



Tutorial de R

Galán, Jhon¹; Zequeira, Yainelis²; Pérez, Eric³; Palomino, Isabella⁴

Programación en lenguajes estadísticos, Docente: Ruiz, Jose
Universidad Nacional de Colombia - Sede La Paz

13 de junio de 2022

Abstract: In this paper you will find a short tutorial on R, which is an environment and programming language usually implemented in statistical computing and graphics. It can also be considered as a different implementation of S, that is, as a language and commercial environment increasingly used as an educational language and research tool. Following the guide: “A (very) short introduction to R”, and other web pages, it is sought that the reader can find, in said writing, a guide to the local installation of R, its alternatives of execution on the web, a description and some examples of the basic data structures in that language and their manipulation, as well as options and examples of data visualization.

Key words: R, Programming language, Alternatives of execution on the web.

Resumen: En el presente escrito encontrará un breve tutorial sobre R, el cual es un entorno y lenguaje de programación usualmente implementado en la computación estadística y gráficos. También puede considerarse como una implementación diferente de S, es decir, como un lenguaje y entorno comercial cada vez más usado como lenguaje educativo y herramienta de investigación. Siguiendo la guía “A (very) short introduction to R”, y demás páginas web, se busca que el lector pueda encontrar, en dicho escrito, una guía para la instalación local de R, sus alternativas de ejecución en la web, una descripción y algunos ejemplos de las estructuras de datos básicas en dicho lenguaje y su manipulación, además de las opciones y ejemplos de visualización de datos.

Palabras clave: R, Lenguaje de programación, Alternativas de ejecución en la web.

1. Introducción

R fue implementado por primera vez a principios de la década de 1990 por Robert Gentleman y Ross Ihaka, ambos profesores de la Universidad de Auckland. Dicho lenguaje se vio fuertemente inspirado en el lenguaje S, el cual fue concebido por John Chambers, Rick Becker, Trevor Hastie, Allan Wilks y otros en Bell Labs a mediados de la década de 1970. Su finalidad era propia para la computación estadística y fue puesto a disposición del público a principios de la década de 1980.

R comenzó como un experimento en el cual se buscaba utilizar los métodos de los implementadores de Lisp para construir un pequeño banco de pruebas, en el cual se probaron diversas ideas sobre cómo una estadística podría concebirse en un entorno; Dicho experimento poseía una sintaxis similar a S. Gracias a los esfuerzos emprendidos mediante el uso del Internet R ha superado sus orígenes, posicionándose como una

potente herramienta educativa que permite el análisis, estadísticas y visualizaciones de datos. R, a su vez, se ha implementado desde el mapeo de amplias tendencias sociales y de marketing en línea hasta el desarrollo de modelos financieros y climáticos que ayudan a impulsar nuestras economías y comunidades.

2. Descarga e instalación R

Para incursionar en este lenguaje de programación te brindamos una guía sencilla y práctica para llevar a cabo su instalación.

- Ingresar al sitio web oficial de R: <https://www.r-project.org/>. Donde se puede descargar legalmente.



[Home]

Download
CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Conferences
Search
Get Involved: Mailing Lists
Get Involved: Contributing
Developer Pages
R Blog

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 4.2.1 (Funny-Looking Kid)** prerelease versions will appear starting Monday 2022-06-13. Final release is scheduled for Thursday 2022-06-23.
- **R version 4.2.0 (Vigorous Calisthenics)** has been released on 2022-04-22.
- **R version 4.1.3 (One Push-Up)** was released on 2022-03-10.
- Thanks to the organisers of userR 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the R Consortium YouTube channel.
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

- Hacer clic en **CRAN mirror**.



[Home]

Download
CRAN

R Project

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

- Luego, seleccionar un sitio de descarga.

CRAN Mirrors	
The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some mirrors are on the status of the mirrors can be found here: status page , mirrors.cran.r-project.org , mirrors.cran.r-project.org	
If you want to host a new mirror at your institution, please have a look at the CRAN Mirror HOWTO	
0-Cloud	https://cloud.r-project.org/
Argentina	https://mirrors.ftpserver.com/cran/
Australia	https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/
Austria	https://mirrors.cran.r-project.org/
Belgium	https://www.derepository.org/cran/ https://derepository.org/cran/
Brazil	https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/
Bulgaria	https://mirrors.cran.r-project.org/
Canada	https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/
Chile	https://mirrors.cran.r-project.org/
China	https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/ https://mirrors.cran.r-project.org/

- Descargar R según tu sistema operativo (Windows, Linux, macOS).

