



Taller 1

Introducción al lenguaje de programación R

Aproximación a las Políticas Públicas desde los datos | PUC | 14 de mayo, 2022

 **José Daniel Conejeros**

 jdconejeros@uc.cl

 **Naim Bro**

 naim.bro@imfd.cl

¿Qué es R?

¿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

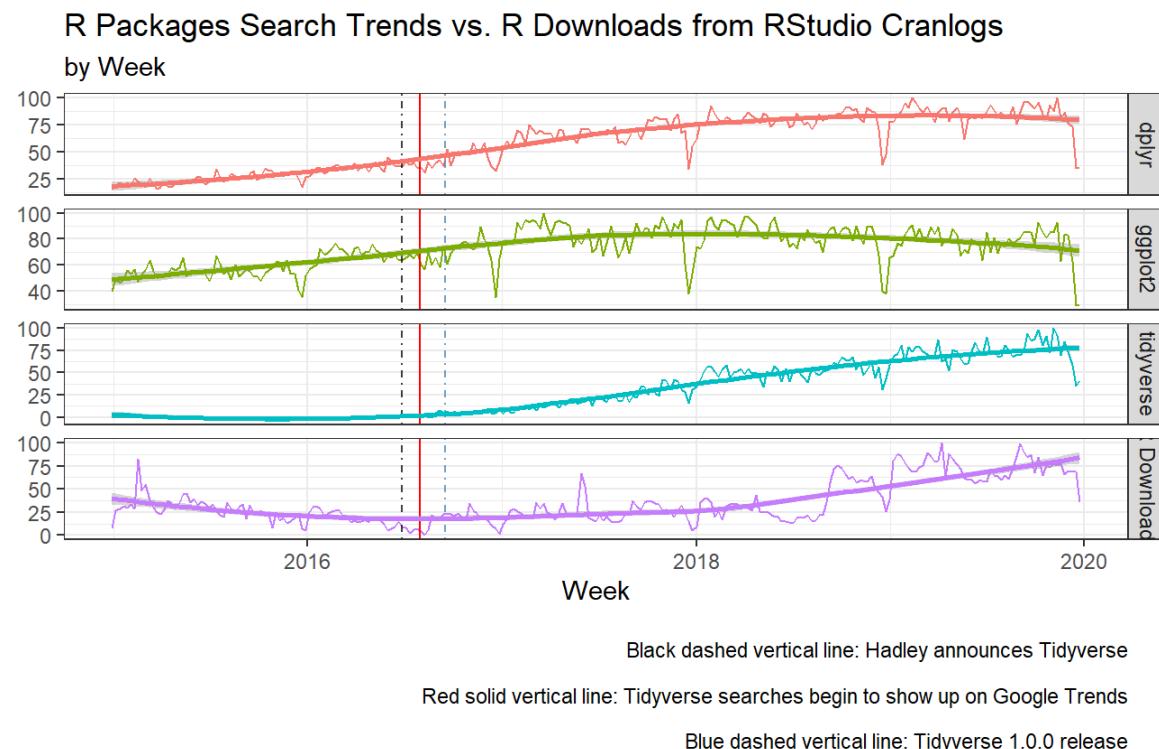
Transparencia

Colaboración

Acceso

¿Qué es R? II

El registro de los análisis queda en texto plano, por lo que es independiente de una plataforma para poder editarlo y permite un control eficiente de versiones (por ejemplo vía Git). Los análisis operan en base a **paquetes o librerías (¡16095 librerías disponibles!)** y su análisis se orienta a **objetos**



Fuente: R's Demise is Highly Overblown. Diciembre 2019

¿Por qué R y Rstudio?

Apunta a distintos **perfiles** (usuarios, desarrolladores), permite el uso de **paquetes**, cuenta con herramientas de **visualización**, tiene un **flujo** completo de trabajo, existe una **comunidad muy activa** y permite el desarrollo de **grupos de interés**.



Instalar R y RStudio

Primero tienes que descargar R desde <https://cran.r-project.org/>. Recuerda seleccionar la versión apropiada a tu computador:

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.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2020-06-06, See Things Now) [R-4.0.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- 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.

