

# Programa de Doctorado

## Introducción a R



## Manejo y visualización de datos en R

Paloma Ruiz Benito, Verónica Cruz Alonso,  
Julen Astigarraga Urcelay

Enero - 2022



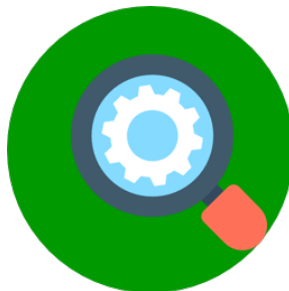
# Definiciones o conceptos clave



¡¡Recuerda!!



Piensa críticamente para  
responder a una pregunta



Ejemplos

# Contenido

## Parte I.- Introducción a R

- Entornos de programación, introducción a R y RStudio.
- Tipos de variables y datos, operaciones aritméticas y lógicas, creación de vectores, matrices, listas y tablas. Selección de datos.

**Día 1**

## Parte II.- Manejo de datos

- Flujo y funciones en la gestión de bases de datos.
- Recomendaciones para la estructura de las bases de datos y creación de código de programación.
- Introducción a la gestión de datos. Recomendaciones para su generación.
- Estructuras de programación: condicionales, bucles y funciones.

## Parte III.- Visualización de datos

- Funciones básicas e introducción a ggplot para la visualización de datos.
- Generación de gráficos unidimensionales: histograma, dispersión, gráfico de cajas y bigotes, etc.
- Generación de gráficos bidimensionales: dispersión, boxplot, gráficos de barras, etc.
- Ejemplos y prácticas de visualización de gráficos en mapas.

## Parte IV.- Trabajo reproducible

- Trabajo reproducible.
- Introducción a git y github.
- Rmarkdown.



# ¿Qué es R?

***R es un lenguaje de programación para análisis estadístico***

- Código libre (gratuito)
- Estado del arte
- Flexible y potente
- Portátil y replicable en otras plataformas
- Ejemplos para “casi” cualquier análisis
- Por y para la comunidad



# ¿Qué es R?

*R es un lenguaje de programación para análisis estadístico*

- Hay una cantidad masiva de paquetes para análisis estadísticos, visualización y manipulación de datos.



1.- **R** es **gratuito** y **flexible**

2.- **R** es de código abierto

3.- **R** es colaborativo

1.- La **ciencia** debe ser **gratuita**

2.- La **ciencia** debe ser de **código abierto**

3.- La **ciencia** debe ser **colaborativa**

4.- La **ciencia** debe ser **replicable**





Parte de R-Ladies - 132 grupos

## R-Ladies Madrid

Madrid, España

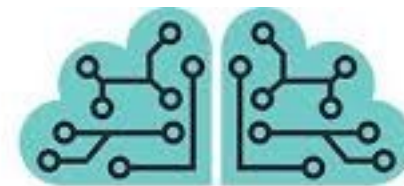
1.101 miembros · Grupo público

Organizado por R-Ladies G. y otras 3 personas

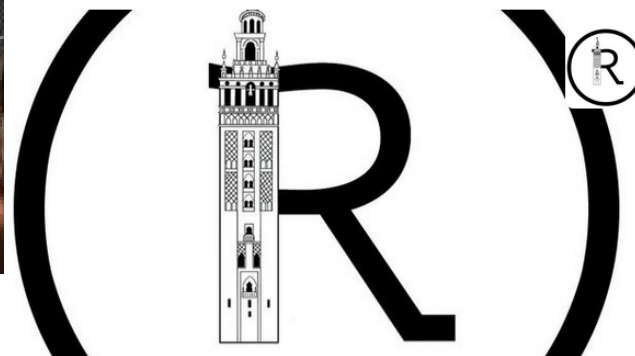
Compartir: [f](#) [t](#) [in](#)

[Sobre nosotros](#) [Eventos](#) [Miembros](#) [Fotos](#) [Conversaciones](#)

Unirse a este grupo



# DATA TONS



Parte de R User Groups - 66 grupos

## Sevilla R users

Sevilla, España

438 miembros · Grupo público

Organizado por Francisco Rodriguez S. y otras 4 personas

Compartir: [f](#) [t](#) [in](#)

[Sobre nosotros](#) [Eventos](#) [Miembros](#) [Fotos](#) [Conversaciones](#)

Unirse a este grupo

# ¿Qué es R?

*R es un lenguaje de programación para análisis estadístico*

¿Quién está de acuerdo con esta afirmación?