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contrib packages. **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

- Hacer clic en base. (ejemplo R for Windows).

R for Windows

Subdirectories:

- [base](#)
- [contrib](#)
- [old-contrib](#)
- [Rtools](#)

Binaries for base distribution. This is what you want to [install R for the first time](#)

Binaries of contributed CRAN packages (for R >= 3.4.x)

Binaries of contributed CRAN packages for outdated versions of R (for R < 3.4.x)

Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

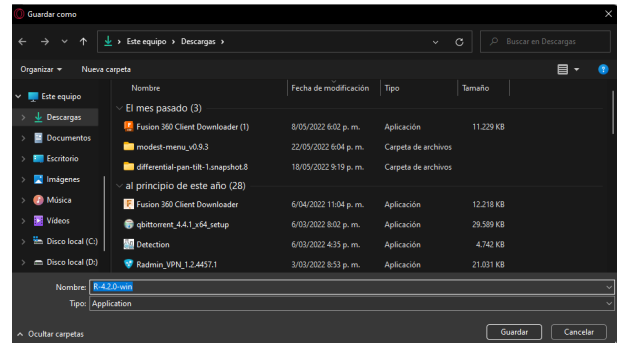
- Descargar la versión actual. (ejemplo R for Windows).

[Download R-4.2.0 for Windows](#) (79 megabytes, 64 bit)

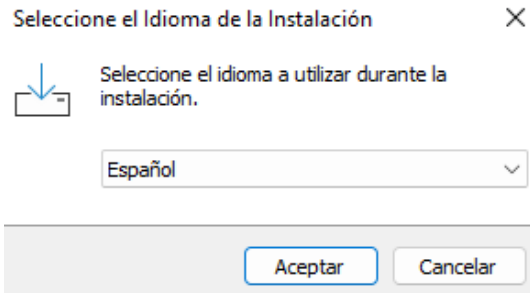
[README on the Windows binary distribution](#)

[New features in this version](#)

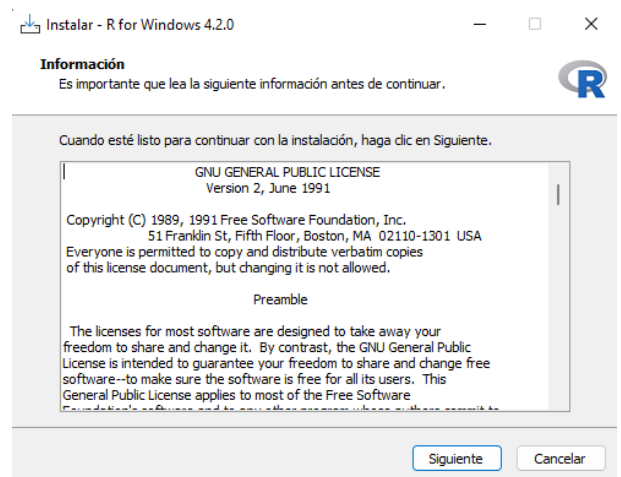
- Seleccionar la ruta de descarga. (en caso de ser necesario).



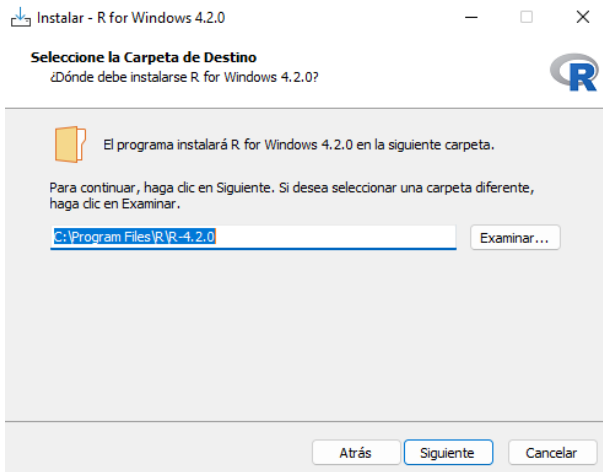
- Ejecutar el instalador y seleccionar el idioma a utilizar durante la instalación.