El instalador para MAC

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.

Package binaries for R versions older than 3.2.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting (<https://cran-archive.r-project.org>) accordingly.

R 4.0.1 "See Things Now" released on 2020/06/06

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-4.0.1.pkg  
in the Terminal application to print the MD5 checksum for the R-4.0.1.pkg image. On Mac OS X 10.7 and later you can also validate the signature using  
pkgutil --check-signature R-4.0.1.pkg
```

Latest release:

[R-4.0.1.pkg](#) (notarized and signed)
SHA1-hash: c21adbc6e3362cac24969dc9916c3bfffde962b8
(ca. 84MB)

R 4.0.1 binary for macOS 10.13 (High Sierra) and higher, signed and notarized package. Contains R 4.0.0 framework, R.app GUI 1.72 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 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 Xcode 10.1 and GNU Fortran 8.2. If you wish to compile R packages from sources, you will need to download and GNU Fortran 8.2 - see the [tools](#) directory.

[NEWS](#) (for Mac GUI)

News features and changes in the R.app Mac GUI

[Mac-GUI-1.70.tar.gz](#)
MD5-hash: b1ef5f285524640680a22965bb88008

Sources for the R.app GUI 1.70 for Mac OS X. This file is only needed if you want to join the development of the GUI, it is not intended for regular users. Read the INSTALL file for further instructions.

Note: Previous R versions for El Capitan can be found in the [el-capitan/base](#) directory.

R binaries for legacy OS X systems.

"Debes descargar y seguir los pasos de instalación"

Click aquí para descargar e instalar

El instalador para WINDOWS

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 >= 2.15.x, managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
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](#).

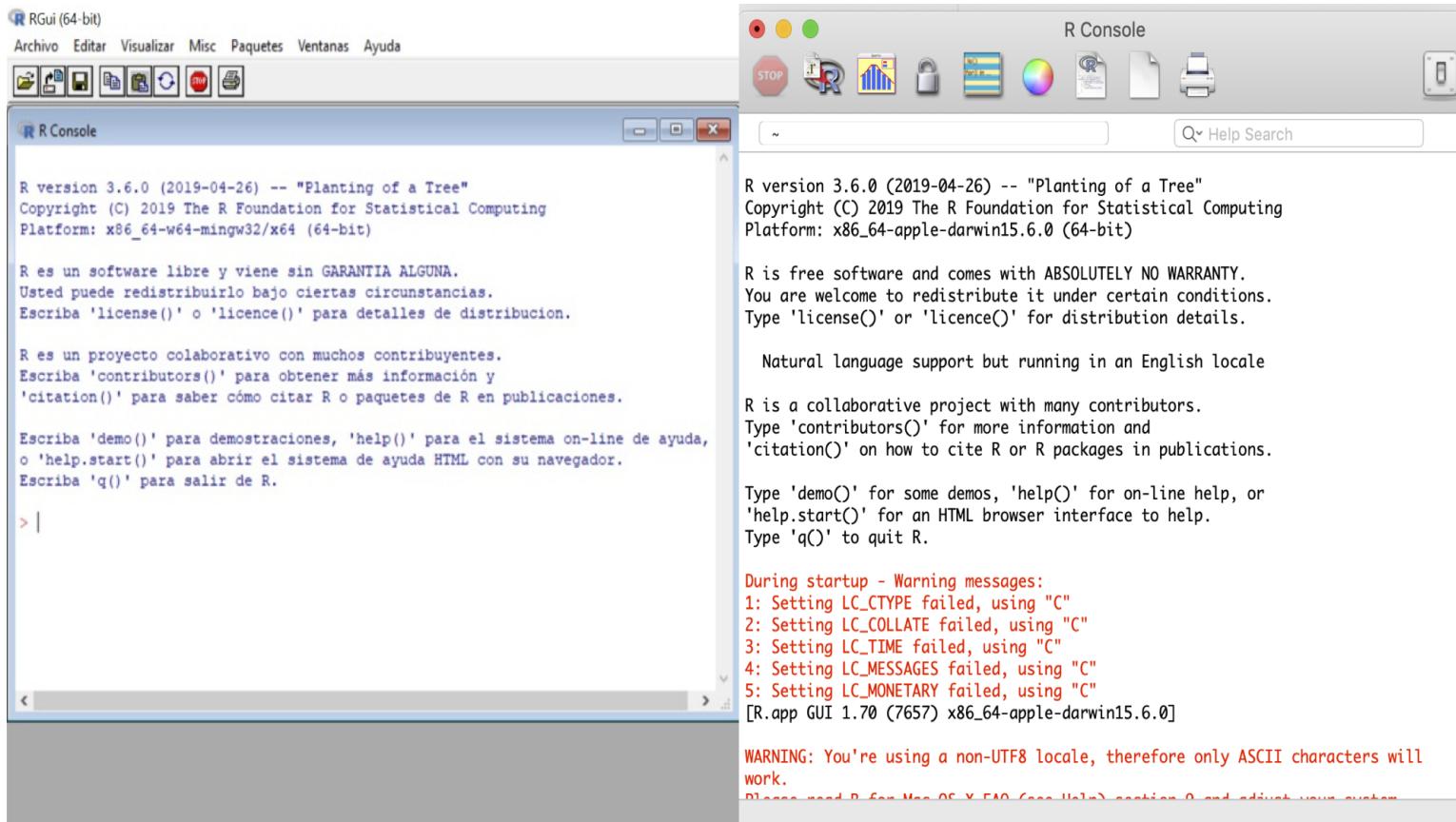
Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

"Debes descargar y seguir los pasos de instalación"

Click aquí para descargar e instalar

Esto es R

Lenguaje y entorno para computación estadística y gráficos:



¿Qué es Rstudio?

Es un entorno de desarrollo integrado (**IDE**) para R que facilita el flujo de trabajo:

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

