

# Ayudantía I: Introducción a R

SOL201S - Análisis Datos III

*true*

*true*

*true*

16/08/2019

## Contents

Objetivos . . . . .	1
¿Qué es R? . . . . .	1
Pasos para instalar R . . . . .	2
¿Qué es RStudio? . . . . .	3
Instalando RStudio . . . . .	3
Conociendo el entorno de RStudio . . . . .	3
R para el análisis de datos sociales . . . . .	3
Interactuando con R y RStudio . . . . .	5
Links de referencia . . . . .	12

## Objetivos

1. Presentar el software estadístico R como una herramienta para el procesamiento y análisis de datos.
2. Presentar RStudio como medio a partir del cual se utiliza R.
3. Identificar los principales conceptos a la hora de comprender el lenguaje de programación en R e interactuar con la plataforma de RStudio.

## ¿Qué es R?

Es un entorno y lenguaje de programación gratuito y de código abierto (freeware). Más que un software estadístico como SPSS o STATA, R es un marco para el análisis estadístico con una fuerte orientación a la ciencia de datos. Se crea bajo un nuevo paradigma de ciencia abierta en 1993 por Ross Ihaka y Robert Gentleman con el foco en 4 aspectos claves:

- **Reproducibilidad**
- **Transparencia**
- **Colaboración**
- **Acceso**

El registro de los análisis queda en formato de texto plano, por lo tanto es independiente de una plataforma para poder editarlo, y además permite un control eficiente de versiones (por ejemplo vía Git). Los análisis operan en base a **paquetes o librerías** (¡actualmente existen más de 3000 librerías disponibles!) y su análisis se orienta principalmente a **objetos** (lo veremos más adelante).

- **Google Scholar Hits**

Fuente: [https://i0.wp.com/r4stats.com/wp-content/uploads/2012/04/Fig\\_2d\\_ScholarlyImpact.png](https://i0.wp.com/r4stats.com/wp-content/uploads/2012/04/Fig_2d_ScholarlyImpact.png)

- **Google Trends Latinoamérica**

Fuente: <https://arcruz0.github.io/libroadp/>

---

## Pasos para instalar R

### 1. Visitar la página de CRAN (Comprehensive R Archive Network)

#### 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](#)
- [Download R for \(Mac\) OS X](#)
- [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.

---

### 2. Seleccionar versión según sistema operativo

#### • Windows:

##### R for Windows

Subdirectories:

<a href="#">base</a>	Binaries for base distribution. This is what you want to <a href="#">install R for the first time</a> .
<a href="#">contrib</a>	Binaries of contributed CRAN packages (for R $\geq$ 2.13.x; managed by Uwe Ligges). There is also information on <a href="#">third party software</a> available for CRAN Windows services and corresponding environment and make variables.
<a href="#">old-contrib</a>	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
<a href="#">Rtools</a>	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.

#### • (Mac) OS X:

---

##### R for Mac OS X

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) [here](#). Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old](#) directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

As of 2016/03/01 package binaries for R versions older than 2.12.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting accordingly.

##### R 3.6.1 "Action of the Toes" released on 2019/07/05

**Important:** since R 3.4.0 release we are now providing binaries for OS X 10.11 (El Capitan) and higher using non-Apple toolkit to provide support for OpenMP and C++17 standard features. To compile packages you may have to download tools from the [tools](#) directory and read the corresponding note below.

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type `md5 R-3.6.1.pkg` in the *Terminal* application to print the MD5 checksum for the R-3.6.1.pkg image. On Mac OS X 10.7 and later you can also validate the signature using `pkgutil --check-signature R-3.6.1.pkg`

##### Latest release:

[R-3.6.1.pkg](#)  
MD5-hash: 279e6662103df6a625b4573143cb995  
SHA1-hash: 4e932f8e5013870d2a9179b54eae277f41657b0  
(ca. 76MB)

**R 3.6.1** binary for OS X 10.11 (El Capitan) and higher, signed package. Contains R 3.6.1 framework, R.app GUI 1.7.0 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 5.2. 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 since it is no longer part of OS X. Always re-install XQuartz when upgrading your macOS to a new major version.