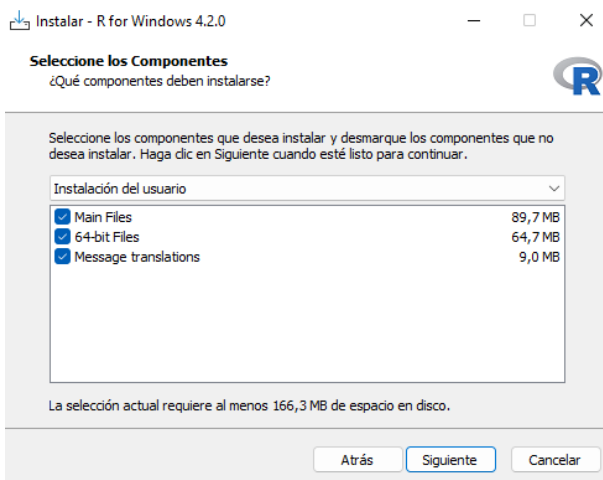
- Dar click en siguiente (leer los términos y condiciones (opcional)).



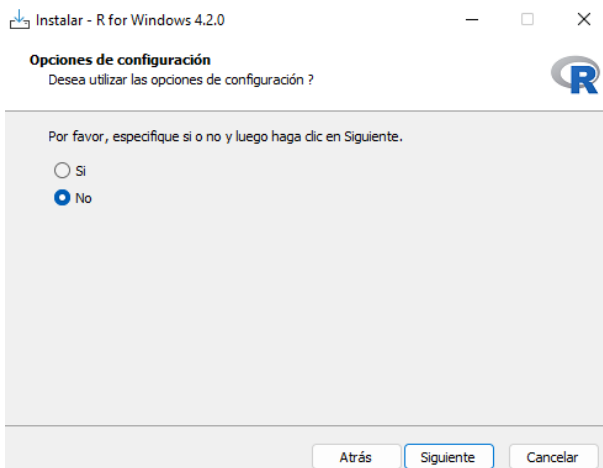
- Seleccionar la ruta de instalación y dar click en siguiente.



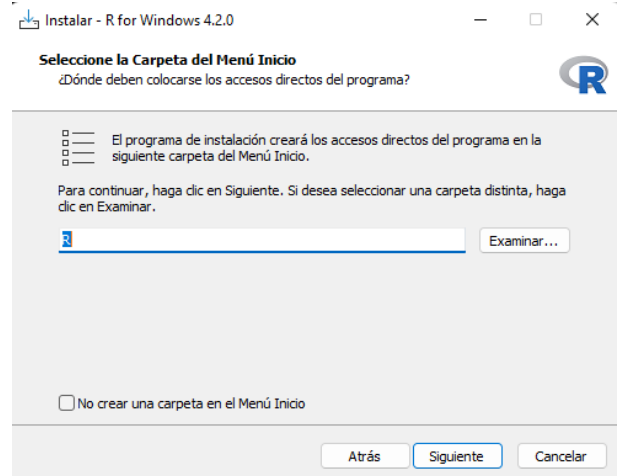
- Seleccionar los componentes y dar click en siguiente.



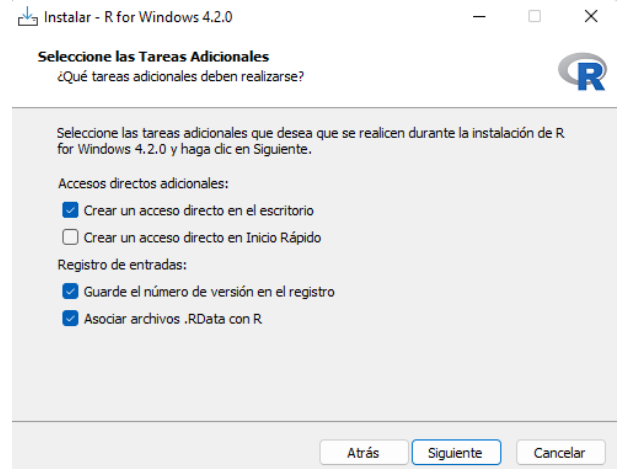
- Seleccionar las opciones de configuración y hacer click en siguiente.