***R se considera un lenguaje de programación en si mismo **que va más allá** del análisis estadístico***

```
> install.packages("lattice")  
> library(lattice)  
> demo (lattice)
```





# ¿Por qué Rstudio?

*R studio es un **IDE** (integrated development environment) gratuito y en código abierto*

## **IDE construido exclusivo para R**

- El resaltado de sintaxis, auto completado de código y sangría inteligente
- Ejecutar código R directamente desde el editor de código fuente
- Salto rápido a las funciones definidas



# ¿Por qué Rstudio?

## Colaboración

- Documentación y soporte integrado
- Administración sencilla de múltiples directorios de trabajo mediante proyectos
- Navegación en espacios de trabajo y visor de datos



# ¿Por qué Rstudio?

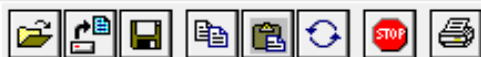
## **Potente autoría y depuración**

- Depurador interactivo para diagnosticar y corregir los errores rápidamente
- Herramientas de desarrollo extensas
- R Markdown y conexión a Git

# El entorno R

RGui (64-bit)

Archivo Editar Visualizar Misc Paquetes Ventanas Ayuda



R Console

```
R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"  
Copyright (C) 2015 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)
```

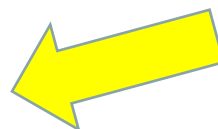
```
R es un software libre y viene sin GARANTIA ALGUNA.  
Usted puede redistribuirlo bajo ciertas circunstancias.  
Escriba 'license()' o 'licence()' para detalles de distribucion.
```

```
R es un proyecto colaborativo con muchos contribuyentes.  
Escriba 'contributors()' para obtener más información y  
'citation()' para saber cómo citar R o paquetes de R en publicaciones.
```

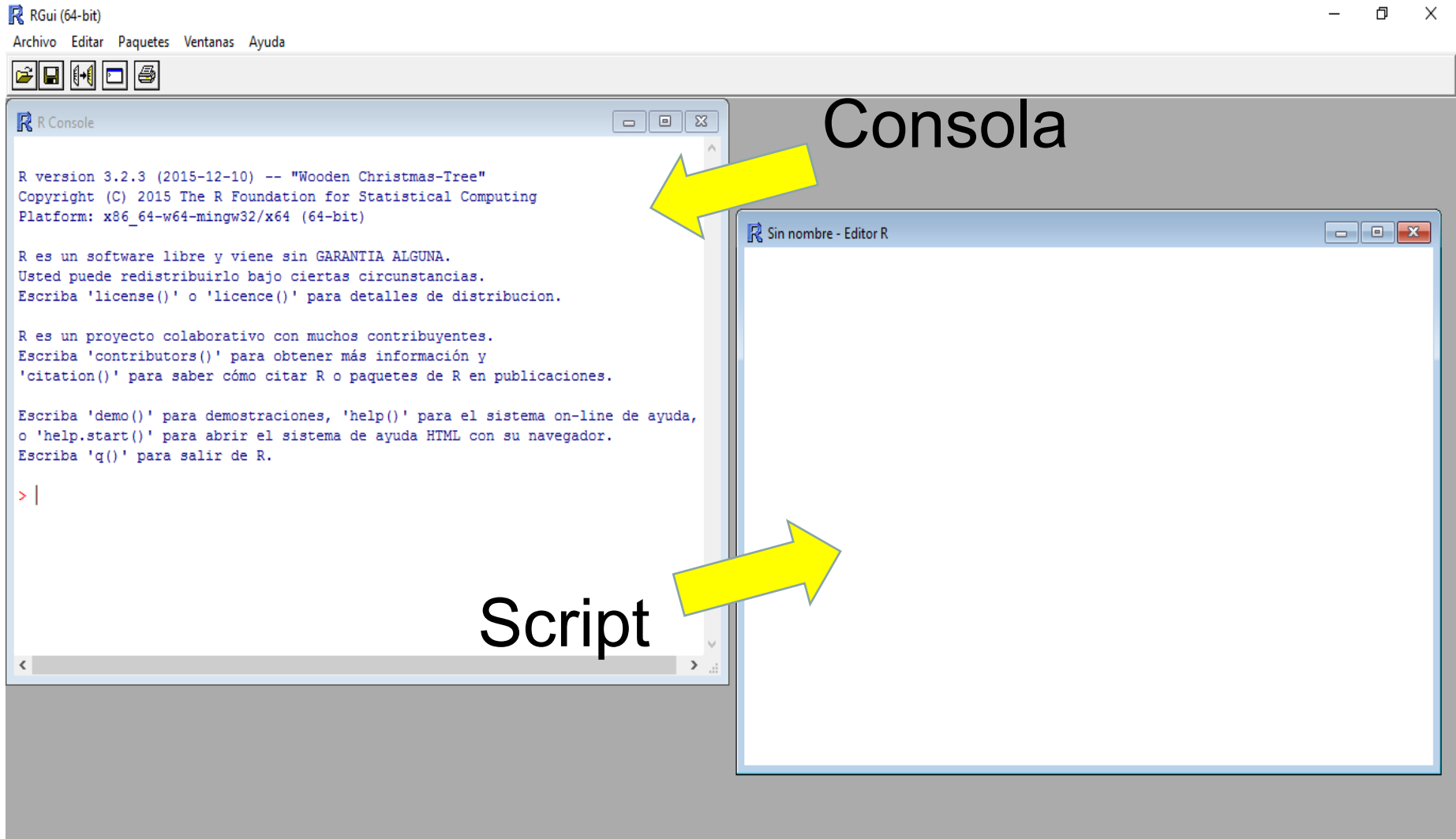
```
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,  
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.  
Escriba 'q()' para salir de R.
```

```
> |
```