**Important:** this release uses Clang 7.0.0 and GNU Fortran 6.1, neither of which is supplied by Apple. If you wish to compile R packages from sources, you will need to download and install those tools - see the [tools](#) directory.

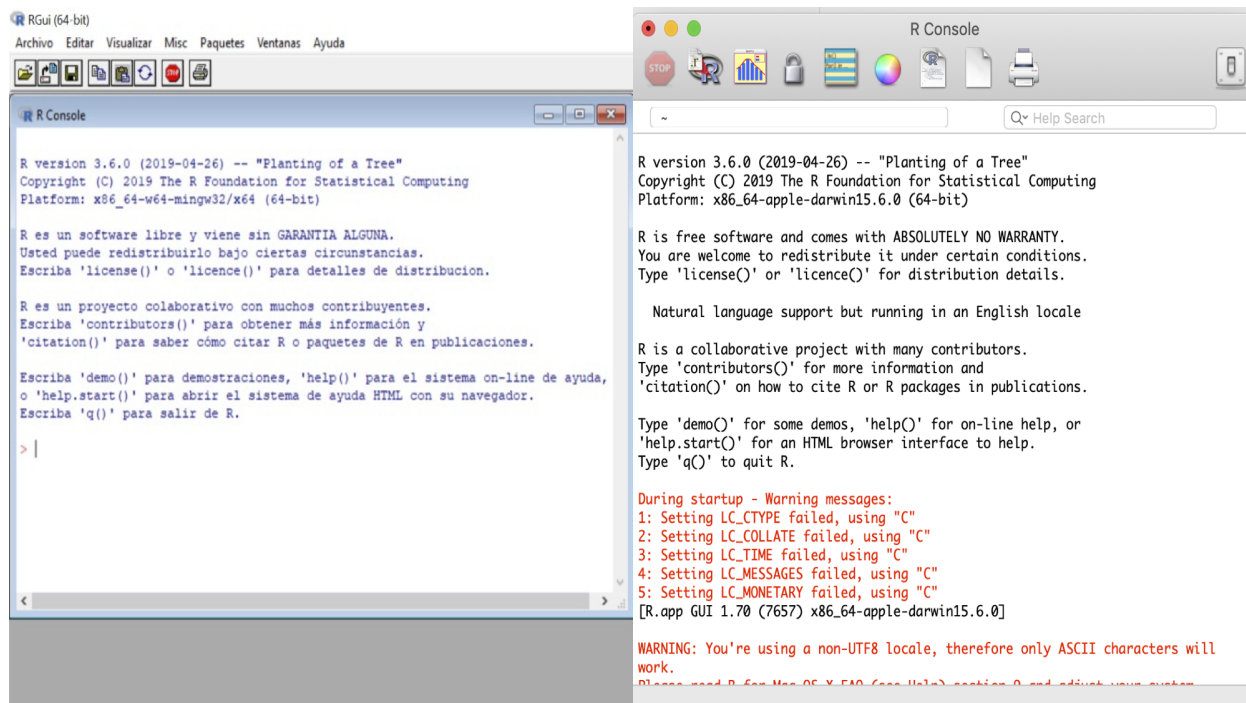


Figure 1: Visualización de R para Windows (izquierda) y Mac (derecha)

### 3. Seguir los pasos de instalación indicados para su sistema operativo y abrir R

## ¿Qué es RStudio?

Es un entorno de desarrollo integrado para R (interfaz gráfica) que permite una navegación más “amigable” y dinámica dentro de la plataforma de R. Cuenta con una consola, editor de resaltado de sintaxis que admite la ejecución directa de código y herramientas para el trazado, el historial, la depuración y la gestión del espacio de trabajo para facilitar el **reporte y reproducibilidad de los análisis**. Esto favorece el desarrollo de procedimientos de ciencia abierta para la investigación.

## Instalando RStudio

### 1. Visitar la página de RStudio.com, descarga y sigue las instrucciones de instalación

## Conociendo el entorno de RStudio

## R para el análisis de datos sociales

Fuente: Fuente: <https://www.r-bloggers.com/i-dont-want-to-learn-r-spss-is-fine-responses/>

## RStudio Desktop 1.2.1335 — Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it [here](#).

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

RStudio 1.2 requires a 64-bit operating system, and works exclusively with the 64 bit version of R. If you are on a 32 bit system or need the 32 bit version of R, you can use an [older version of RStudio](#).

## Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.2.1335 - Windows 7+ (64-bit)	126.9 MB	2019-04-08	d0e2470f1f8ef4cd35a669aa323a2136
RStudio 1.2.1335 - macOS 10.12+ (64-bit)	121.1 MB	2019-04-08	6c570b0e2144583f7c48c284ce299eef
RStudio 1.2.1335 - Ubuntu 14/Debian 8 (64-bit)	92.2 MB	2019-04-08	c1b07d0511469abfe582919b183eee83
RStudio 1.2.1335 - Ubuntu 16 (64-bit)	99.3 MB	2019-04-08	c142d69c210257fb10d18c045fff13c7
RStudio 1.2.1335 - Ubuntu 18/Debian 10 (64-bit)	100.4 MB	2019-04-08	71a8d1990c0d97939804b46cfb0aea75
RStudio 1.2.1335 - Fedora 19/RedHat 7 (64-bit)	114.1 MB	2019-04-08	296b6ef88969a91297fab6545f256a7a
RStudio 1.2.1335 - Debian 9 (64-bit)	100.6 MB	2019-04-08	1e32d4d6f6e216f086a81ca82ef65a91
RStudio 1.2.1335 - OpenSUSE 15 (64-bit)	101.6 MB	2019-04-08	2795a63c7efd8e2aa2dae86ba09a81e5
RStudio 1.2.1335 - SLES/OpenSUSE 12 (64-bit)	94.4 MB	2019-04-08	c65424b06ef6737279d982db9eefcae1

Figure 2: Fuente: <https://www.rstudio.com/resources/cheatsheets>

## Áreas

de

comparación

Comentarios

Curva de aprendizaje	La interfaz gráfica de <b>SPSS</b> permite el uso más intuitivo del programa, mientras que <b>R</b> está orientado al código. Si bien la curva de aprendizaje de <b>R</b> es mayor, una vez comprendido su lenguaje se vuelve cada vez más simple incorporar nuevos conocimientos.
Variedad de técnicas	Cuando el objetivo de la investigación exige una técnica estadística más avanzada, <b>R</b> tiene ventajas sobre <b>SPSS</b> . Al poseer diversos paquetes estadísticos, la cantidad de técnicas de análisis es superior y como es abierto, cada vez personas al rededor del mundo contribuyen con nuevas herramientas. Por ejemplo, toda la rama de técnicas de análisis factorial no se encuentran disponibles en <b>SPSS</b> .
Herramientas gráficas	Las capacidades gráficas de <b>R</b> son sumamente superiores que las de <b>SPSS</b> , ofreciendo una diversidad amplia de herramientas para poder construir y personalizar las gráficas. <b>SPSS</b> ofrece opciones más estándar.
Interacción con otros programas	<b>R</b> brinda la posibilidad de importar datos y/o leer datos desde diferentes formatos porque es compatible con todas las demás interfaces. Así, a través de este se puede trabajar con datos en formatos de <b>SPSS</b> , <b>STATA</b> , <b>SAS</b> , <b>Excel</b> , etc.
Masividad del programadominios	En Chile, las ciencias sociales y el mercado laboral usan predominantemente <b>SPSS</b> para el análisis de datos, pero el programa <b>STATA</b> también fue creció en masividad. En estos mismos programadominios, <b>R</b> ha aumentado fuertemente su popularidad en los últimos años, principalmente dado que es libre y un programa muy rico en términos de herramientas.

Áreas de comparación	Comentarios
Costo	SPSS es un programa pagado. Cuesta 99 dólares en su versión base y se pueden agregar hasta económicos paquetes con funciones adicionales por 79 dólares cada uno. R por su parte, junto con todas sus funciones y librerías, es gratuito, y RStudio también.
Filosofía del diseño	A diferencia de SPSS, R puede incorporar múltiples funcionalidades según necesidad. Esto se hace a partir de la descarga de paquetes que permiten el uso de variadas herramientas, y exportar los análisis en variados formatos hacia diversos programas computacionales (word, excel, etc.). Otra gran diferencia es que SPSS tiene código cerrado. En palabras simples, esto significa que no es posible saber exactamente qué está haciendo el programa –a no ser que se especifique en algún documento– para calcular los análisis estadísticos.