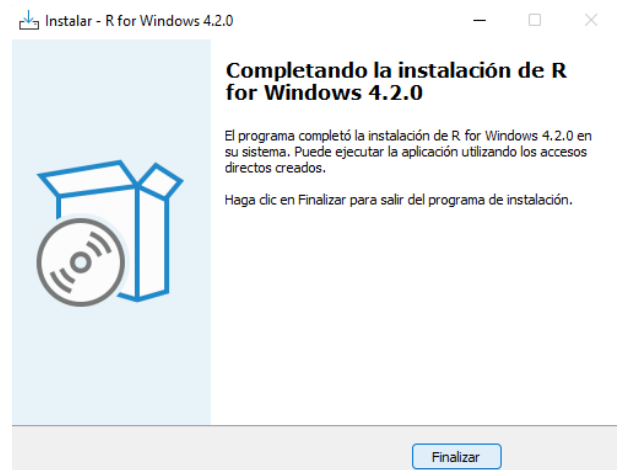
- Seleccionar la carpeta del menú de inicio y hacer click en siguiente.



- Seleccionar las tareas adicionales y hacer click en siguiente.



- Una vez terminado el proceso de instalación hacer click en finalizar.

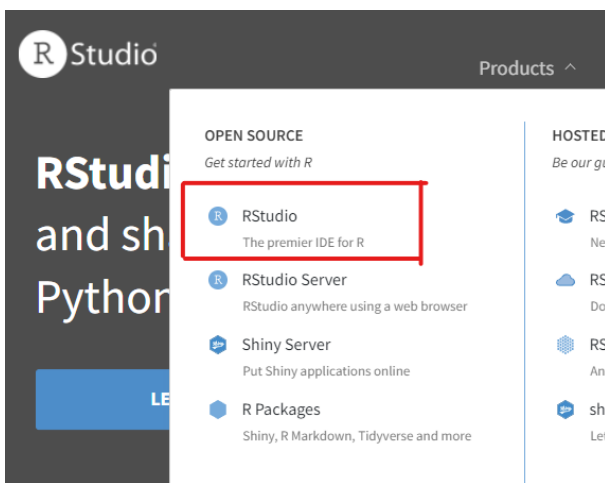


3. Descarga e instalación de Rstudio

Ahora procederemos a instalar Rstudio, el cual es un IDE (entorno de desarrollo integrado) que cuenta con

diversas herramientas que nos brindan un manejo de R más cómodo y versátil.

- Ingresar al sitio web oficial de RStudio: <https://www.rstudio.com> pasar el cursor sobre la pestaña de productos y seleccionar RStudio.



- Hacer click en la versión de escritorio.

There are two versions of RStudio:



- Elegir la opción Open Source.



Open Source Edition

- Access RStudio locally
- Syntax highlighting, code completion, and smart indentation
- Execute R code directly from the source editor
- Quickly jump to function definitions
- View content changes in real-time with the Visual Markdown Editor
- Easily manage multiple working directories using projects
- Integrated R help and documentation
- Interactive debugger to diagnose and fix errors
- Extensive package development tools

Overview

Support	Community forums only
License	AGPL v3
Pricing	Free

[DOWNLOAD RSTUDIO DESKTOP](#)

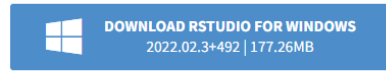
- Elegir el plan gratuito.

RStudio Desktop	RStudio Desktop Pro
Open Source License	Commercial License
Free	\$995 /year
DOWNLOAD	BUY
Learn more	Learn more

- Descargar la versión de escritorio recomendada para tu sistema operativo..

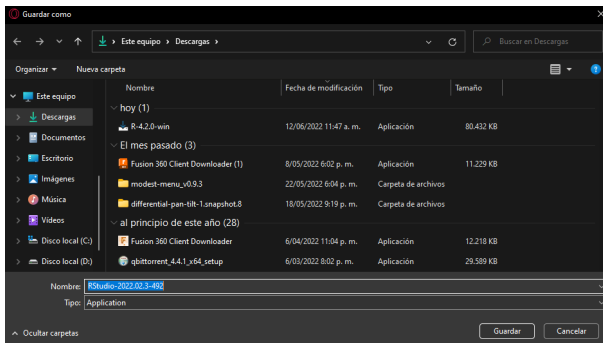
RStudio Desktop 2022.02.3+492 - Release Notes

1. Install R. RStudio requires R 3.3.0+.
2. Download RStudio Desktop. Recommended for your system:

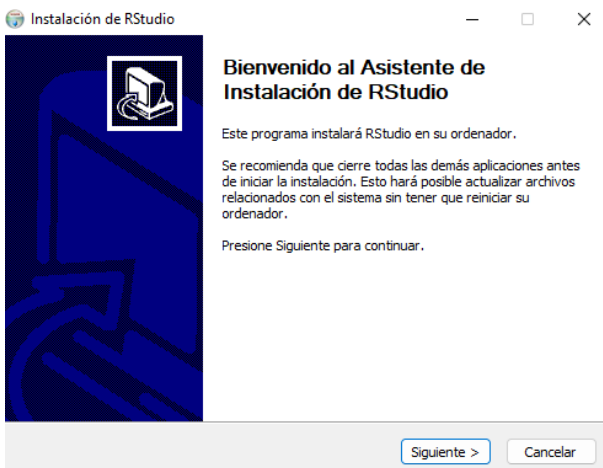


Requires Windows 10/11 (64-bit)