Consola



# El entorno R



# Rstudio

The screenshot shows the RStudio interface with several components highlighted by red boxes and numbered annotations:

- 1** Scripts abiertos. (Scripts open)
- 2** Ventana de código (script) (Code window)
- 3** Visor de objetos en la memoria (Object viewer)
- 4** Consola (Console)
- 5** Panel de información (Information panel)

The code in the script window is as follows:

```
1  
2  
3 #-----  
4 library(caret)  
5 library(raster)  
6  
7 carpeta<-"E:\\Logistico\\"  
8  
9  
10 #defino las variables y c  
11 modelos<-paste(carpeta,"m  
12 predicciones<-paste(carpe  
13  
14 rasters<-paste(carpeta,"M  
15 rasters2<-paste(carpeta,"I  
16  
17 dir.create(file.path(mode  
18 dir.create(file.path(pred  
19  
20  
21 #cargó los rasters  
22 setwd(rasters)  
23 l<-list.files(pattern = g  
24 r <- stack(l)  
25  
26 setwd(rasters2)  
27  
28
```

The console output is as follows:

```
15:42 [Untitled] R Script  
Loading required package: ggplot2  
> library(raster)  
Loading required package: sp  
>  
> carpeta<-"E:\\Logistico\\"  
>  
> #defino las variables y creo las carpetas que se necesitan  
> modelos<-paste(carpeta,"modelos\\",sep="")  
> predicciones<-paste(carpeta, "predicciones\\",sep="")  
>  
> rasters<-paste(carpeta,"Month\\", sep="")  
> rasters2<-paste(carpeta,"Month2\\", sep="")  
>  
> dir.create(file.path(modelos), showWarnings = FALSE)  
> dir.create(file.path(predicciones), showWarnings = FALSE)  
>
```

The Environment pane shows the following values:

Object	Value
carpeta	"E:\\Logistico\\"
modelos	"E:\\Logistico\\modelos\\"
predicciones	"E:\\Logistico\\predicciones\\"
rasters	"E:\\Logistico\\Month\\"
rasters2	"E:\\Logistico\\Month2\\"

The Packages pane shows the following user library:

Name	Description	Version
abind	Combine Multidimensional Arrays	1.4-5
acepack	ACE and AIVAS for Selecting Multiple Regression Transformations	1.4.1
aplpack	Another Plot PACKAGE: stem, leaf, bagplot, faces, spin3R, plotsummary, plotnulls, and some slider functions	1.3.0
archivist	Tools for Storing, Restoring and Searching for R Objects	2.3.1
arm	Data Analysis Using Regression and Multilevel/Hierarchical Models	1.10-1
assertthat	Easy Pre and Post Assertions	0.2.0
automap	Automatic interpolation package	1.0-14
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.2
base64enc	Tools for base64 encoding	0.1-3
BBmisc	Miscellaneous Helper Functions for B. Bischl	1.11
BH	Boost C++ Header Files	1.66.0-1
BiasedUm	Biased Um Model Distributions	1.07
bibtex	Bibtex Parser	0.4.2
bindr	Parametrized Active Bindings	0.1.1
bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2
Bioconductor	Install/Update Bioconductor, CRAN, and github Packages	1.24.0
bit	A Class for Vectors of 1-Bit Booleans	1.1-14
bit64	A S3 Class for Vectors of 64bit Integers	0.9-7
bitops	Bitwise Operations	1.0-6

# Rstudio

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1\* Encinas\_size\_BAI\_Resilience.R Gbif.R JessicaTFMscript.R

Source on Save

Run Source

```
46 library(plyr)
47 revalue(size2$Ecotype, c("GR"="Árida", "HR"="Local"))
48 levels(size2$Ecotype) <- c("Árida", "Local")
49 histogram(~ size2$Dmm | size2$Ecotype, col="lightblue", ylab="Porcentaje", xlab="Diámetro(mm)")
50 hist
51
52
53 ##### Creating data set with BAI values #####
54 # Calculating BAI
55 growth <- read.csv("Encinas_growthT.csv") # Read datafile with ring width data (trees in columns and years
56 BAI_encinas <- bai.in(growth) # calculate BAI increment from pith to bark (square of input unit, i.e. mm2)
57 BAI_encinas <- BAI_encinas/100 # convert to cm2
58 row.names(BAI_encinas) <- c(2008:2015) # Give names to rows
59 drops <- "ID"
60 BAI_encinas <- BAI_encinas[, !(names(BAI_encinas) %in% drops)] # Prepare data to export (eliminate ID row)
61 write.table(BAI_encinas, file="BAI_encinasT.csv", sep=";") # Export data to an excel file
62 BAI <- unlist(BAI_encinas[1:29], use.names = FALSE) # Create a vector by joining columns consecutively
63 # "Detrending" data (These data have not been used in models)
64 <
```

50:1 Size in February 2016

R Script

Console C:/Users/Enrique/Desktop/Data analysis/Encinas/

```
> t.test(Dmm~Ecotype, data=size[size$Dmm<=85,])
```

Welch Two Sample t-test

data: Dmm by Ecotype

t = 2.5095, df = 60.729, p-value = 0.01477

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

1.857298 16.430702

sample estimates:

mean in group GR mean in group HR

47.47971 38.33571

```
> par(mfrow=c(1,2))
```

```
> boxplot(Hcm~Ecotype, data=size, ylab="Height (cm)", xlab="Ecotype", col="light green")
```

```
> boxplot(Dmm~Ecotype, data=size, ylab="Diameter (mm)", xlab="Ecotype", col="light green")
```

Environment History

Import Dataset

Global Environment

Data

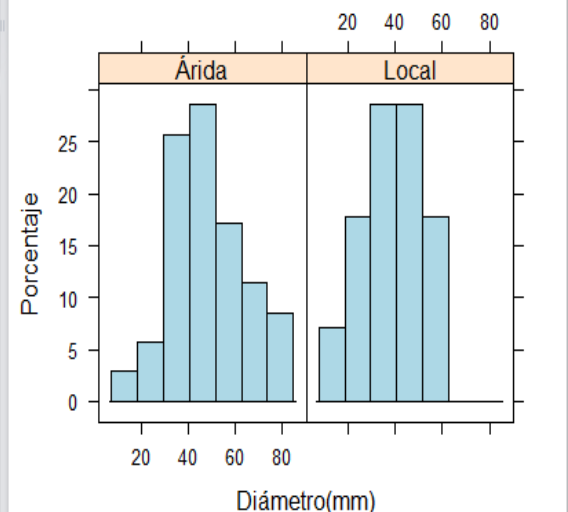
clima	180 obs. of 13 variables
growth	8 obs. of 30 variables
size	70 obs. of 7 variables
size2	69 obs. of 7 variables