Abrir en una nueva ventana
Pestaña de navegación
Guardar
Encontrar y reemplazar
Compilar como cuaderno
Ejecutar el código seleccionado

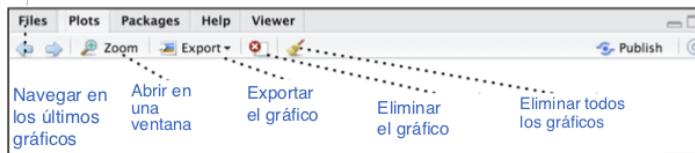
Soporte R

Historial de comandos
Importar datos con asistente
Mostrar presentaciones de diapositivas .Rpres
Archivo > Nuevo archivo > Presentación R

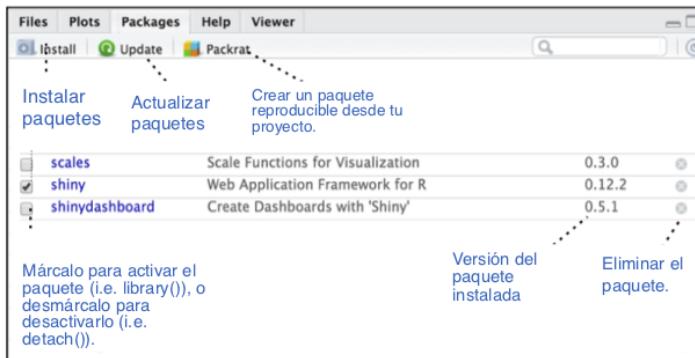
The screenshot displays the RStudio IDE interface. On the left, the code editor shows R code with syntax highlighting and hover documentation. A tooltip for line 12 explains the use of Alt + arrastrar. The code includes diagnostic symbols and a snippet completion example. The middle section shows the R Console with command history and a cursor in the middle of a line. The bottom right shows the file browser with a context menu open over a folder named 'IDEcheatsheet'. The top navigation bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. The status bar at the bottom right shows the date and time (Apr 13, 2016, 11:17 AM).

¿Qué es Rstudio? II

GRÁFICOS



PAQUETES



AYUDA



Descargar e instalar Rstudio

RStudio Desktop 1.3.959 - [Release Notes](#)

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



Requires macOS 10.13+ (64-bit)



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](#).

[Links de Descarga](#)

OS	Download	Size	SHA-256
Windows 10/8/7	 RStudio-1.3.959.exe	171.41 MB	3d493ae5
macOS 10.13+	 RStudio-1.3.959.dmg	148.57 MB	7c5b695d

Click aquí para instalar

Trabajar en la nube: Rstudio Cloud

Si no queremos/podemos instalar R en nuestros equipos, tenemos la opción de usar su versión en la nube. Para empezar a trabajar con RStudioCloud necesitamos [crear una cuenta aquí](#).

The screenshot shows the RStudio Cloud interface with the following sections:

- Left Sidebar:** Includes links for "Your Workspace", "New Space", "Learn" (with "Guide", "What's New", "Primers", "Cheat Sheets"), "Feedback and Questions", "Info" (with "Plans & Pricing", "Terms and Conditions", "System Status"), and a "Cloud" section.
- Header:** Shows "Your Workspace / Mi Proyecto".
- Top Bar:** Includes "File", "Edit", "Code", "View", "Plots", "Session", "Build", "Debug", "Profile", "Tools", "Help", "Console", "Terminal" (active), and "Jobs".
- Console:** Displays R version 4.0.0 output:

```
R version 4.0.0 (2020-04-24) -- "Arbor Day"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```
- Environment:** Shows an empty environment.
- Files:** A file browser showing a "Cloud" folder containing ".Rhistory" (0 B, Jun 11, 2020, 10:58 PM) and "project.Rproj" (205 B, Jun 11, 2020, 10:58 PM).

¿Qué son los objetos?

Estructuras de datos que contienen información:

```
objeto ← c("info1", "info2", "info3")
objeto
```

```
[1] "info1" "info2" "info3"
```

```
ventas_chocolate ← c(254,203,182,50)
ventas_chocolate
```

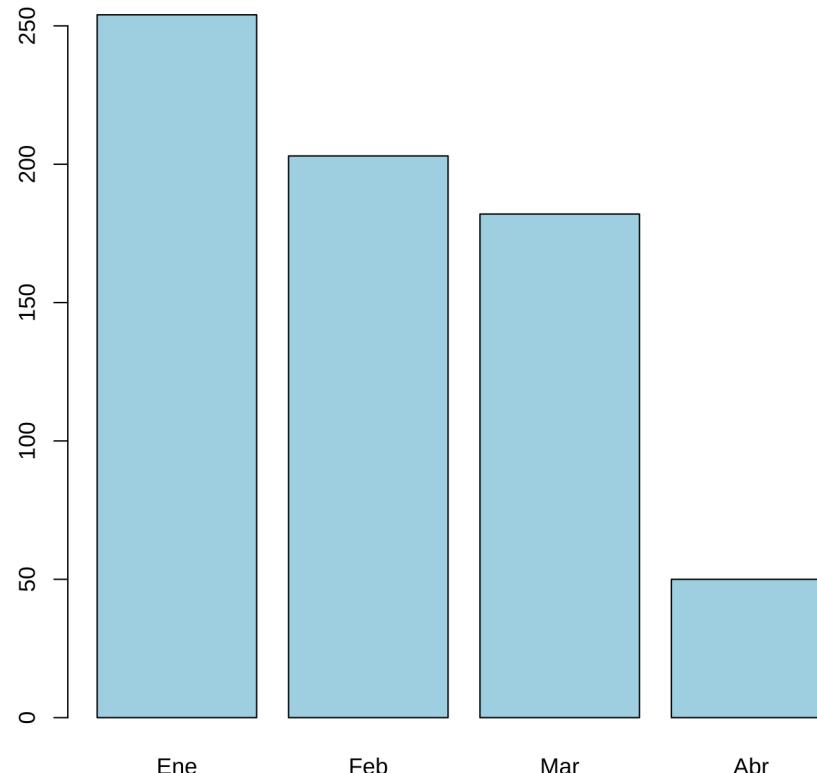
```
[1] 254 203 182 50
```

```
meses ← c("Ene", "Feb", "Mar", "Abr")
meses
```

```
[1] "Ene" "Feb" "Mar" "Abr"
```

