

Introducción a R y RStudio

Introducción

Índice

Ideas clave	3
1.1. Presentación	3
1.2. ¿Qué es R?	3
1.3. Un poco de historia	4
1.4. ¿Por qué usar R?	5
1.5. El entorno de trabajo RStudio	6
1.6. Instalación de R y RStudio	7
1.7. Formato del código en el texto	13
1.8. La ayuda en línea de R	14
1.9. Referencias bibliográficas	16

1.1. Presentación

El presente material pretende ser un material introductorio al lenguaje de programación R, dirigido a personas que nunca han usado R ni ningún otro lenguaje de programación, y que no tienen (*no se requiere*) conocimiento previo de probabilidad y estadística.

El objetivo no es que aprendas probabilidad y estadística, ni sus aplicaciones específicas a ciencia de datos, aprendizaje automático, econometría, bioestadística u otras. El objetivo de estos apuntes es que adquieras los fundamentos del uso de R como un lenguaje de programación, desde sus conceptos más elementales, hasta la definición de funciones y generación de gráficos. Para ello, el material estará organizado en una lógica de aprendizaje incremental. Es decir, se enseñarán herramientas de R y RStudio para el análisis de datos mediante diversos ejemplos, comenzando desde los usos más simple hasta cuestiones de mayor complejidad.

1.2. ¿Qué es R?

Empezaremos diciendo que R es un lenguaje de programación interpretado, de distribución libre, bajo Licencia GNU, para el análisis estadístico y gráfico de datos desarrollado por *The R Foundation for Statistical Computing*. Está disponible para todos los sistemas operativos y puede descargarse de forma gratuita desde el sitio web <http://www.r-project.org/>. De la misma forma, RStudio puede descargarse libremente desde el sitio web <https://www.rstudio.com/>.

1.3. Un poco de historia

El lenguaje de programación R fue desarrollado por los neozelandeses **Ross Ihaka** y **Robert Gentleman**, en la Universidad de Auckland, Nueva Zelanda, a principios de la década de 1990, como una reimplementación del lenguaje S. Sus inventores diseñaron el lenguaje para combinar las fortalezas de dos lenguajes existentes, S (Becker et al, 1988) y Scheme (Steel y Sussman, 1975), en sus propias palabras, «The resulting language is very similar in appearance to S, but the underlying implementation and semantics are derived from Scheme. In fact, we implemented the language by first writing an interpreter for a Scheme subset and then progressively mutating it to resemble S» (Ihaka and Gentleman, 1996). El resultado fue bautizado como R «in part to acknowledge the influence of S and in part to celebrate our own efforts» (Ihaka y Gentleman, 1996).

El código fuente de R se publicó por primera vez bajo la Licencia pública general de GNU (GPL) en 1995. Desde mediados de 1997, ha existido el Equipo central de desarrollo de R (el *R Core Team*), que actualmente consta de 20 miembros, y sus nombres están disponibles al escribir `contributors()` en una sesión R. La *Comprehensive R Archive Network* (CRAN) se anunció oficialmente el 23 de abril de 1997. CRAN almacena los archivos ejecutables, el código fuente, la documentación y los paquetes aportados por los usuarios de R. CRAN originalmente tenía 3 espejos y 12 paquetes contribuidos. Actualmente (agosto de 2022) cuenta con 104 espejos y 24.809 paquetes aportados.

La primera versión de R se difundió rápidamente y la expansión es hoy irrefrenable. Desde su creación R se alimenta y crece con los trabajos de investigadores provenientes de prácticamente todas las ramas del conocimiento. Las aportaciones desinteresadas de funciones y librerías de propósito tanto general como específico hacen de R en un entorno dinámico formado por una comunidad en continuo movimiento que se inscribe dentro de la filosofía del software libre.

En la actualidad, el desarrollo de este lenguaje de programación se mantiene activa. La versión más reciente de R al momento de escribir este documento es la 4.2.1 “Funny-Looking Kid” fue publicada en junio de 2022 y diariamente son publicados nuevos paquetes y sus respectivas actualizaciones.

1.4. ¿Por qué usar R?

R es sin duda es el entorno estadístico y de análisis de datos más potente y completo. En R es posible realizar desde procedimientos de recolección, limpieza, todo tipo de análisis y comunicación de datos.

