 (Top Level) R Script
```

Entorno de desarrollo de RStudio

RStudio está compuesto de múltiples ventanas. Algunas de las más relevantes son:

- **Ambiente**
Permite visualizar las variables almacenadas.

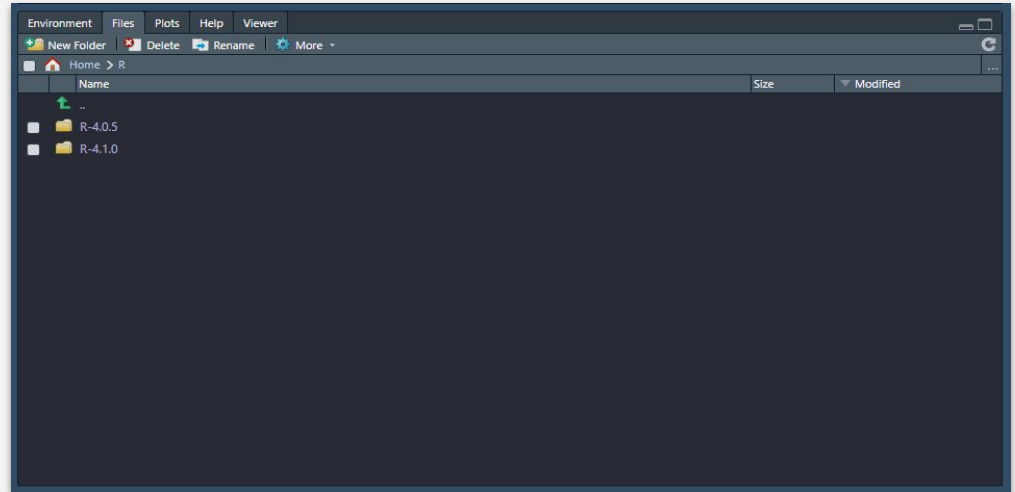


Entorno de desarrollo de RStudio

RStudio está compuesto de múltiples ventanas. Algunas de las más relevantes son:

- **Archivos**

Funciona como un explorador de archivos del equipo local de cómputo, pero permite crear carpetas, y abrir, eliminar y renombrar archivos y carpetas.

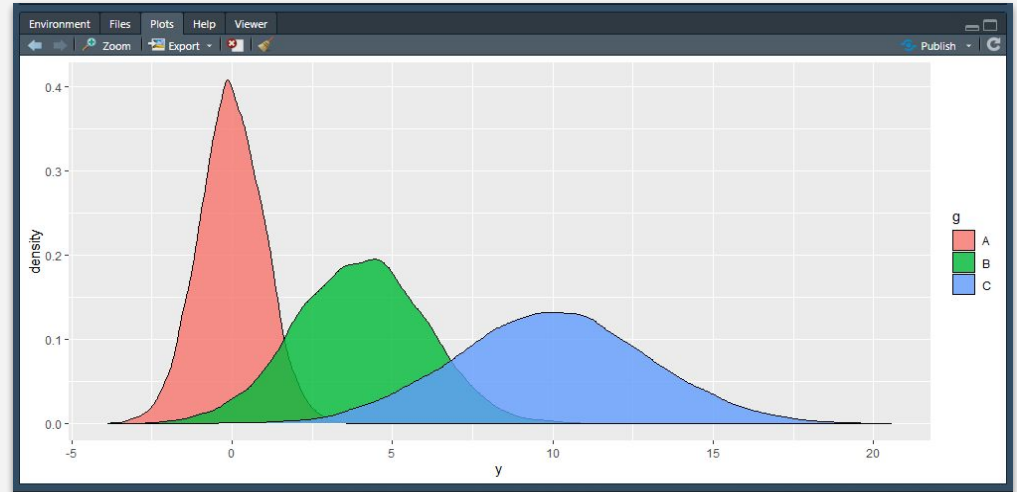


Entorno de desarrollo de RStudio

RStudio está compuesto de múltiples ventanas. Algunas de las más relevantes son:

- **Gráficos**

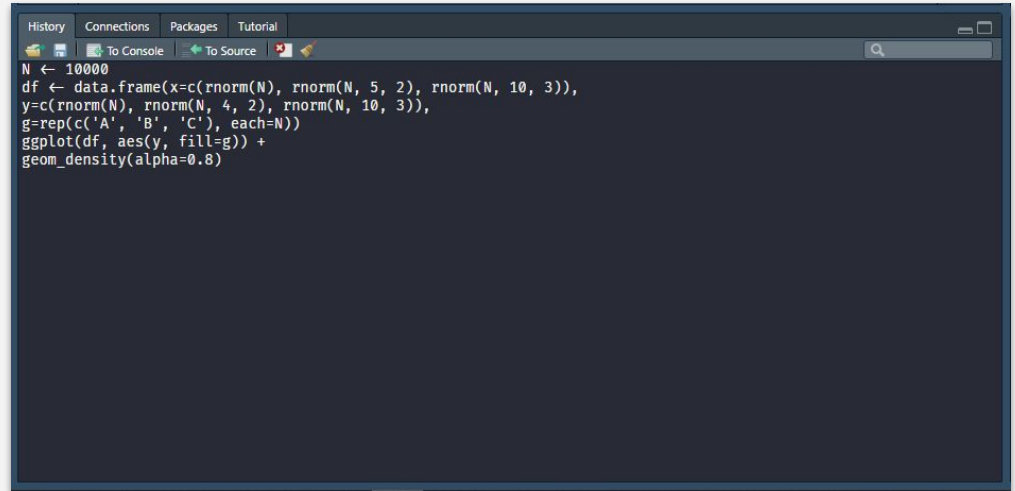
Visualizados responsivo de gráficos resultantes.



Entorno de desarrollo de RStudio

RStudio está compuesto de múltiples ventanas. Algunas de las más relevantes son:

- **Historia**
Historial de líneas de código ejecutadas.