¿Qué son los objetos?

```
barplot ← barplot(ventas_chocolate, names.arg = meses, col="lightblue")
```



Funciones y argumentos

Ejecutan una acción sobre nuestros datos. Algunas requieren **inputs** (argumentos) que van dentro del paréntesis.

```
?sum  
sum( ... , na.rm = FALSE)  
  
Arguments  
... numeric or complex or logical vectors.  
na.rm logical. Should missing values (including NaN) be removed?
```

```
ventas_chocolate
```

```
[1] 254 203 182 50
```

```
sum(ventas_chocolate)
```

```
[1] 689
```

Funciones y argumentos

En lo relacionado con el análisis de datos, la lógica de R es bastante amigable;

funcion(objeto, ...)

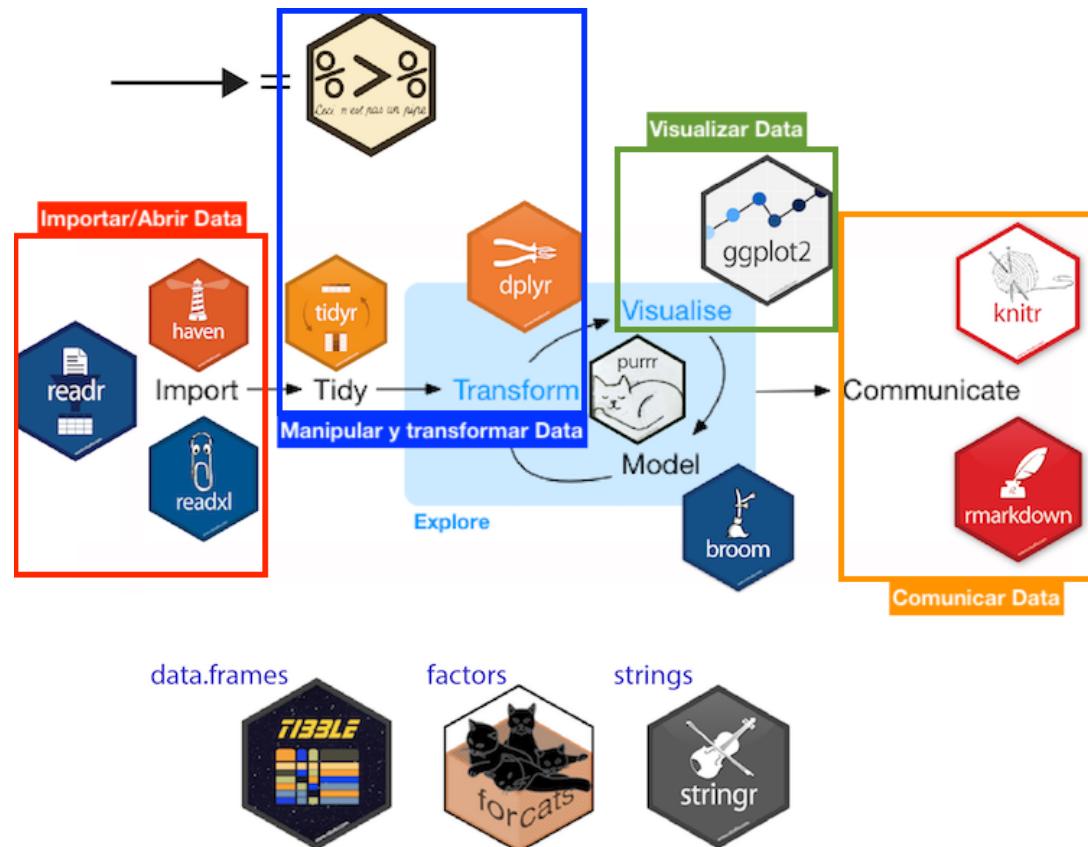


Acción

Estructura de datos
(recibe la acción)

Los ... corresponden a los argumentos que modifican la acción a realizar, dependiendo de la función esta puede o no tener argumentos y estos pueden ser tanto obligatorios como opcionales.