A diferencia de otros programas, para utilizar R necesitamos del uso pleno de la modalidad de sintaxis, esto significa que no emplearemos botones para realizar los análisis, sino que nos comunicaremos con el software de manera directa, a partir de lenguaje (código) computacional. Esto conlleva que el dominio de la sintaxis sea exigente y requiera mucha práctica. Su curva de aprendizaje suele ser bastante empinada al comienzo, pero el esfuerzo inicial al poco tiempo es recompensado sobradamente por las ventajas que ofrece.

Hay muchos motivos por los que utilizar R, de los que destacamos los siguientes:

- ▶ Es un lenguaje inicialmente pensado para la estadística que permite manipular los datos rápidamente y de forma precisa.
- ▶ Se puede automatizar fácilmente, gracias a la creación de *scripts* que automatizan procesos, por ejemplo, leer datos o hacer operaciones con los datos, y hacerlo siempre de forma automática.
- ▶ Puede leer prácticamente cualquier tipo de datos.
- ▶ La visualización de datos en R puede ser simple y muy poderosa.
- ▶ La comunidad R en línea es una de las más amigables e inclusivas de todas las comunidades de programación.

- ▶ Es flexible y extensible debido a la existencia de paquetes creados por expertos de la comunidad R.
- ▶ El entorno de desarrollo integrado (IDE) de RStudio es una herramienta muy poderosa para programar con R porque todo su código, resultados y visualizaciones están juntos en un solo lugar. Además, con [RStudio Cloud](#) puedes programar en R usando RStudio directamente desde un navegador web.
- ▶ R es un entorno de software gratuito para descargar y usar, lo que lo hace más adecuado para particulares que las costosas licencias empresariales como SPSS o SAS.

R es de confianza no solo en el mundo académico, sino también en grandes empresas, como Google, Facebook, Airbnb, Uber y más. Se utiliza en salud, consultoría, gobierno, seguros, energía, finanzas, medios de comunicación, casi en todas partes.

1.5. El entorno de trabajo RStudio

La interfaz de usuario de R no es muy amigable por lo que lo más común es usar R a través de RStudio.

RStudio es un ambiente de desarrollo integrado (o IDE, por su sigla en inglés: *Integrated Development Environment*) para programar en R. Un IDE es un software que nos ayuda programar de una manera más simple y eficiente. Entre algunas de sus múltiples funciones, nos sugiere qué deberíamos continuar escribiendo, nos marca dónde nos hemos confundido y administra nuestros archivos en proyectos. En pocas palabras, RStudio es un programa para manejar R y utilizarlo de manera más cómoda (al menos en algunos aspectos). Básicamente se trata de una máscara para visualizar el software que tiene como principales ventajas (1) el orden y (2) la visualización de los procesos que son llevados a cabo con R, todo de manera simultánea.

RStudio divide la pantalla en ciertos paneles o ventanas que nos facilitan bastante el flujo de trabajo al tener, en simultáneo, el código que estamos escribiendo, los resultados que obtenemos, los gráficos, etc. Las mismas cosas se pueden hacer más rápido, ¡eso es todo! En el tema siguiente, veremos en detalle cuáles son esos paneles y cómo los utilizarlos.

1.6. Instalación de R y RStudio

La manera de instalar R cambia dependiendo del sistema operativo utilices, pero todas tienen en común el uso de **CRAN**.

CRAN es un acrónimo de *Comprehensive R Archive Network* que es el repositorio central de software de R. Está compuesto por una serie de servidores en todo el mundo también conocidos como sitios espejo que son utilizados para distribuir tanto R como sus paquetes. Si no quieres preocuparte por elegir un servidor que esté cerca de tu lugar de residencia, puedes utilizar el servidor en la nube, que lo identificará automáticamente por ti. Para ello, abre el navegador de tu preferencia, dirígete al sitio oficial del servidor en la nube de R, <https://cloud.r-project.org> para comenzar el proceso de descarga.

En el siguiente video “*Instalación de R y RStudio*” se muestra el proceso a realizar para instalar estos programas en Windows:



Accede al vídeo

1.6.1. Descargar e instalación de R en Windows

En la pantalla que se abre siguiendo al enlace anterior, elige [Download R for Windows](#) (ver Figura 2).

The Comprehensive R Archive Network

Download and Install R

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

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [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.

Figura 2: Enlace de descarga del instalador de R para Windows.

Al seleccionar la opción [Download R for Windows](#) se accederá a la siguiente página en la cual se debe hacer clic en el enlace [install R for the first time](#), para descargar el instalador.