## Interactuando con R y RStudio

### Sintaxis:

En el archivo de código simple `Rscript` hay caracteres especiales para trabajar y se guarda por defecto con extensión `.R`.

```
# : Comentarios que no se ejecutan como comandos
# + : Sigue el comando en la próxima línea
# ; : Para escribir más de una función en la misma línea
```

### Aritmética Básica:

Se puede utilizar para calculadora (este es un simple uso para una plataforma multianálisis).

```
2+3 #Suma
## [1] 5
2-3 #Resta
## [1] -1
2*3 #Multiplicación
## [1] 6
2/3 #División
## [1] 0.6666667
2^3 #Potencia
## [1] 8
```

### Funciones:

Se pueden utilizar funciones predeterminadas en R.

```
log(1000) #Por DEFAULT es logaritmo natural, en base a euler 2,718.
## [1] 6.907755
```

## Documentos y apps

**RStudio** desarrolla herramientas gratuitas y abiertas para R. Su entorno de desarrollo integrado (IDE) facilita el análisis de datos con R.

También ofrece: muchos paquetes R (e.g. tidyverse, sparklyr, ggplot2, dplyr), incluidos **Shiny** (crea aplicaciones web sencillas en R) y **R Markdown** (te permite convertir tus análisis en documentos, informes, presentaciones y paneles de alta calidad; compartirlos y reproducirlos).



## Escribe tu código

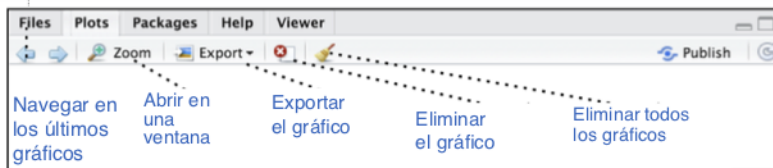
## Soporte R

The screenshot shows the RStudio IDE interface with the following annotations:

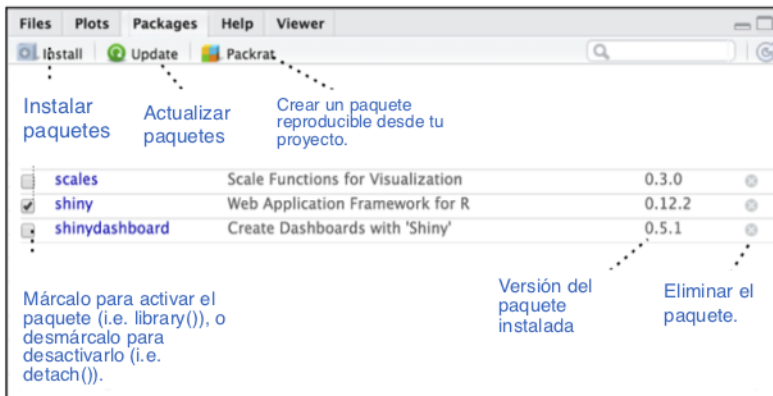
- File menu:**
  - Abrir en una nueva ventana** (Open in a new window)
  - Guardar** (Save)
  - Encontrar y reemplazar** (Find and replace)
  - Compilar como cuaderno** (Compile as notebook)
  - Ejecutar el código seleccionado** (Run selected code)
  - Importar datos con asistente** (Import data with wizard)
  - Historial de comandos** (Command history)
  - Mostrar presentaciones de diapositivas .RPres** (Show .RPres slide presentations)
  - Archivo> Nuevo archivo> Presentación R** (File > New file > R Presentation)
- Editor:**
  - Pestaña de navegación** (Navigation tab)
  - Cursor del usuario** (User cursor)
  - Ejecutar el código previo** (Run previous code)
  - Ejecutar el código completo** (Run full code)
  - Esquema del archivo** (File schema)
  - Selección: Alt + arrastre del mouse.** (Selection: Alt + mouse drag)
  - Código de diagnóstico que aparece en el margen.** (Diagnostic code that appears in the margin. Pass the mouse over the diagnostic symbols to obtain more details.)
  - Resultado de sintaxis.** (Syntax result)
  - Autocompletar mediante tabulación:** nombres de funciones, rutas de archivos, argumentos y más. (Autocomplete by tab: function names, file paths, arguments, etc.)
  - Cambia el tipo de archivo** (Change file type)
  - Saltar a la función en el archivo** (Jump to function in file)
- Environment pane:**
  - Carga el área de trabajo** (Load workspace)
  - Guarda el área de trabajo** (Save workspace)
  - Borra todos los objetos** (Delete all objects)
  - Buscador del área de trabajo** (Workspace search)
  - Elegir el entorno para mostrar** (Choose environment to show)
  - Mostrar objetos como lista o cuadrícula** (Show objects as list or grid)
- Files pane:**
  - Crear una carpeta** (Create a folder)
  - Subir un archivo** (Upload a file)
  - Eliminar un archivo** (Delete a file)
  - Renombrar un archivo** (Rename a file)
  - Ruta del directorio** (Directory path)
  - Directorio de trabajo** (Working directory)
  - Maximiza/minimiza los paneles** (Maximize/minimize panels)
  - Presiona ↑ para ver el historial de comandos** (Press ↑ to see command history)
  - Arrastra los límites** (Drag boundaries)
  - Buscar de archivos. Haga clic en el nombre del archivo o directorio para abrirlo.** (Search for files. Click on the file or directory name to open it.)