Values

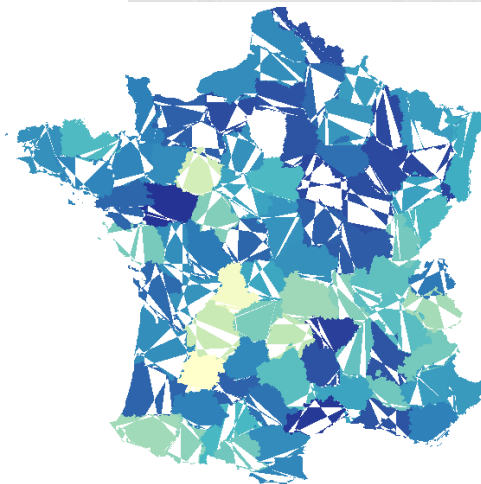
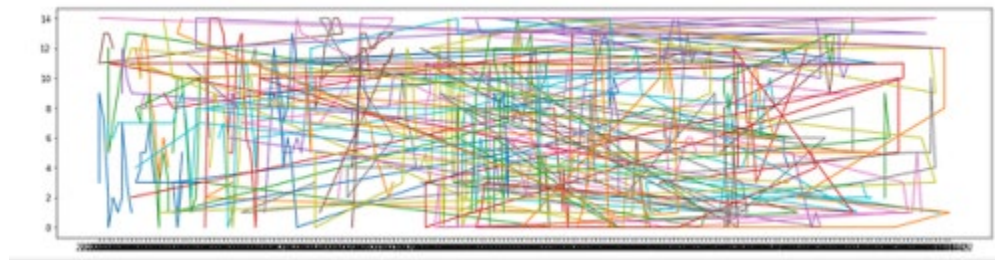
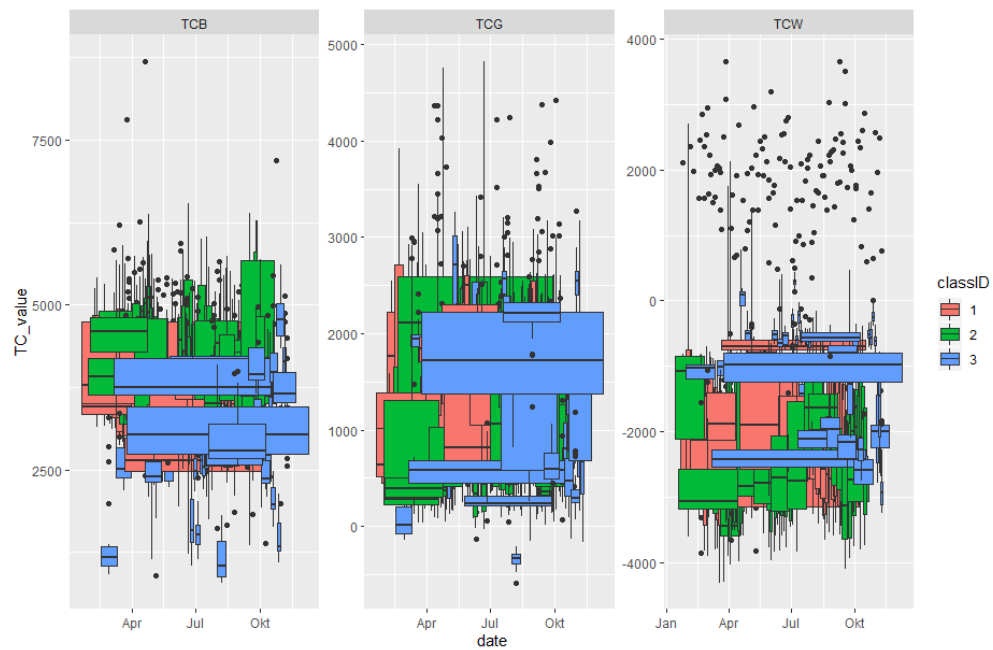
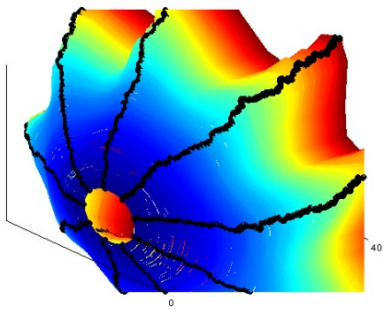
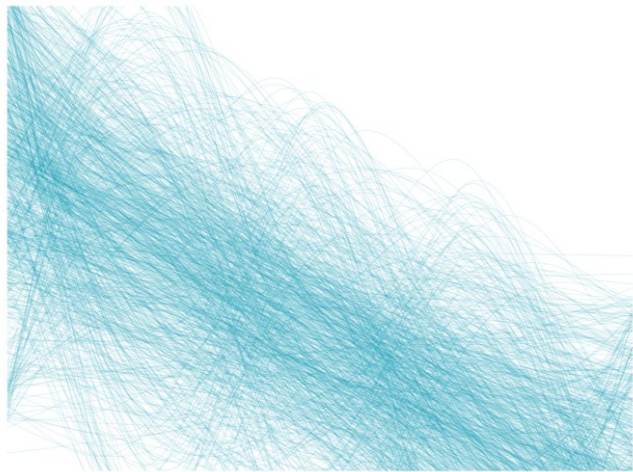
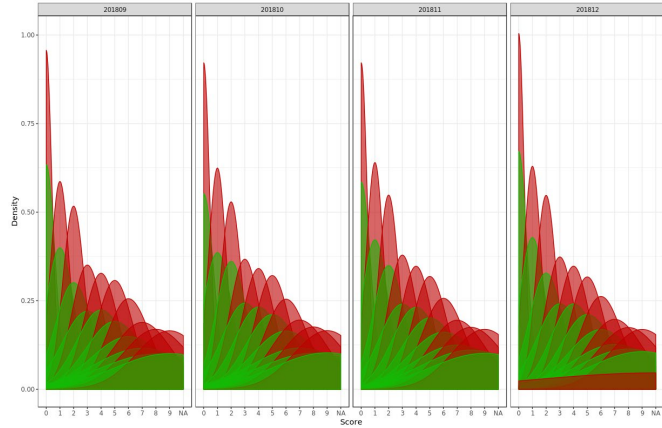
a	num [1:2(1d)]	16.6 12.3
b	int [1:2(1d)]	35 28