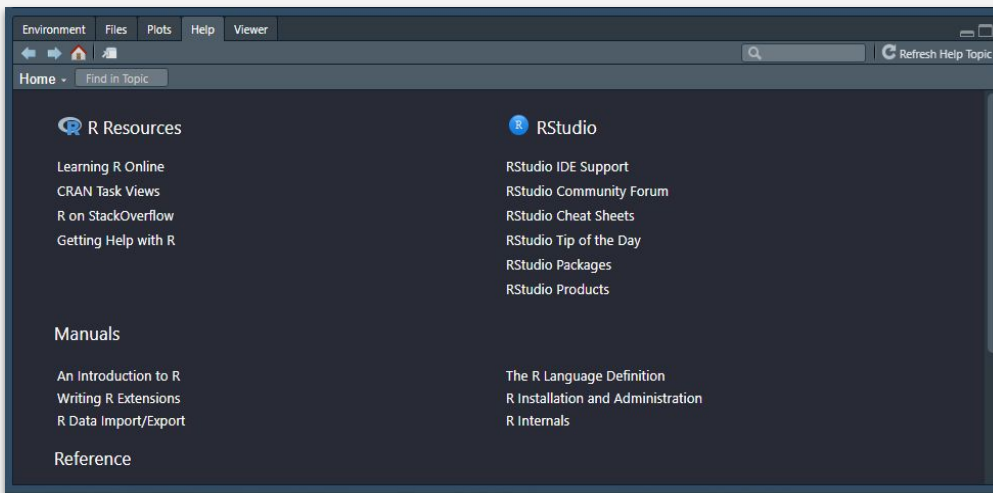
```
History Connections Packages Tutorial
N ← 10000
df ← data.frame(x=c(rnorm(N), rnorm(N, 5, 2), rnorm(N, 10, 3)),
y=c(rnorm(N), rnorm(N, 4, 2), rnorm(N, 10, 3)),
g=rep(c('A', 'B', 'C'), each=N))
ggplot(df, aes(y, fill=g)) +
geom_density(alpha=0.8)
```


Entorno de desarrollo de RStudio

RStudio está compuesto de múltiples ventanas. Algunas de las más relevantes son:

- **Ayuda**

Acceso directo a documentación de R, RStudio y paquetes.



Agenda

Definiendo R

La importancia de R

Scripts de R

Entorno de desarrollo de RStudio

Directorios de trabajo **IMPORTANTE**

Proyectos de R

Directorios de trabajo

Un día estarás analizando datos en R y recordarás que hay otras cosas por hacer. Entonces, tendrás que cerrar RStudio, y no regresarás a tu análisis de datos hasta el día siguiente.

Un día estarás trabajando en varios análisis de datos en R simultáneamente y desearás mantenerlos separados.

