

# Tutorial de programación con R

## Programming tutorial with R

 Ronald Rivas<sup>1</sup>, Fabián Olivella<sup>1</sup>, and Luis Carmona<sup>1</sup>

<sup>1</sup> Universidad Nacional de Colombia; Escuela de pregrados. Estudiantes de ingeniería biológica.

### Resumen

En muchos artículos académicos es común apreciar el uso de la estadística y el análisis de datos para llegar a una conclusión basada en los datos recopilados de la experimentación. El software estadístico R es uno de los más flexibles, potentes y profesionales que existen actualmente para realizar tareas estadísticas, desde las más elementales hasta las más avanzadas. Está desarrollado y soportado por una comunidad académica a nivel mundial, además es gratuito al público que lo necesite. Entre algunas de las características que hacen a R un lenguaje llamativo y altamente utilizado están que, cuenta con una cantidad de documentación en línea grande además de apoyo en línea. Cada día R mejora añadiendo numerosas librerías y paquetes gratis que expanden su capacidad. R al ser un lenguaje de programación de uso mayormente estadístico cuenta con un gran soporte de estructura de datos que le permiten resolver de forma sencilla y eficiente problemas de alto nivel, además de esto también es capaz de crear variedad de gráficos con una alta calidad. Este documento tiene como objetivo dar información sobre este lenguaje, proporcionando breves explicaciones sobre su instalación, estructuras de datos básicas y métodos u opciones de visualización de datos con la finalidad de proporcionar a las personas principiantes en el lenguaje unas bases para empezar.

**Palabras clave:** Tutorial; R Studio; Programación; Lenguaje R; análisis.

### Abstract

In many academic papers it is common to see the use of statistics and data analysis to reach a conclusion based on data collected from experimentation. The statistical software R is one of the most flexible, powerful and professional software currently available to perform statistical tasks, from the most elementary to the most advanced. It is developed and supported by a worldwide academic community, and it is free to the public that needs it. Among some of the features that make R a striking and highly used language are that it has a large amount of online documentation and online support. Every day R improves by adding numerous free libraries and packages that expand its capabilities. R being a programming language of mostly statistical use has a great support of data structure that allows it to solve in a simple and efficient way high level problems, besides this it is also able to create a variety of graphics with a high quality. This document aims to give information about this language, providing brief explanations about its installation, basic data structures and data visualization methods or options in order to provide beginners in the language a basis to get started.

**Keywords:** Tutorial; R Studio; Programming; Language R; analysis.

### Instalación de R y R Studio

R es un lenguaje de programación gratuito y se encuentra disponible para todos aquellos que lo requieran. Es útil para el análisis y procesamiento de datos estadísticos. R Studio, es el IDE que será usado para utilizar R de forma más cómoda, por el conjunto de herramientas que este contiene y facilitan su uso (Contento, 2020) (UNIR, 2019).

## Instalación de R

1. Para instalar este lenguaje de programación, el primer paso a seguir es entrar al sitio web de R en el equipo en el cual deseen instalarlo para luego dar click en el enlace que dice *download R* que se encuentra el texto titulado como "Getting Started". El link del sitio es <https://www.r-project.org/>.



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

R Foundation

Foundation Board

Members

Donors

Donate

Help With R

Getting Help

# 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 4.2.0 \(Vigorous Calisthenics\)](#) 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 useR! 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](#)

## News via Twitter



The R Foundation Retweeted



R Contributors

@R\_Contributors

Save the date! We are running a Bug BBQ as a satellite event to [@\\_useRconf](#).

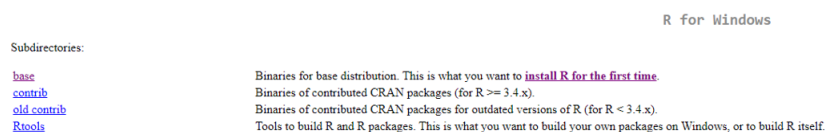
Join any of 3 online sessions over June 23-24 to help work on open bugs in R - both novice and experienced contributors welcome!