Files Plots Packages Help Viewer

Zoom Export Publish

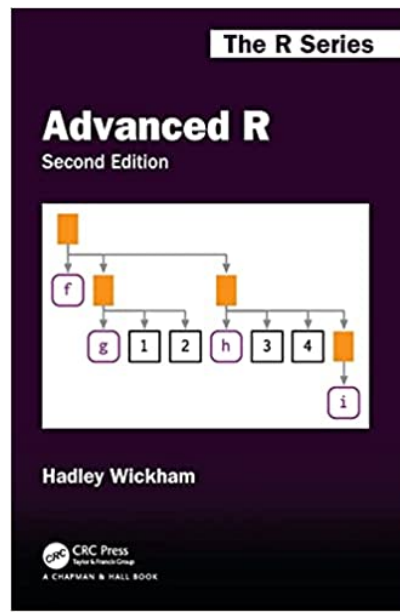
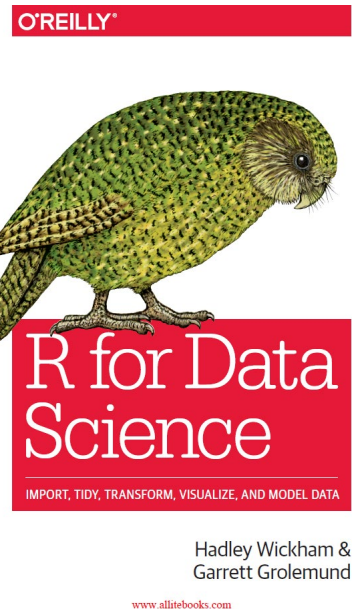
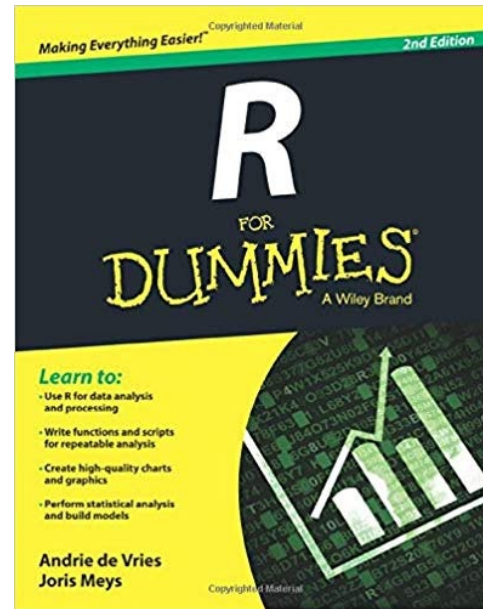
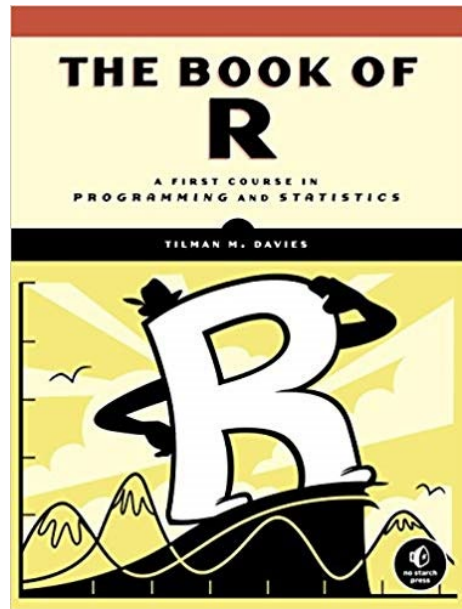
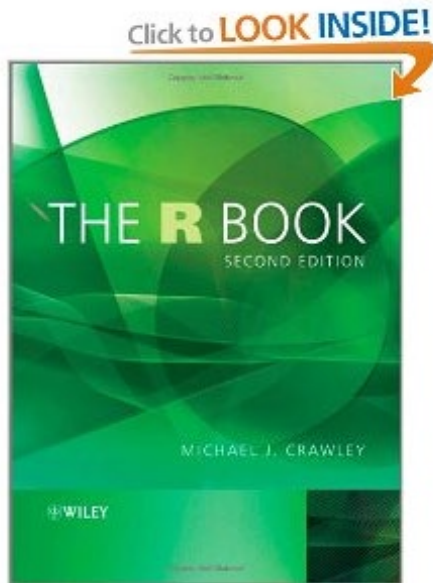