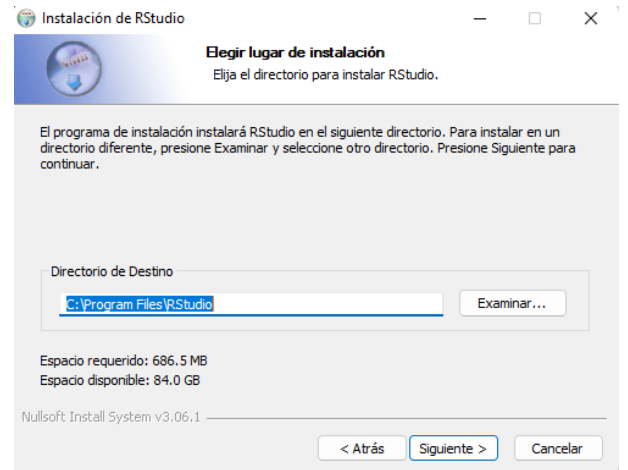
- Seleccionar la ruta de descarga. (en caso de ser necesario).



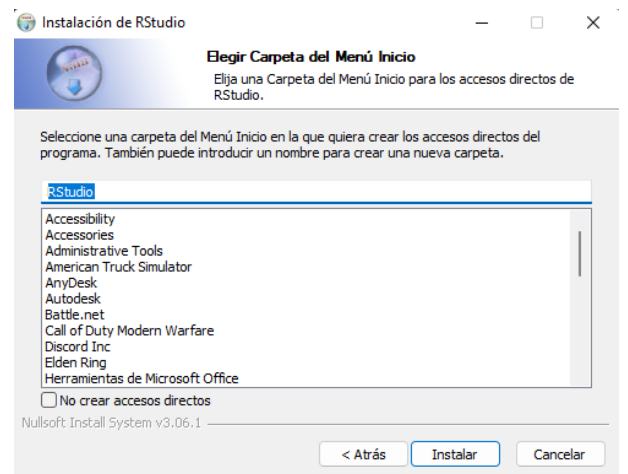
- Descargar el instalador, ejecutarlo y hacer click en siguiente.



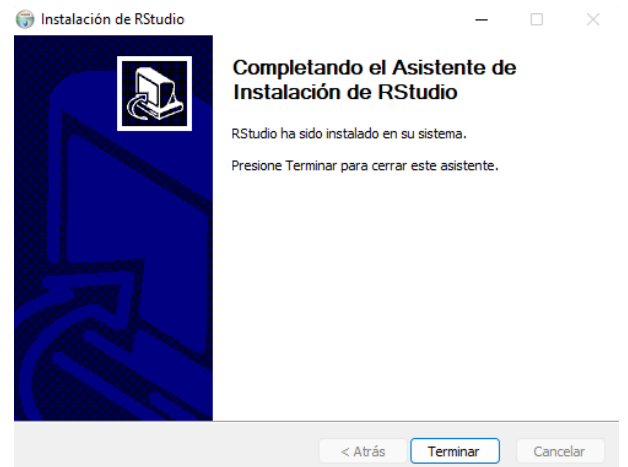
- Seleccionar la ruta de instalación.



- Elegir carpeta del menú de inicio.



- Una vez terminado el proceso de instalación hacer click en finalizar.



4. Alternativas de ejecución de R en la web

Otra forma de utilizar R es ejecutándolo de manera remota, es decir, en línea. En estas alternativas en-

contramos diversas opciones, siendo unas mejores que otras ya sea por la opción de visualización de datos o la necesidad de registrarse para poder emplear dicha herramienta

Entre estas opciones nos encontramos con [Rdrr.io](#), el cual presenta una interfaz sencilla pero con el que podremos no simplemente realizar operaciones sino también importar librerías, paquetes y realizar gráficas para la visualización de datos y esto sin necesidad de un registro previo.