Uso de librerías o Paquetes



Puedes revisar las principales librerías dando [click aquí](#)
Puedes revisar todas las librerías de tidyverse dando [click aquí](#)

¿Cómo instalar librerías o paquetes?

Imagina que vas a un supermercado a comprar una ampolleta:

Podemos instalar librerías desde **CRAN** (Comprehensive R Archive Network). Repositorio oficial de paquetes de R (debe ser aprobados).

```
install.packages("nombrepaket")
```

También podemos instalar librerías desde **GitHub** que es una plataforma de desarrollo de software.

```
install.packages("remotes")
remotes::install_github("usuario/paquete")
```

Cada vez que necesites la ampolleta debes encenderla:

Una vez instaladas nuestra librería en **R** deben ser "llamadas" a la sesión de trabajo:

```
library(paquete)
```

Recuerda que debes cargar tus librerías para cada sesión de trabajo

Ejemplo

Generar matrices de datos

```
id ← c(1, 2, 2, 4, 4)
edad ← c(23, 45, 67, 201)
sexo ← c(1, 0, 1, 0, NA)
peso ← c(80, 60, 70, 50, 55)
altura ← c(180, 160, 200, 140, 300)

data ← as.data.frame(cbind(id, edad, sexo,
                           peso, altura))
data ← as.data.frame(data)
data
```

	id	edad	sexo	peso	altura
1	1	23	1	80	180
2	2	45	0	60	160
3	2	67	1	70	200
4	4	201	0	50	140
5	4	23	NA	55	300

Explorar y guardar datos

```
dim(data)
```

```
[1] 5 5
```

```
colnames(data)
```

```
[1] "id"      "edad"     "sexo"     "peso"     "altura"
```

```
head(data)
```

	id	edad	sexo	peso	altura
1	1	23	1	80	180
2	2	45	0	60	160
3	2	67	1	70	200
4	4	201	0	50	140
5	4	23	NA	55	300

```
write.csv(data, "data.csv")
```

Validación, Manipulación y transformación

```
library(dplyr)
table(duplicated(data$id))
```

```
FALSE  TRUE
 3      2
```

```
data[duplicated(data$id), ]
```

	id	edad	sexo	peso	altura
3	2	67	1	70	200
5	4	23	NA	55	300

```
data ← data %>%
  select(id, edad, sexo, altura, peso) %>
  arrange(edad) %>%
  filter(edad!=201) %>%
  mutate(pes_alt = peso/altura,
        id=c(1:4))
```

Estructuras de control, loops e iteraciones

```
# Loops
for (i in 1:4){
  print(paste0("Su IMC es: ", data$pes_alt[i]))
}
```

```
[1] "Su IMC es: 0.4444444444444444"
[1] "Su IMC es: 0.1833333333333333"
[1] "Su IMC es: 0.375"
[1] "Su IMC es: 0.35"
```

```
# IF - ESLE-IF - ELSE
a ← data$pes_alt[2]
if (a > 0.40) {
  print("sobre peso")
} else {
  print("bajo peso")
}
```

```
[1] "bajo peso"
```

```
a
```

Descriptivos y funciones

```
summary_replica <- function(x) {  
  tabla <- round(c(Mean = mean(x, na.rm =  
    Sd = sd(x, na.rm = T),  
    Min = min(x, na.rm = T  
    Median = as.numeric(me  
    Max = max(x, na.rm = T  
    n =length(x)), 2)  
  return(tabla)  
}  
  
summary_replica(data$peso)
```