More details: [contributor-t-project.org/events/bug-bbq](#)

2. Una vez realizado el paso anterior se abrirá una pestaña con varios enlaces de descarga. Se debe seleccionar el primer enlace que está debajo de donde dice **0-cloud**.

CRAN Mirrors	
The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: <a href="#">main page</a> , <a href="#">windows release</a> , <a href="#">windows old release</a>	
If you want to host a new mirror at your institution, please have a look at the <a href="#">CRAN Mirror HOWTO</a>	
0-Cloud	Automatic redirection to servers worldwide, currently sponsored by Rstudio
Argentina	Universidad Nacional de La Plata
Australia	CSIRO AARNET School of Mathematics and Statistics, University of Melbourne Curtin University
Austria	Wirtschaftsuniversität Wien
Belgium	Patrick Wessa Belnet, the Belgian research and education network
Brazil	Universidade Federal do Paraná Oswaldo Cruz Foundation, Rio de Janeiro University of São Paulo, São Paulo University of São Paulo, Piracicaba
Bulgaria	Sofia University
Canada	Simon Fraser University, Burnaby Manitoba Unix User Group University of Toronto University of Waterloo
Chile	Departamento de Ciencias de la Computación, Universidad de Chile
China	TUNA Team, Tsinghua University Beijing Foreign Studies University Peking University University of Science and Technology of China KoDSoS in Hong Kong Elite Education Lanzhou University Open Source Society eScience Center, Nanjing University Shanghai Jiao Tong University Shanghai University of Science and Technology (SUSTech)

3. Dependiendo del sistema operativo (Windows, Linux o Mac en este caso trabajaremos con Windows) elegiremos la opción de descarga, así como también elegiremos el subdirectorio “base” y daremos click en “*Download R-4.2.0 for Windows*” con un peso de 79MB.

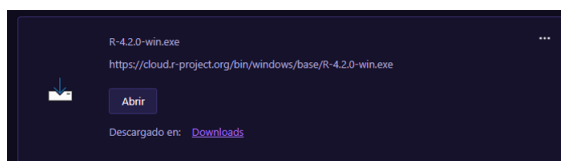


[Download R-4.2.0 for Windows](#) (79 megabytes, 64 bit)

[README on the Windows binary distribution](#)

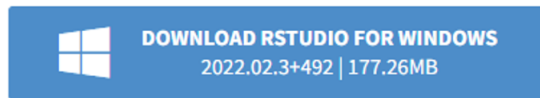
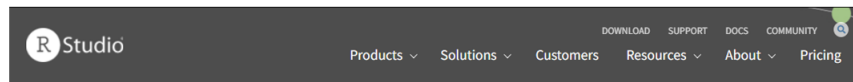
[New features in this version](#)

4. Se ejecuta el archivo con la extensión “.exe” obtenido al dar en el enlace y acomoda las opciones de instalación según lo necesite o simplemente deja las que están por defecto y ya tendría instalado *R* en su computador.



### Instalación de R Studio

1. Para la instalación de R Studio, primero se debe ingresar a la página oficial de el IDE haciendo la búsqueda en el navegador o entrando al siguiente enlace <https://www.rstudio.com>.
2. Una vez dentro de la página, se debe buscar el apartado que dice “**download**”, para luego dar click en “**download for windows**” (en caso de tener el sistema operativo windows). Luego de hacer eso comenzará a descargarse un archivo con la extensión “.exe”.