Interfaz:

```
library(ggplot2)

# Use stdout as per normal...
print("Hello, world!")

# Use plots...
plot(cars)

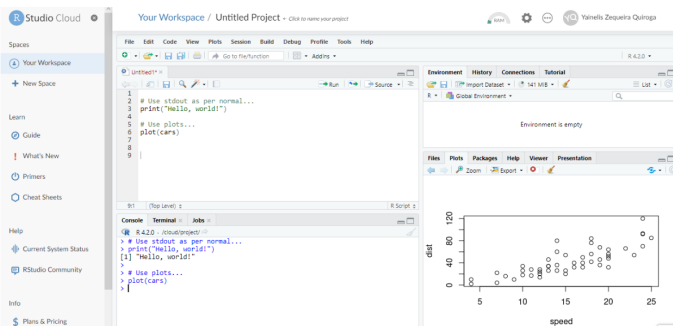
# Even ggplot!
qplot(mpg, data = mtcars, colour = factor(cyl))
```

Run (Ctrl-Enter)

Any scripts or data that you put into this service are public.

Otra opción bastante conveniente es [Rstudio Cloud](#). No obstante, para utilizar esta alternativa primero debemos registrarnos en el sitio web, una vez hecho podremos emplear las características propias de Rstudio pero de manera limitada (según el plan seleccionado al momento del registro)

Interfaz:



5. Estructuras de datos básicas

Las estructuras de datos son una forma determinada de organizar datos para emplearlos de manera más eficiente. Dentro de las estructuras de datos más básicas en R encontramos las siguientes:

- **Vector:** Un vector es la estructura de datos más sencilla en R. Un vector es una colección de uno o más datos del mismo tipo. Los vectores se utilizan de acuerdo a tres propiedades: tipo, largo y atributo.

- **Creación de vectores:** Para crear un vector se utiliza la función `c()` (combinar). Una vez llamada esta función, se debe entregar como argumento los elementos que deseamos combinar en un vector, separados por comas.

Vector numérico `c(1, 2, 3, 5, 8, 13)`

Imprime : [1] 1, 2, 3, 5, 8, 13

Vector de cadena de texto
`c("árbol", "casa", "persona")`

Imprime: [1] árbol , casa , persona

Para agregar un elemento a un vector ya existente, se combina el vector original con los nuevos elementos y se asigna el resultado al vector original.

`vector <- c("árbol", "casa", "persona")`

`vector <- c(vector, "gato", "perro")`

vector

Imprime: [1] árbol , casa , persona , gato , perro

En este orden de ideas, se pueden crear vectores que son combinaciones de vectores.

`v <- c(1, 3, 5)`

`u <- c(2, 4, 6)`

`b <- c(v, u)`

b

Imprime: [1] 1 3 5 2 4 6

- **Vectorización de operaciones:** Existen algunas operaciones que al aplicarlas a un vector, se aplican a cada uno de sus elementos. A este proceso le llamamos vectorización. Las operaciones aritméticas y relacionales pueden vectorizarse.

Por ejemplo, creamos un vector numérico:

`vector <- c(2, 3, 6, 7, 8, 10, 11)`

Si aplicamos operaciones aritméticas, obtenemos un vector con un resultado por cada elemento

Operaciones aritméticas

```
vector + 2
Imprime: [1] 4 5 8 9 10 12 13
vector * 2
Imprime: [1] 4 6 12 14 16 20 22
vector %% 2
Imprime: [1] 0 1 0 1 0 0 1
```

Al aplicar operaciones relacionales, obtenemos un vector de TRUE y FALSE, uno para cada elemento comparado

```
vector > 7
FALSE FALSE FALSE FALSE TRUE TRUE
TRUE
vector < 7
TRUE TRUE TRUE FALSE FALSE FALSE
FALSE
vector == 7
FALSE FALSE FALSE TRUE FALSE FALSE
FALSE
```

- **Matriz:** Es una estructura de datos que almacena elementos del mismo tipo, es decir, elementos homogéneos. En R, una matriz es un vector que tiene asociado dos dimensiones (filas y columnas).