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to install R for the first time .
contrib	Binaries of contributed CRAN packages (for R \geq 3.4.x).
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 3.4.x).
Rtools	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.

Figura 3: R base para Windows.

La siguiente página (Figura 4) ofrecerá la opción de descarga de R para Windows. Al presionar el enlace, automáticamente comenzará la descarga del instalador, es decir, un archivo con extensión “.exe” en tu computadora.

R-4.2.1 for Windows

[Download R-4.2.1 for Windows](#) (79 megabytes, 64 bit)
[README on the Windows binary distribution](#)
[New features in this version](#)

Figura 4: Descarga del instalador de R para Windows.

Luego ir a la carpeta de Descargas (o donde se haya descargado), ejecutarlo y seguir las instrucciones de instalación que va ofreciendo el software (poner *Continuar* a todo), hasta completar la instalación. Una vez concluido el proceso, el software habrá quedado instalado en tu computadora.

1.6.2. Descarga e instalación de RStudio en Windows

Una vez instalado R se podrá instalar RStudio, un entorno de desarrollo integrado para el lenguaje R que lo vuelve más fácil de utilizar. Se lo debe descargar de la página <https://www.rstudio.com/>.

La lógica de instalación de RStudio es similar a lo ya presentado para R. En la página oficial de RStudio se debe acceder a la pestaña *Products* y seleccionar la opción *RStudio*.

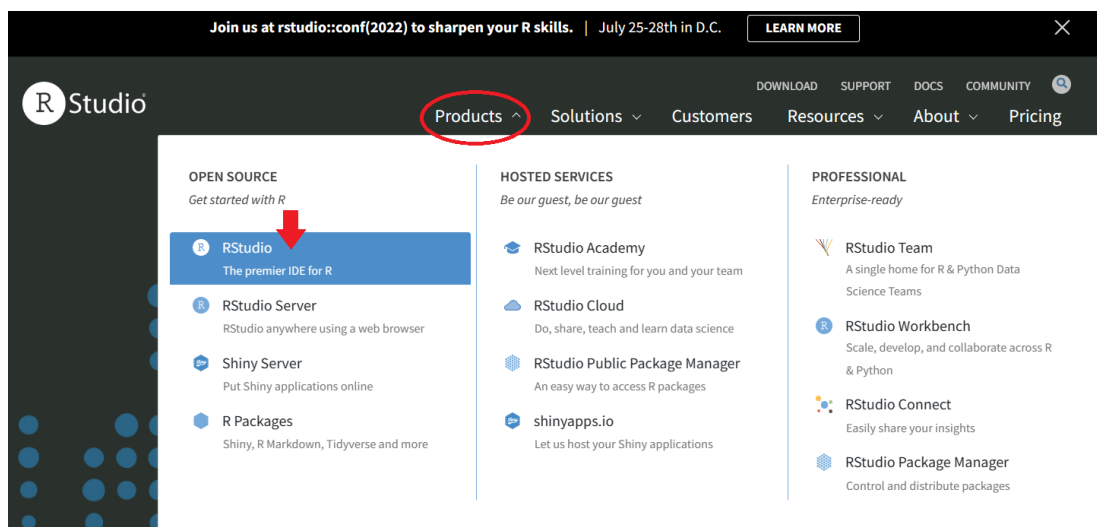


Figura 5: Página de descarga de RStudio.


Dentro de esa página hay que acceder al apartado de descarga del instalador presionando el enlace **RStudio Desktop** (Figura 6).


There are two versions of RStudio:



Figura 6: Enlace de descarga de RStudio Desktop.

En la página a la que nos lleva el enlace anterior, clicamos en **Download RStudio Desktop** para acceder a la página de descarga de las diferentes versiones de los instaladores.

 **Studio Desktop**

	Open Source Edition	RStudio Desktop Pro
Overview	<ul style="list-style-type: none">• 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	<p>All of the features of open source; plus:</p> <ul style="list-style-type: none">• A commercial license for organizations not able to use AGPL software• Access to priority support• RStudio Professional Drivers• Connect directly to your RStudio Workbench  instance remotely
Support	Community forums only	<ul style="list-style-type: none">• Priority Email Support• 8 hour response during business hours (ET)
License	AGPL v3	RStudio License Agreement
Pricing	Free	\$995/year

DOWNLOAD RSTUDIO DESKTOP

←

DOWNLOAD FREE RSTUDIO DESKTOP PRO TRIAL

Figura 7: Acceso a los instaladores de RStudio Desktop.