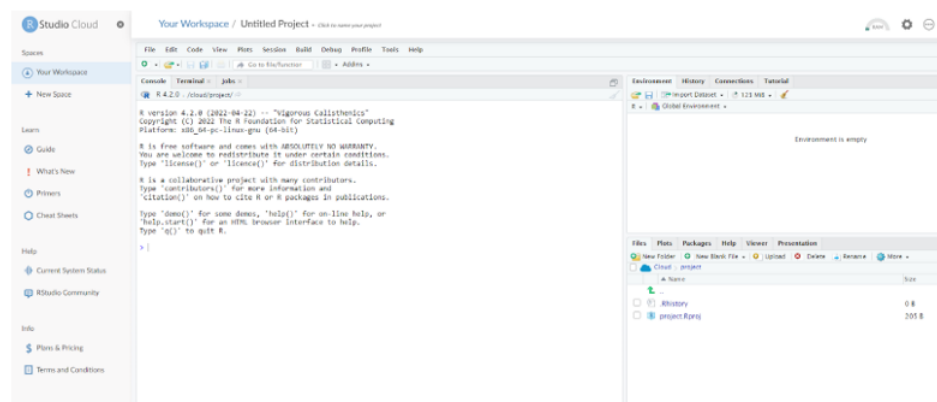
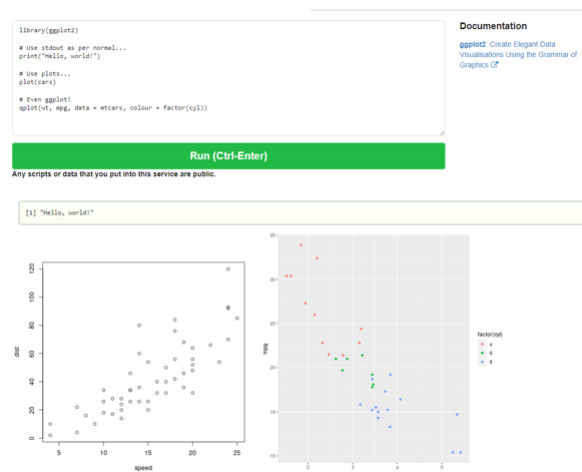


Requires Windows 10/11 (64-bit)

- Una vez termine la descarga del archivo mencionado en el paso anterior, se deberá dar click derecho sobre él para posteriormente darle en *ejecutar como administrador*. Luego se deberán aceptar los términos y condiciones además de elegir las opciones que más convengan o simplemente dejar las que están por defecto.

## Alternativas de R en la web

Existen algunas alternativas al método de instalación local de R y RStudio los cuales vendrían a ser los compiladores Online de R, entre ellos las páginas **rdrr.io** y **RStudio Cloud** que además de compilar códigos de R, usar tanto plot como ggplot, permiten buscar paquetes de R y encontrar documentación para el lenguaje de R.



## Estructura de datos y su manipulación

Las estructuras de datos son aquellas que contienen la información, aunque también son herramientas empleadas en el análisis estadístico o para resolver problemas de alta complejidad de una forma más sencilla. Al hacer uso de R se hacen uso de estas estructuras, algunas estructuras de datos que nos encontraremos en R serán vectores, matrices o data frames (**Platzi**, 2017).

### *Vectores*

Los vectores son las estructuras más sencillas y de una dimensión empleados en R, estos están compuestos por uno o muchos datos de diferentes tipos. Para crear vectores se utiliza principalmente la función `c()` que hace referencia a combinar o concatenar, una vez creado nuestro vector podremos visualizar uno o todos sus elementos y asimismo poder cambiarlos (**Vega**, 2018).

```
1 > vec1 = c(3, 4, 1, 2, 12)
2 > vec1
3 [1] 3 4 1 2 12
4 > vec1[5]
5 [1] 12
6 > vec1[3] = 19
7 > vec1
8 [1] 3 4 19 2 12
```

Al hacer uso de vectores podemos también añadir nuevos elementos a los mismos creando un nuevo vector o mediante la función `append()` :

```
1 > Vec2= c(1, 3, 5)
2 > Vec3= c(2, 4, 6)
3 > Vec4= c(Vec2, Vec3)
4 Vec4
5 [1] 1 2 3 4 5 6
```

Así como también realizar operaciones entre vectores:

```
1 > vec1= c(1, 2, 3)
2 > vec1 * 2
3 [1] 2 4 6
4 > vec1 + 2
5 [1] 3 4 5
```