La función **matrix** permite una matriz, dando como resultado un vector columna con tantas filas como la longitud del vector indique.

```
datos = 1 : 10
matrix(datos)
```

Con los parámetros **ncol** y **nrow** es posible establecer el número de filas y el número de columnas, además, con el parámetro **byrow** = FALSE/TRUE se puede indicar si la matriz está ordenada por filas o por columnas.

```
datos = 1 : 10
matrix(datos, ncol = 5, byrow = FALSE) # por columnas.
matrix(datos, ncol = 5, byrow = TRUE) # por filas.
```

Si se tienen los datos almacenados en vectores,

con las funciones **cbind** y **rbind** se pueden concatenar columnas y filas respectivamente.

```
v = c("manzana", "pera", "uva")
v1 = c("piña", "mango", "fresa")
```

```
rbind(v, v1)
```

```
vec1 = c(1, 2, 3, 4, 5)
vec2 = c(4, 9, 7, 5, 1)
c = cbind(vec1, vec2)
dim(c)
length(c)
r = rbind(vec1, vec2)
dim(r)
length(r)
```

La función **rbind** también se puede utilizar para apilar o combinar matrices.

```
vec1 = matrix(c(2, 7, 1, 3, 6, 1), ncol = 3, byrow = FALSE)
```

```
vec2 = matrix(c(3, 7, 6, 3, 5, 9), ncol = 3, byrow = FALSE)
```

```
rbind(vec1, vec2)
```

La función **length** indica cuántos elementos tiene una matriz.

Con la función **dim** se puede conocer las dimensiones de la matriz. La función **dim** también se puede utilizar para crear una matriz en R.

```
A = c(4, 9, 7, 5, 1, 8)
dim(A) = c(3, 2)
```

```
A
```

Con las funciones **rownames** y **colnames** se puede cambiar el nombre de las filas y el nombre de las columnas respectivamente.


```
B = matrix(c(2,4,6,8,10,12,14,16),nrow =
4,ncol = 2,byrow = TRUE)
```

```
rownames(B) =
c("Fila1","Fila2","Fila3","Fila4")
```

```
colnames(B) = c("Columna1","Columna2")//
B
```

```
rownames(B) = NULL
colnames(B) = NULL
```

B

Para eliminar los nombres basta usar el valor **NULL** o usar la función **unname** para borrar todos los nombres.

```
unname(B)
```

Otras funciones con matrices.

Función	
diag	diagonal de una matriz
solve	inversa de una matriz
%*%	producto matricial
t	transpuesta de una matriz
+, -, *, /	operaciones aritméticas a una matriz

Algunos ejemplos

```
a = matrix(1 : 9,nrow = 3,ncol = 3)
b = matrix(1 : 18,nrow = 3,ncol = 6)
```

a

b

```
c = a %* %b #multiplicación de matrices
```

c

```
d = t(b) #transpuesta de una matriz
```

d

```
e = diag(a) #diagonal de una matriz
```

e

```
diag(a) = 0#diagonal reemplazada por ceros
a
```

```
diag(a) = 1 : 3#diagonal reemplazada por un
vector
```

a

```
matriz = matrix(1 : 8,nrow = 2,ncol =
4)matriz
```

```
m1 = matriz * 2
```

m1

```
m2 = matriz + 1
```

m2

```
m3 = matriz/3
```

m3

```
m3 = matriz - 1
```

m3

```
m3 = matriz^2
```

m3

- **Data Frames:** Los data frames son estructuras de datos muy usados en estadística, los cuales cuentan con dos dimensiones (rectangulares) como una matriz, con la diferencia de que pueden contener datos de diferentes tipos, por lo tanto, son heterogéneas.

Internamente un data frame se implementa como una lista, cada componente de la lista corresponde a una columna de dicho data frame. Todas las columnas deben tener la misma longitud. El hecho de que un data frame sea semánticamente similar a una matriz, pero esté implementado como una lista, va a producir una gran riqueza sintáctica a la hora de acceder a sus elementos.

En términos generales, los renglones en un data frame representan casos, individuos u observaciones, mientras que las columnas representan

atributos, rasgos o variables.

```
notas = data.frame(
  "Número" = 1 : 4,
  "Alumno" = c("Jhon", "Isa", "Eric", "Yai"),
  "calificación" = c(96, 75, 88, 55),
  "Valoración" = (c("S", "Ba", "A", "B"))
)
```

En general, se puede aplicar la mayor parte de las funciones asociadas a matrices y a vectores.