Dentro de la nueva página que se abrirá, se encuentra la siguiente sección donde se puede descargar el *installer* de RStudio para Windows (también se encuentran para otros sistemas operativos).

All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).


OS	Download	Size	SHA-256
Windows 10/11	RStudio-2022.07.1-554.exe 	190.14 MB	5ab6215b
macOS 10.15+	RStudio-2022.07.1-554.dmg	221.04 MB	7b1a2285
Ubuntu 18+/Debian 10+	rstudio-2022.07.1-554-amd64.deb	132.91 MB	74b9e751
Ubuntu 22	rstudio-2022.07.1-554-amd64.deb	145.33 MB	92f2ab75
Fedora 19/Red Hat 7	rstudio-2022.07.1-554-x86_64.rpm	103.29 MB	0fc15d16
Fedora 34/Red Hat 8	rstudio-2022.07.1-554-x86_64.rpm	149.77 MB	0c4ef334
OpenSUSE 15	rstudio-2022.07.1-554-x86_64.rpm	133.76 MB	45f277d0

Figura 8: Descarga del instalador de RStudio Desktop.

A continuación, ir a la carpeta de Descargas (o donde hayas descargado el archivo) y ejecuta el instalador. Luego de seguir todas las instrucciones del proceso de instalación el programa habrá quedado instalado correctamente. Es necesario enfatizar que en lo posterior **solamente se ejecutará RStudio**, y nunca el software R. Este último queda instalado y RStudio se encarga de ejecutarlo.

1.6.3. Descargar e instalación de R en macOS

Como lo indicamos, dirígete al sitio oficial del servidor en la nube de R y haz clic en el enlace [Download R for macOS](#).

The Comprehensive R Archive Network

Download and Install R

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

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [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.

Figura 9: Enlace de descarga del instalador de R para macOS.

En la página que se abre (Figura 10) selecciona la versión de R para descargar su archivo de instalación.

↓
[R-4.2.1.pkg](#) (notarized and signed)
SHA1 hash: R34fc96e0dd1919325594e01381a273073a3a
(ca. 90MB) for Intel Macs

Latest release:

R 4.2.1 binary for macOS 10.13 (High Sierra) and higher, **Intel 64-bit** build, signed and notarized package. Contains R 4.2.1 framework, R.app GUI 1.79 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 6.7. The latter two components are optional and can be omitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

Note: the use of X11 (including `tcltk`) requires [XQuartz](#) to be installed (version 2.7.11 or later) since it is no longer part of macOS. Always re-install XQuartz when upgrading your macOS to a new major version.

This release supports Intel Macs, but it is also known to work using Rosetta2 on M1-based Macs. For native Apple silicon arm64 binary see below.

Important: this release uses Xcode 12.4 and GNU Fortran 8.2. If you wish to compile R packages from sources, you may need to download GNU Fortran 8.2 - see the [tools](#) directory.

Figura 10: Instalador de R para macOS.

Desde allí se puede acceder a la última versión de R si es que es compatible con la correspondiente versión del sistema operativo macOS (en este caso macOS 10.13 *High Sierra* o superior). Haciendo clic sobre el enlace, comienza la descarga.

Una vez descargado, busca el instalador descargado y haciendo doble clic sobre tal archivo se abrirá la ventana de instalación. La interfaz gráfica te irá guiando, para continuar con la instalación. Usa las opciones recomendadas por el instalador si tienes duda.

Para corroborar la correcta instalación de R, puedes chequear que el programa se encuentre disponible en el *launchpad*, en donde aparecen todas las aplicaciones instaladas en el ordenador.

1.6.4. Descargar e instalación de RStudio en macOS

Para la instalación de RStudio en macOS hay que seguir el mismo tipo de procedimiento ya indicado para Microsoft Windows. En la página de descargas de los instaladores de RStudio Desktop (Figura 8) se debe seleccionar la opción denominada macOS 10.15+. Al hacer clic, se descargará automáticamente la última versión disponible, para la última actualización de sistema operativo.

Una vez descargado, inicia el instalador, generalmente ubicado en su carpeta de descargas, haciendo doble clic sobre el archivo lo que abrirá la ventana de instalación. En ella deberás arrastrar el icono de RStudio al de aplicaciones (Applications) para concretar la instalación del software.