### *Matrices*

Las matrices son en realidad un tipo de vector pero con algo más en particular, a diferencia de los vectores que son unidimensionales las matrices son vectores bidimensionales. Para crear matrices se utiliza la función `matrix()` en donde al igual que en un vector se ingresan los datos y también como las matrices son de 2 dimensiones se puede determinar tanto el número de columnas (`ncol`) y el de filas (`nrow`) (**Torfs & Brauer**, 2014) (**Paradis**, 2003).

```
1 mat=matrix(data=c(3, 2, 1, 4, 5, 6, 9, 8, 7),ncol=3)
2 > mat
3      [,1] [,2] [,3]
4 [1,] 3    2    1
5 [2,] 4    5    6
6 [3,] 9    8    7
```

Al emplear matrices podemos realizar diversas acciones como con los vectores operaciones aritméticas.

```
1 mat2 = matrix(1:9, nrow = 3, ncol = 3)
2 mat2
3      [,1] [,2] [,3]
4 [1,]  1   4   7
5 [2,]  2   5   8
6 [3,]  3   6   9
7 mat2 * 3
8      [,1] [,2] [,3]
9 [1,]  3  12  21
10 [2,]  6  15  24
11 [3,]  9  18  27
```

### *Data frames*

Una forma sencilla de ver a los data frames es como matrices con títulos o nombres en sus columnas. Los data frames o marco de datos son estructuras de datos bidimensionales (al igual que las matrices) que a diferencia de las matrices estos sí pueden contener datos heterogéneos, son muy utilizados en los análisis estadísticos tanto así que te resultará familiar a los datos presentes en hojas de excel (**Paradis**, 2003) (**Navarro**, 2016).

Para crear un data frame se utiliza la función `data.frame` en donde se tendrá que ingresar la cantidad de vectores correspondientes a las columnas (estos vectores deben tener la misma cantidad de datos).

```
1 RegistroEdades <- data.frame("entero" = 1:4,
2      "Individuo" = c("Carlos", "Brayan", "Nicolas", "Vanessa"),
3      "Edad" = c(16, 18, 20, 15))
4 > RegistroEdades
5      Entero Individuo Edad
6 [1]  1      Carlos   16
7 [2]  2      Brayan   18
8 [3]  3     Nicolás   20
9 [4]  4      Vanessa   15
```

Al igual que con las otras estructuras de datos, con los data frames también se pueden hacer manipular para modificar la información que este lleva dentro, como por ejemplo, para añadir una nueva columna al data frame:

```
1 RegistroEdades$"Decimales" = c(1.2, 3.4, 4.5, 5.6)
2 > RegistroEdades
3      Entero Individuo Edad Decimales
4 [1]  1      Carlos   16    1.2
5 [2]  2      Brayan   18    3.4
6 [3]  3     Nicolás   20    4.5
7 [4]  4      Vanessa   15    5.6
```

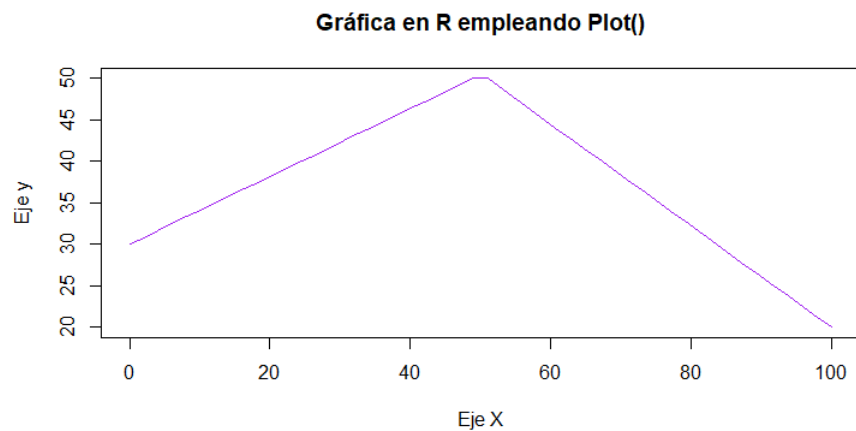
## Visualización

El graficar o visualizar los datos nos permite observar el comportamiento de los mismos, es una actividad que toma una gran importancia a la hora de hacer análisis estadísticos o proponer hipótesis, debido a esto R posee una gran variedad de librerías así como también varias opciones de gráficos como `plot()`, `hist()`, `barplot()`, `pie()`, `boxplot()` (**Torfs & Brauer**, 2014).