`dim(notas)` # dimensiones

`length(notas)` # tamaño del data frame = número de columnas

`ncol(notas)` # número de columnas

`nrow(notas)` # número de filas

`colnames(notas)` # nombres de las columnas

`rownames(notas)` =
`c("Primero", "Segundo", "Tercero", "Cuarto")`

`notas`

- **Propiedades de un dataframe:** Al igual que con una matriz, también se pueden aplicar operaciones aritméticas a un data frame. Los resultados que obtendremos dependerán del tipo de datos de cada columna. R nos devolverá todas las advertencias que ocurran como resultado de las operaciones realizadas, por ejemplo, aquellas que hayan requerido una coerción.

```
mi_df <- data.frame(
  "entero" = 1 : 4,
  "factor" = c("a", "b", "c", "d"),
  "numero" = c(1.2, 3.4, 4.5, 5.6),
  "cadena" = as.character(c("a", "b", "c", "d"))
)
```

```
mi_df * 2
##Warning in Ops.factor(left, right) :
*not
meaningful for factors
##Warning in Ops.factor(left, right) :
*not
meaningful for factors
## entero factor numero cadena
##1 2 NA 2.4 NA
##2 4 NA 6.8 NA
##3 6 NA 9.0 NA
##4 8 NA 11.2 NA
```

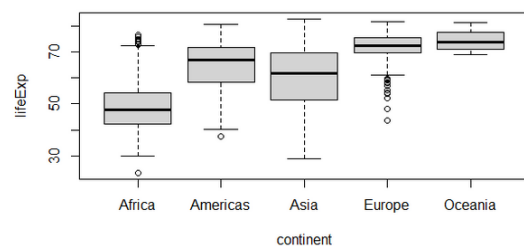
6. Visualización de datos

R se caracteriza por su gran variedad de opciones y herramientas que permiten una visualización de datos gráficamente, ya sean con valores registrados en el mismo programa o de manera externa..

Dentro de las formas de visualización de datos también tenemos las gráficas como los diagramas de dispersión, diagramas de cajas, histogramas, entre otros.

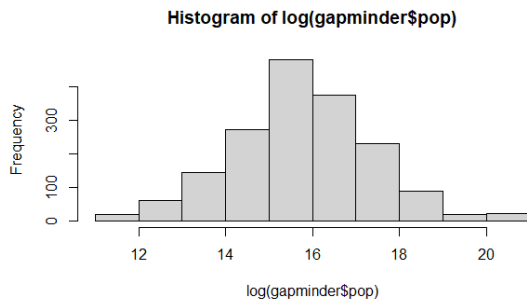
Diagrama de cajas: Es un método estandarizado para representar una serie de datos a través de sus cuartiles. En R, la forma de generar un diagrama de cajas es de la siguiente forma:

`(lifeExp ~ continent)`



Histograma: Es una representación gráfica en forma de barras donde la superficie de cada barra corresponde a la frecuencia de los valores representados. En R, un histograma se obtiene de la siguiente manera:

`(log(gapminder$pop))`

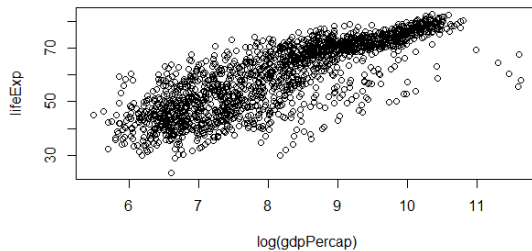


June 13, 2022, from <https://r-coder.com/matrices-r/>

UJAEN. (n.d.). *Tipos de datos* : matrices, listas y data frames. Programación con R. Retrieved June 13, 2022, from <http://www4.ujaen.es/fmartin/R/tipos-de-datos-matrices-listas-y-data-frames-1.html>

Diagrama de dispersión: Es un diagrama que muestra los valores de dos variables para un conjunto de datos, utilizando coordenadas cartesianas. En R, llamamos un diagrama de dispersión como se muestra a continuación:

(lifeExp ~ log(gdpPercap))



7. Referencias

BOOK DOWN. (n.d.). *Data frames*. R para principiantes principiantes. Retrieved June 13, 2022, from <https://bookdown.org/jboscomendoza/r-principiantes4/data-frames.html>

Hornik, K., Leisch, F. (2001, Enero 1). *What is R?* <https://www.r-project.org/doc/Rnews/Rnews2001-1.pdf>

Ihaka, R. (n.d.). *R : Past and Future History*. <https://www.stat.auckland.ac.nz/ihaka/downloads/Interface98.pdf>

Microsoft R Application Network. (n.d.). *What is R?*. MRAN. Microsoft R Application Network. Retrieved June 10, 2022, from <https://mran.microsoft.com/documents/what-is-r>

R CODER. (n.d.). *MATRICES en R con la función matrix()* [CONEJEMPLOS]. R Coder. Retrieved