Figure 3: Fuente: <https://www.rstudio.com/resources/cheatsheets>

## GRÁFICOS



## PAQUETES



## AYUDA

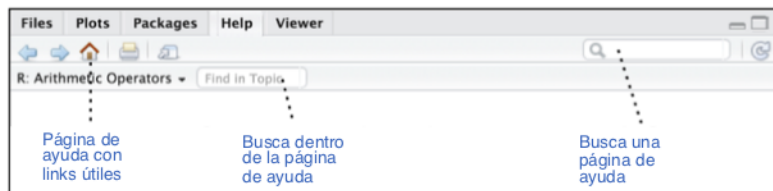


Figure 4: Fuente: <https://www.rstudio.com/resources/cheatsheets>

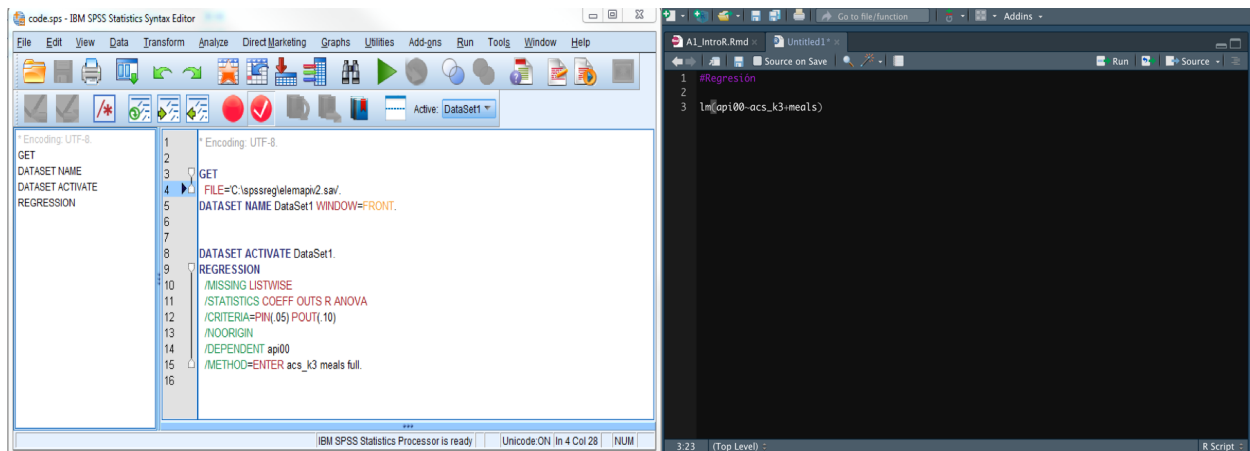


Figure 5: Visualización código (sintaxis) de una regresión lineal en SPSS (izquierda) y en R (derecha)

```
log(2,718) #Cercano a 1 porque euler elevado a 1 = euler

## [1] 0.1053981
log(100, base=10)

## [1] 2
log10(100)

## [1] 2
help("log") #Para saber argumentos
?log

example("log")

##
## log> log(exp(3))
## [1] 3
##
## log> log10(1e7) # = 7
## [1] 7
##
## log> x <- 10~-(1+2*1:9)
##
## log> cbind(x, log(1+x), log1p(x), exp(x)-1, expm1(x))
##
##      x
## [1,] 1e-03 9.995003e-04 9.995003e-04 1.000500e-03 1.000500e-03
## [2,] 1e-05 9.999950e-06 9.999950e-06 1.000005e-05 1.000005e-05
## [3,] 1e-07 1.000000e-07 1.000000e-07 1.000000e-07 1.000000e-07
## [4,] 1e-09 1.000000e-09 1.000000e-09 1.000000e-09 1.000000e-09
## [5,] 1e-11 1.000000e-11 1.000000e-11 1.000000e-11 1.000000e-11
## [6,] 1e-13 9.992007e-14 1.000000e-13 9.992007e-14 1.000000e-13
## [7,] 1e-15 1.110223e-15 1.000000e-15 1.110223e-15 1.000000e-15
## [8,] 1e-17 0.000000e+00 1.000000e-17 0.000000e+00 1.000000e-17
## [9,] 1e-19 0.000000e+00 1.000000e-19 0.000000e+00 1.000000e-19
```



```
#1e-3 = 0,003
#7.2e4 = 72000

args("log")

## function (x, base = exp(1))
## NULL

log(100,10) #= log(100, base=10)

## [1] 2
```

---

## Vectores y variables:

Objeto unidimensional constituido por elementos del mismo tipo.

```
c(1) #Vector de un elemento