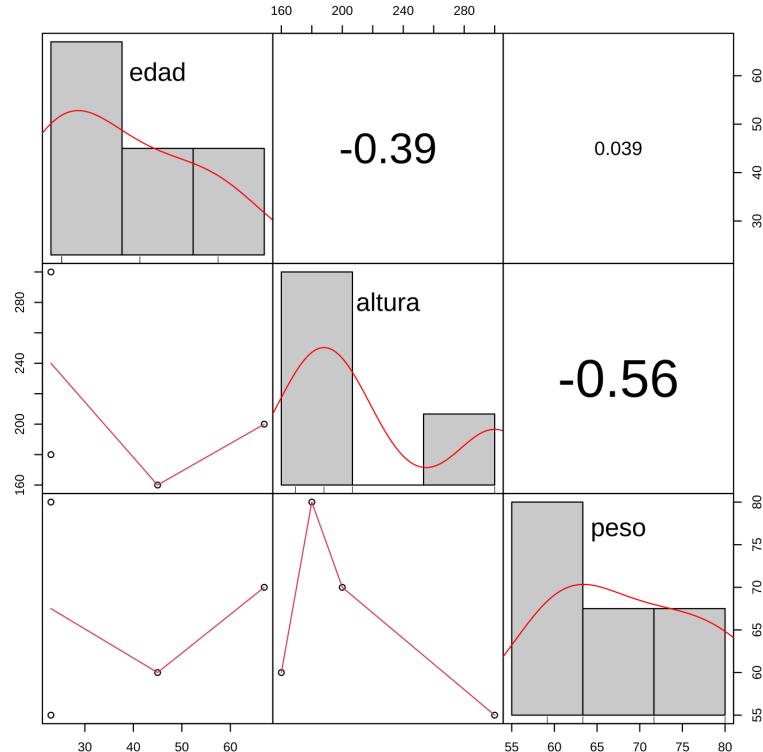
Mean	Sd	Min	Median	Max	n
66.25	11.09	55.00	65.00	80.00	4.00

```
summary_replica(data$altura)
```

Mean	Sd	Min	Median	Max	n
210.00	62.18	160.00	190.00	300.00	4.00

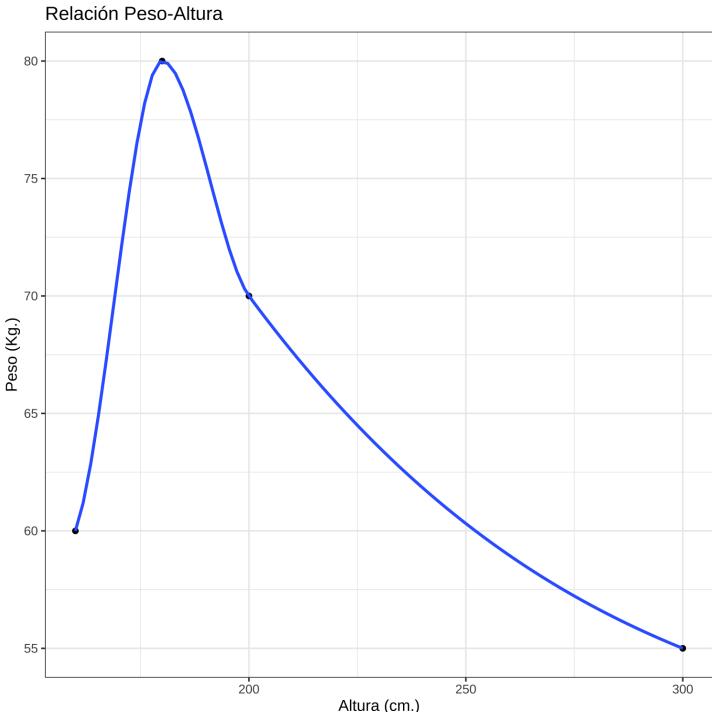
Asociaciones

```
library(PerformanceAnalytics)  
chart.Correlation(data[, c(2,4,5)],  
  histogram = T, pch=19)
```



Visualización

```
library(ggplot2)
data %>% ggplot(aes(x=altura,y=peso)) +
  geom_point() + geom_smooth() +
  labs(title="Relación Peso-Altura",
       x="Altura (cm.)", y = "Peso (Kg.)"
  theme_bw()
```