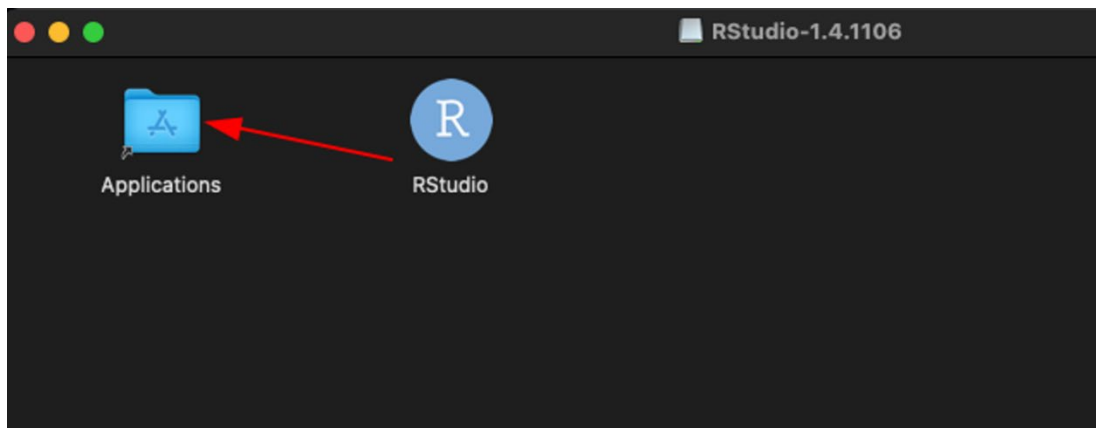


Figura 11: Instalación de RStudio en macOS.

Para corroborar la correcta instalación de RStudio, puedes chequear que el programa se encuentre disponible en el *launchpad*, en donde aparecen todas las aplicaciones instaladas en el ordenador.

1.7. Formato del código en el texto

A lo largo del apunte, los ejemplos del código R aparecerán sobre un fondo gris. Los comentarios estarán resaltados en color verde. El resultado de lo que se envía en la consola aparecerá con `##` delante del código para hacer la distinción entre el código y el resultado del código. A continuación, se muestra un fragmento de código con las explicaciones correspondientes.

```
# Este es un comentario.
# En R los comentarios empiezan a partir del carácter '#'.
# -----
# Asignar variables: son todos equivalentes
x = 10
x <- 10
10 -> x
# Ahora se imprimirá el valor de la variable dos veces, la primera
# vez se hará de manera explícita por medio de la función print(),
# como sigue:
print(x)
## [1] 10
# o, simplemente escribiendo en consola:
x
## [1] 10
# Notar que las impresiones o resultados de las operaciones
```

```
# aparecen como comentarios, pero iniciados con '##', seguido de [1] si es  
# un resultado en una sola línea.
```

1.8. La ayuda en línea de R

Ayuda de una función específica

En muchas ocasiones necesitaremos ayuda sobre cómo funciona una determinada función, cuáles son sus argumentos, etc. La ayuda en línea de R proporciona información muy útil de cómo utilizar las funciones. Para obtener la información de una función en específico, por ejemplo `mean()`, el comando es:

```
help(mean) # o help("mean")  
# Una forma abreviada sería  
?mean
```

La ejecución de cualquiera de las líneas anteriores, mostrara dentro de R, ayuda para la función `mean()` (cálculo del promedio). Al llamar la ayuda, se abrirá una ventana en la pestaña *Help* con información general sobre la función. Un fragmento de la salida de ayuda que ofrece el sistema para ese tema (`mean()`) se muestra en la Figura 12.