### ***Plot***

Representa una de las formas más simples de graficar en R, las formas de graficar empleando esta función dependerá del tipo de datos que se le ingresen. La función plot() pedirá valores con respecto a 2 variables tal como plot(x, y) en donde x corresponde al eje X y la Y a su respectivo eje (**Santana & Hernández, 2016**).

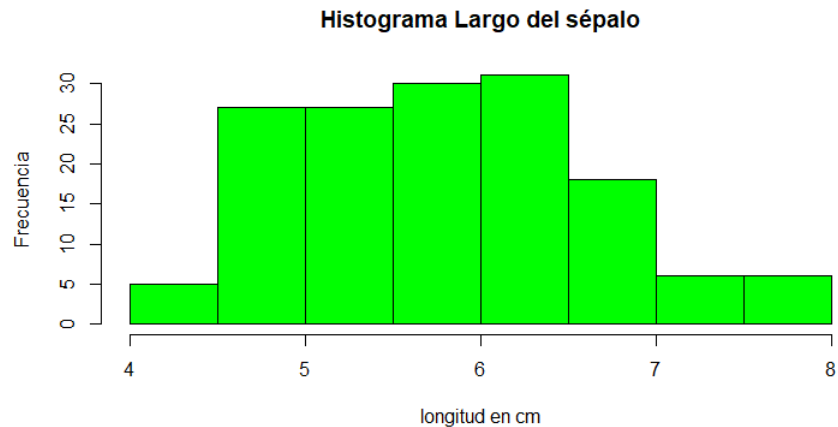
```
1 x <- seq(0,100,length=50)
2 y <- c(seq(30,50, length=25), seq(50,20, length=25))
3 x
4 y
5 plot(x,y, xlab= "Eje X", ylab= "Eje y", type= "l", col= "Purple",
      main= "Gráfica en R empleando Plot()")
```



### ***Hist***

Representa un histograma para diferentes variables continuas teniendo en cuenta la distribución de frecuencia de los datos continuos, permitiéndonos de esta forma identificar asimetrías o valores atípicos (**Santana & Hernández, 2016**).

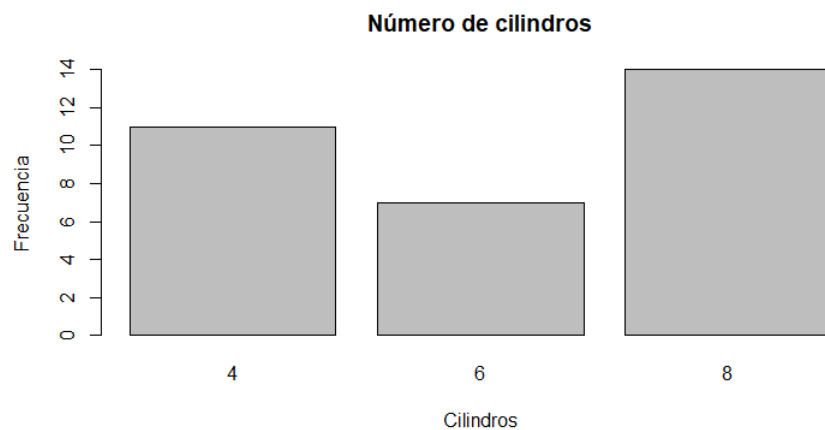
```
1 iris
2 hist(iris$Sepal.Length, main="Histograma Largo del sépalo",
3      xlab="longitud en cm", ylab="Frecuencia", col="green")
```