## [1] 1

c(1, 2 , 3 , 4) #Crear un vector

## [1] 1 2 3 4

1:4

## [1] 1 2 3 4

4:1

## [1] 4 3 2 1

seq(1, 4)

## [1] 1 2 3 4

seq(4, 1)

## [1] 4 3 2 1

-1:2

## [1] -1 0 1 2

seq(-1, 2)

## [1] -1 0 1 2

seq(2, 8, by=2)

## [1] 2 4 6 8

seq(0, 1, by=0.1)

## [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

seq(0, 1, length=11)

## [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

---

## Operaciones con vectores numéricos

```
#Operaciones aritméticas
```

```
c(1, 2, 3, 4)/2
```

```
## [1] 0.5 1.0 1.5 2.0
```

```
(1:4)/2
```

```
## [1] 0.5 1.0 1.5 2.0
```

```
(1:4)*(4:1)
```

```
## [1] 4 6 6 4
```

```
log(c(0.1, 1, 10, 100), base=10)
```

```
## [1] -1 0 1 2
```

```
c(1, 2, 3, 4) + c(4, 3)
```

```
## [1] 5 5 7 7
```

```
c(1, 2, 3, 4) + c(4, 3, 2) #Produce warning cuando no son múltiplos
```

```
## Warning in c(1, 2, 3, 4) + c(4, 3, 2): longer object length is not a  
## multiple of shorter object length
```

```
## [1] 5 5 5 8
```

```
(1:4)*(1:6) #Warning
```

```
## Warning in (1:4) * (1:6): longer object length is not a multiple of shorter  
## object length
```

```
## [1] 1 4 9 16 5 12
```

```
(1:4)*(1:2)
```

```
## [1] 1 4 3 8
```

---

## Matrices:

Objeto bidimensional constituido por filas y columnas de elementos del mismo tipo.

```
x <- matrix(1:9,3,3)
```

```
x
```

```
##      [,1] [,2] [,3]
```

```
## [1,]    1    4    7
```

```
## [2,]    2    5    8
```

```
## [3,]    3    6    9
```

```
y <- matrix(1:8,2,4,byrow = F) #Genera una matriz con 2 filas y 4 columnas que se irá completando por c  
y
```