Reporte

```
Session7_Reporte_ref.Rmd x
1 ...
2 title: "Taller R - Sesión 7: Reportería con Rmarkdown"
3 author: "José Daniel Conejeros"
4 date: "Julio 2020"
5 output: pdf_document
6 header-includes:
7   - \usepackage[spanish,es-tabla]{babel}
8   - \usepackage[utf8]{inputenc}
9   urlcolor: blue
10  fontsize: 10pt
11 ...
12
13 El **objetivo** de esta sesión es introducir a la reportería de análisis estadísticos con Rmarkdown.
14
15 ## I. ¿Qué es R Markdown?
16
17 Es un espacio para elaborar documentos en PDF, HTML, MS-WORD (entre muchas otras) con una sintaxis de formato simple.
resultados y texto con el fin de crear documentos para que puedan ser reproducibles.
18 \vspace{0.3 cm}
```

Taller R - Sesión 7: Reportería con Rmarkdown

José Daniel Conejeros

Julio 2020

El **objetivo** de esta sesión es introducir a la reportería de análisis estadísticos con Rmarkdown.

I. ¿Qué es R Markdown?

Es un espacio para elaborar documentos en PDF, HTML, MS-WORD (entre muchas otras) con una sintaxis de formato simple. Dado que R es un entorno de Desarrollo Integrado, R Markdown proporciona un marco unificado para la ciencia de datos combinando código, resultados y texto con el fin de crear documentos para que puedan ser reproducibles.

Recomendaciones

1. Utilizar un nombre que tenga alguna relación con los datos que contiene el objeto.
2. Evitar caracteres especiales, como ñ, tildes o espacios.
3. Para separar palabras se puede utilizar un guión bajo (proyeccion_enero) o mayúscula inicial (ProyeccionEnero). Lo importante es ser consistente en la opción elegida.

Errores Comunes

La mayoría de los errores que cometemos son por problemas de tipeo:

- Escribimos mal el nombre de una función u objeto. Nos falta cerrar un paréntesis.
- Nos falta una coma.

En caso de que falte un paréntesis o una coma, el editor de RStudio nos lo advertirá.

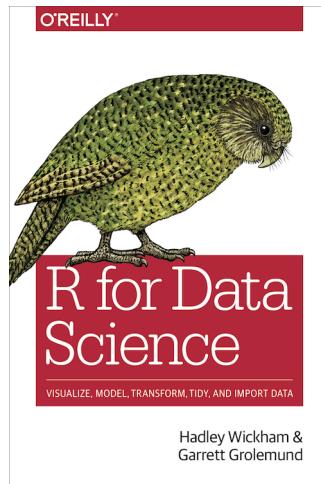
A veces se generan problemas porque olvidamos correr una línea de código o porque sobreescrivimos un objeto. En esos casos, lo mejor es reiniciar R y volver a ejecutar el código desde el principio.

¡Recuerda que equivocarse es normal!

Buscadores como Google ayudan mucho a resolver errores, saber como buscar soluciones a tus dificultades de manera eficiente.

Referencias

R para Ciencia de Datos



Versión en español (aún en desarrollo).

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

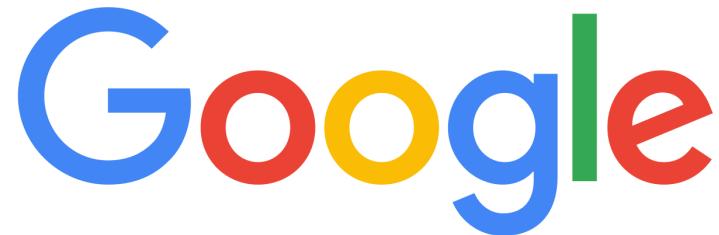


Repositorio con información.

Otros manuales



Para resolver dudas





Taller 1

Introducción al lenguaje de programación R

 **José Daniel Conejeros**

 jdconejeros@uc.cl |  JDConejeros |  Material del taller

 **Naim Bro**

 naim.bro@imfd.cl |  naimbro |  Programa del curso

Tema adaptado de Esteban Rucán Carrasco:  errucan@uc.cl