### ***Barplot***

Es utilizado para representar diagramas de barras, parecido al histograma con la diferencia de que este gráfica variables discretas en barras separadas en lugar de pegadas (**Santana & Hernández, 2016**).

```
1 mtcars
2 x<- table(mtcars$cyl)
3 barplot(x, main="Número de cilindros", xlab= "Cilindros",
4         ylab="Frecuencia")
```



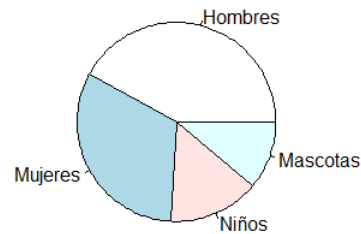
### ***Pie***

Nos ofrece la misma información que un gráfico de barras (función barplot()) pero en forma de diagrama de torta o gráfico circular (**Santana & Hernández, 2016**).

```
1 Porcentajes<- c(42,32,15,11)
2 labels<- c("Hombres","Mujeres","Niños","Mascotas")
3 pie(Porcentajes,labels,main="Distribución de población doméstica")
```



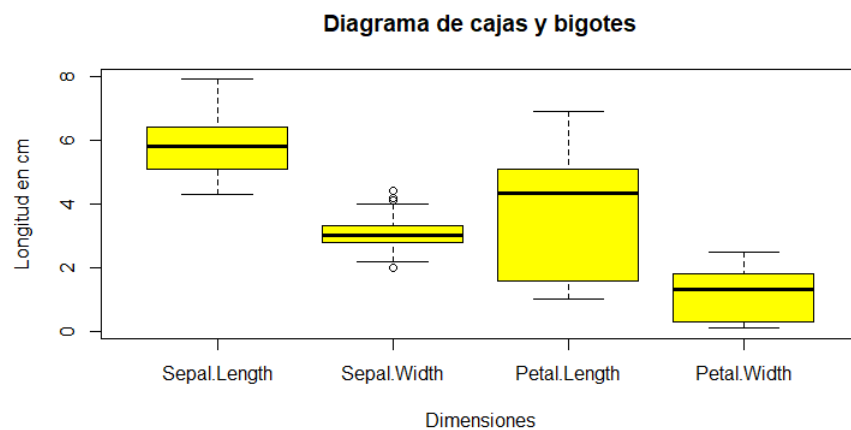
### Distribución de población doméstica



### Boxplot

Representa una gráfica de cajas y bigotes, esta gráfica asu vez nos muestra 5 números importantes en el análisis estadístico siendo estos el número o valor mínimo, el primer cuartil, el segundo cuartil que corresponde a la mediana, el tercer cuartil y el valor máximo (Santana & Hernández, 2016).

```
1 iris
2 boxplot(iris[, -5], main="Diagrama de cajas y bigotes",
3         xlab="Dimensiones", ylab="Longitud en cm", col="yellow")
```



### References

- Contento, M. R.** (2020). *Estadística con aplicaciones en r*. Universidad de Bogotá Jorge Tadeo Lozano. doi: 10.21789/9789587252729
- Navarro, D.** (2016). *Learning statistics with r: A tutorial for psychology students and other beginners*. Retrieved from <https://learningstatisticswithr.com/lsr-0.6.pdf>
- Paradis, E.** (2003). *R para principiantes*. Retrieved from <https://cran.r-project.org/doc/contrib/rdebuts.es.pdf>

- Platzi.** (2017). *¿qué son y para qué sirven las estructuras de datos?* Retrieved from <https://platzi.com/blog/estructuras-de-datos-que-son/>
- Santana, A., & Hernández, C. N.** (2016). *Gráficos en r: introducción.*
- Torfs, P., & Brauer, C.** (2014, 3). *A (very) short introduction to r.* Retrieved from <https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf>
- UNIR.** (2019, 11). *Lenguaje r, ¿qué es y por qué es tan usado en big data?* Retrieved from <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>
- Vega, M.** (2018). *R para principiantes.* Retrieved from <https://bookdown.org/jboscomendoza/r-principiantes4/quien-usa-r.html#>