```
##      [,1] [,2] [,3] [,4]
```

```
## [1,]    1    3    5    7
```

```
## [2,]    2    4    6    8
```

```
z <- matrix(1:8,2,4,byrow = T) #Genera la matriz completándola por filas
```

```
z
```

```
##      [,1] [,2] [,3] [,4]
```

```
## [1,] 1 2 3 4
## [2,] 5 6 7 8
```

---

### Objetos:

R es un programa orientado a objetos, los que son creados por funciones, que en su forma más general sería: `Objeto <- función` o de manera equivalente `Objeto = función`. Hay diferentes tipos de objetos: vectores, factores, matrices, marco/base de datos (entre otros) Veamoslo en el siguiente código:

```
##Crear objetos
x <- c(1,2,3,4)
z=c(1,2,3,4)

edad <- c(23, 45, 67, 89)
sexo <- c(1, 1, 0, 1)

edad
```

```
## [1] 23 45 67 89
```

```
sexo
```

```
## [1] 1 1 0 1
```

---

### Fijar directorio de trabajo:

Por temas de reproducibilidad, se recomienda establecer directorio de trabajo (ubicación en el pc), donde se buscan y guardan los archivos que se vayan utilizando y/o creando durante el análisis.

```
getwd() # Se obtiene el directorio de trabajo actual
setwd("ruta") # Establecer directorio de trabajo
#Windows
setwd("~/Dropbox/SOL201S_Datos_III/SOL201S_2019/Ayudantia/SOL201S_Datos_3/Ayudantia1") #
#Mac
setwd("~/Dropbox/SOL201S_Datos_III/SOL201S_2019/Ayudantia/SOL201S_Datos_3/Ayudantia1")
```

---

### Librerías o Paquetes:

Conjunto de funciones que tienen una relación entre ellas y que usualmente vienen acompañadas de ficheros de ayuda (documentación). Algunas librerías vienen preinstaladas, otras específicas hay que instalarlas de acuerdo a las necesidades del usuario.

```
library() #Puedo revisar los paquetes instalados
install.packages("libreria") #Para instalar
library("libreria") #Las librerías se instalan sólo una vez, pero deben ser cargadas si se quieren utilizar

install.packages("dplyr") #Para manipulación de datos
install.packages("car") #"Companion to Applied Regression" (Fox & Weisberg)
```

Pueden encontrar información de las librerías oficiales de R disponibles según disciplina académica en el siguiente link.

Además, para evitar lo de instalar/cargar, hay una librería en R que se llama “pacman” (package manager) que cumple la función de cargar las librerías, o en su defecto, instalarlas y cargarlas.