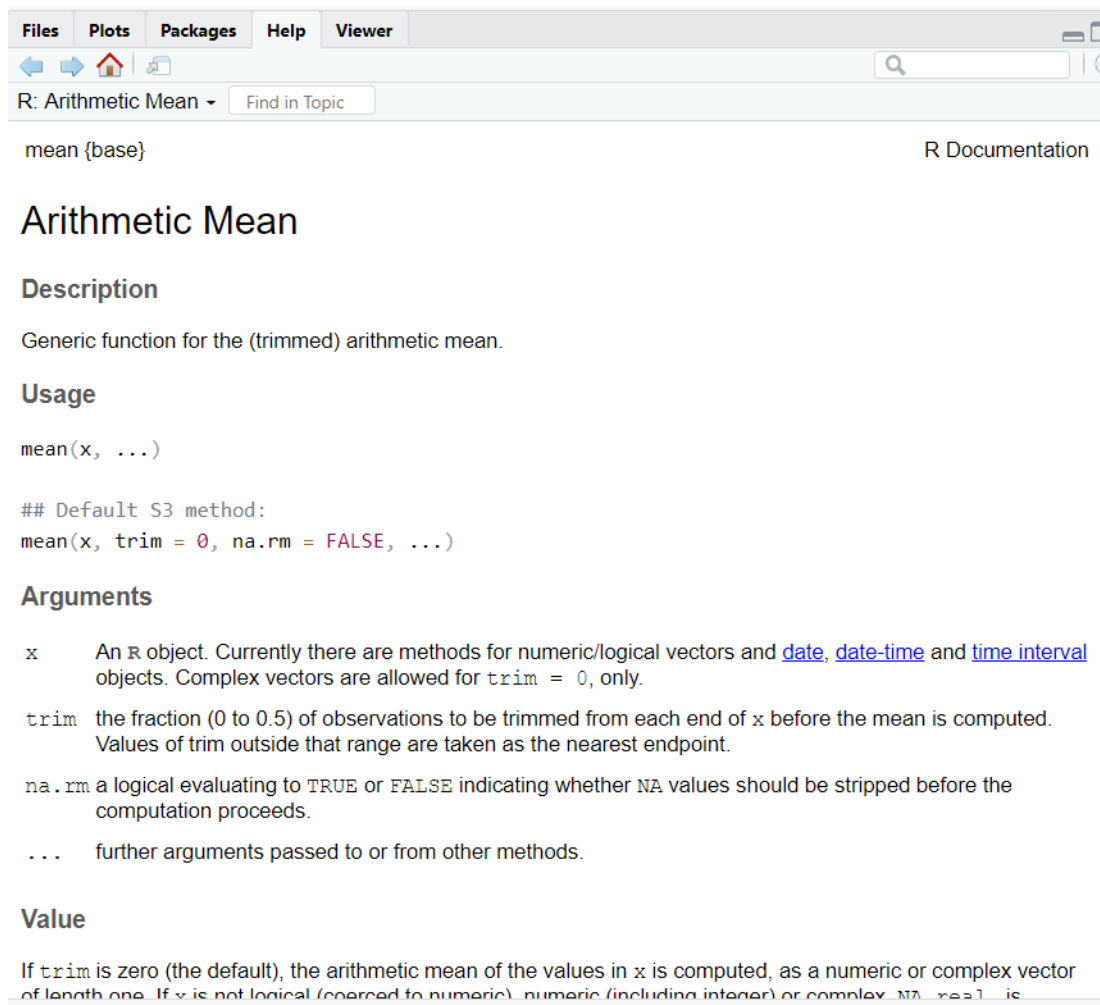


Figura 12: Ventana de ayuda para la función mean().

Es muy recomendable acostumbrarse a usar la ayuda de R, ya que, usualmente, cada explicación viene acompañada de ejemplos de uso.

Ayuda de un paquete

Equivalentemente, podemos obtener ayuda detallada sobre un paquete de R especificándolo en el argumento `package`.

```
help(package = "stats")
```

La función `help.search`

A veces no sabemos el nombre exacto de la función de la que se quiere obtener la documentación. En estos casos la función `help.search()` busca la documentación en todos los paquetes instalados que contengan cierta palabra, **aunque no estén cargados en la memoria**. Por ejemplo, si no sabes qué función utilizar para calcular la mediana de un conjunto de datos, podríamos utilizar el siguiente código:

```
help.search("median")
# Dentro de la función también puedes especificar el paquete
# en el que quieres que se realice la búsqueda.
help.search("median", package = "stats")

# Equivalentemente, se podría usar el operador '??'.
??median

# Si quieres obtener más información sobre los argumentos de la función
# help.search ejecuta ?help.search.
```

1.9. Referencias bibliográficas

Becker, R.A., Chambers, J.M., y Wilks, A.R. (1988). *The new S language: A programming environment for data analysis and graphics*. Pacific Grove, CA: Wadsworth & Brooks/Cole.

Ihaka, R. Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
<https://www.tandfonline.com/doi/abs/10.1080/10618600.1996.10474713>

Sussman, G.J., Steele, G.L. (1998). Scheme: A Interpreter for Extended Lambda Calculus. *Higher-Order and Symbolic Computation* 11, 405–439.
<https://doi.org/10.1023/A:1010035624696>