Un día necesitarás traer datos del mundo exterior a R y luego enviar resultados numéricos y cifras de R de vuelta al mundo exterior.

Para manejar estas situaciones de la vida real, debes preguntarte lo siguiente:

1. **¿Qué de mi análisis es "real", es decir, qué guardaré como un registro duradero de lo que sucedió?**
2. **¿Dónde "vive" mi análisis?**



1. ¿Qué de mi análisis de datos es "real", es decir, qué guardaré como un registro duradero de lo que sucedió?

Los scripts de R.

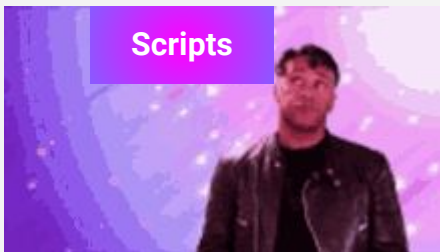
.....

¿Las líneas de código que he corrido en la consola? No.

¿Las variables almacenadas en mi ambiente? No.

¿La historia de cómo fui editando mi código? No.

¿Las líneas de código que he copiado y pegado de múltiples páginas en Internet que he guardado en mis favoritos? No.



Recordatorio

Un script de R es un archivo de texto que contiene líneas de código escritas en el lenguaje de programación de R.

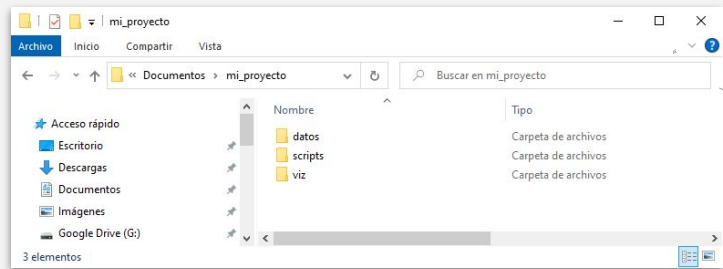
2. ¿Dónde “vive” mi análisis de datos?

En el directorio de trabajo.

.....

El directorio de trabajo es una carpeta en tu equipo local de cómputo y es en donde R busca los archivos que le pidas que cargue y donde colocará los archivos que le pidas que guarde.

Hay un directorio de trabajo establecido por defecto, pero tú puedes siempre restablecerlo a la carpeta de tu proyecto.



Un solo directorio de trabajo (es decir, en una sola carpeta) **por cada** proyecto de análisis de datos.

Agenda

Definiendo R

La importancia de R

Scripts de R

Entorno de desarrollo de RStudio

Directorios de trabajo

Proyectos de R

IMPORTANTE

Proyectos de R: la mejor *buena práctica*

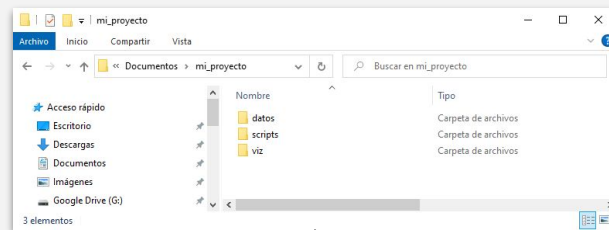
Trabajar proyectos de análisis de datos en un directorios de trabajo independientes es una práctica tan común, que RStudio tiene su propia implementación de *Proyectos*.

Con RStudio puedes:

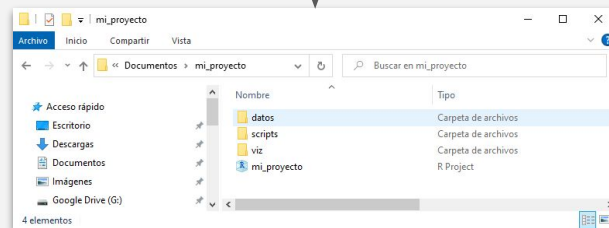
- volver un directorio de trabajo existente un Proyecto.
- crear un Proyecto nuevo que creará un directorio de trabajo nuevo.

Los Proyectos de R te permiten:

- cargar tu directorio de trabajo directamente en el ambiente RStudio en un solo click.
- organizar aún mejor tus proyectos de análisis de datos.



Switching to project ~/mi_proyecto...



Vamos a la práctica.