```
install.packages("pacman") #Sólo la primera vez. Este es un paquete que nos permite administrar otros p  
pacman::p_load(dplyr,  
               car) #Cada vez  
#El camino más tradicional:  
library(dplyr)  
library(car)
```

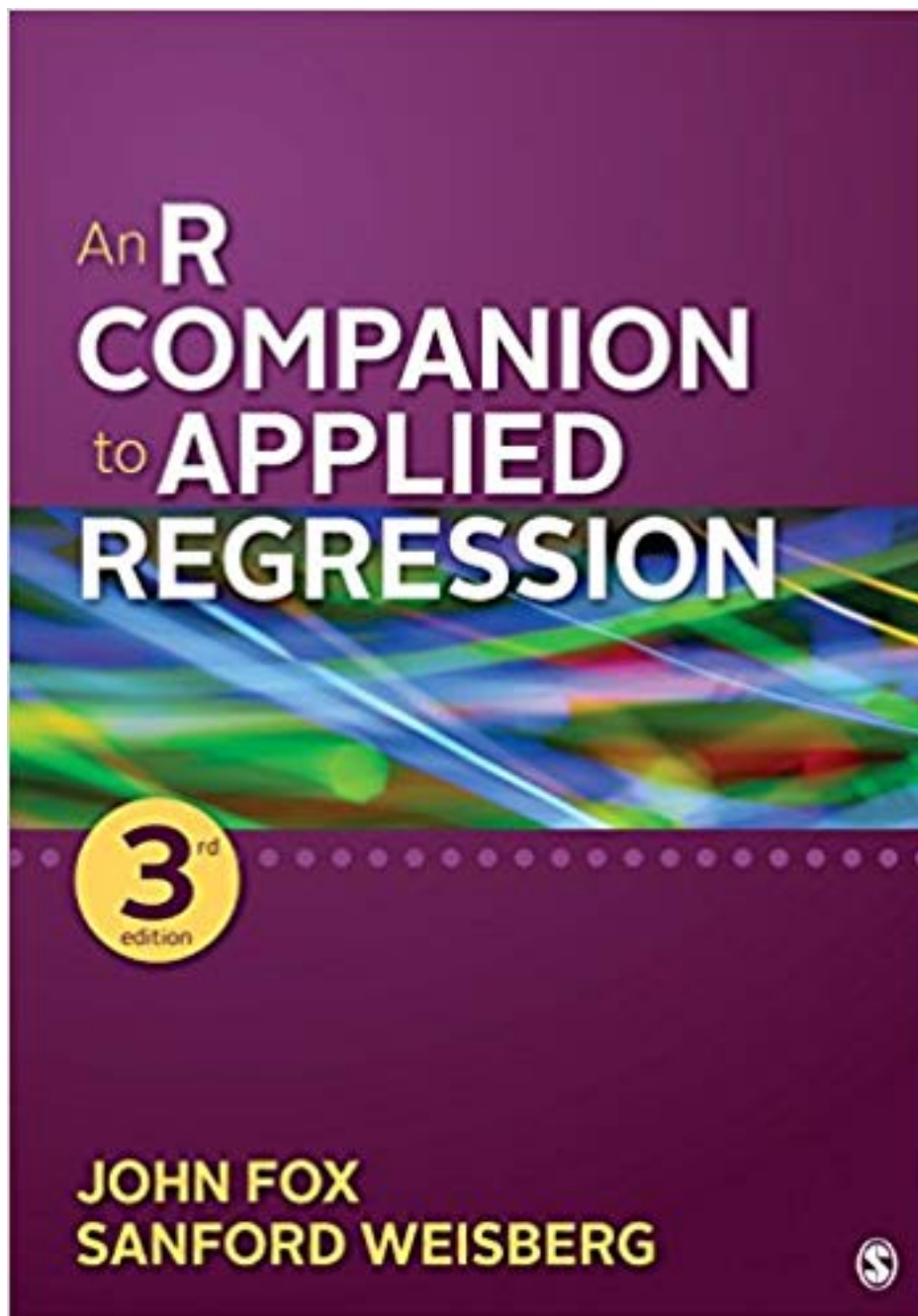
---

## Links de referencia

En R hay múltiples referencias, paquetes, herramientas y formas de hacer las cosas. Para no intimidarse con tal magnitud de información les dejamos algunos links que pueden ser de ayuda:

### Libro base del curso

Fox, John y Sanford Weisberg (2019). An R Companion to Applied Regression. Third Edition. SAGE.



#### Manuales:

Son excelentes herramientas para iniciarse en R, te recomendamos revisarlos con calma y tener mucha paciencia y perseverancia. Es un poco difícil entender la lógica del software en un principio pero solo es cosa de práctica.

- Boccardo & Ruiz (2018). Uso de RStudio para Estadística Univariada en Ciencias Sociales: Manual de apoyo docente para Estadística Descriptiva. Elaborado por sociólogos es una buena fuente para introducirse al mundo de R.
- R para Ciencia de Datos: Libro base para el uso de R, aquí podrán ver que la plataforma de Rstudio no es solo una plataforma de análisis estadístico, sino que de procesamiento de datos y reporte.
- AnalizaR Datos Políticos: Manual con herramientas y tips prácticos para analizar datos políticos.

- RStudio Cheat Sheets: Otros recursos en **inglés**.

### **Foros:**

Para los que alguna vez fuimos nuevos en **R** y **RStudio** nos sirvió bastante ver las preguntas y respuestas de otras personas en todo el mundo. De hecho, casi todas nuestras preguntas ya fueron respondidas por otras personas (una de las ventajas del paradigma de la ciencia abierta) por lo que solo debemos buscar. Aunque, ojo, la mayoría de ellas están en inglés:

- RStudio Community
- Stackoverflow

### **Ayudantías:**

Puedes encontrar las carpetas con las ayudantías y bases de datos en el siguiente Link

La próxima ayudantía veremos:

- Lectura de Bases de datos
- Procesamientos de datos (recodificación)
- Estadística Descriptiva
- Introducción a modelos de regresión lineal