<http://accidental-art.tumblr.com/>

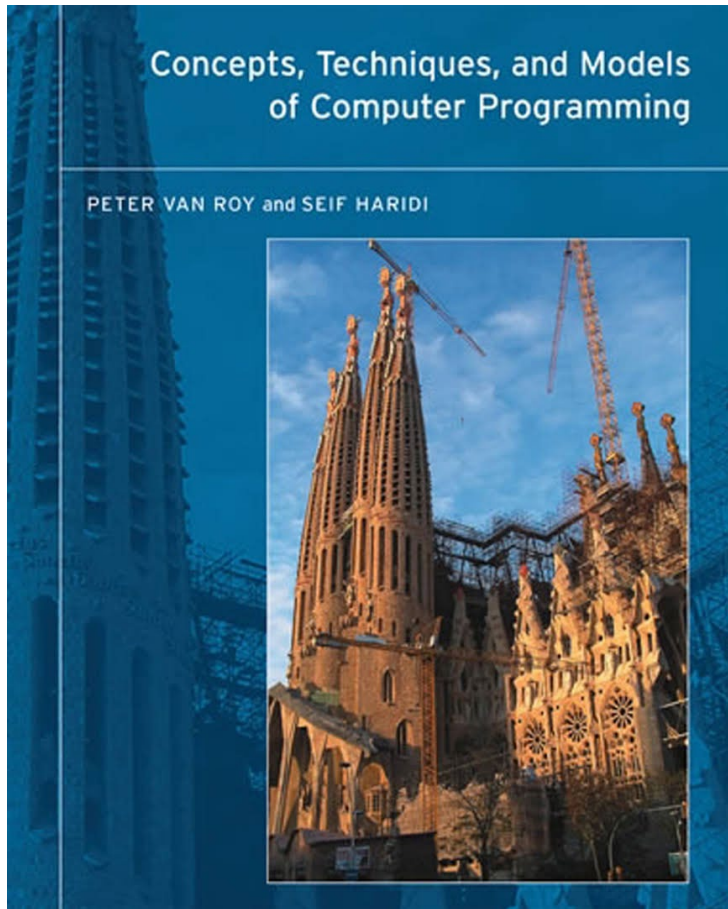
# Bibliografía recomendada



<http://adv-r.had.co.nz/>

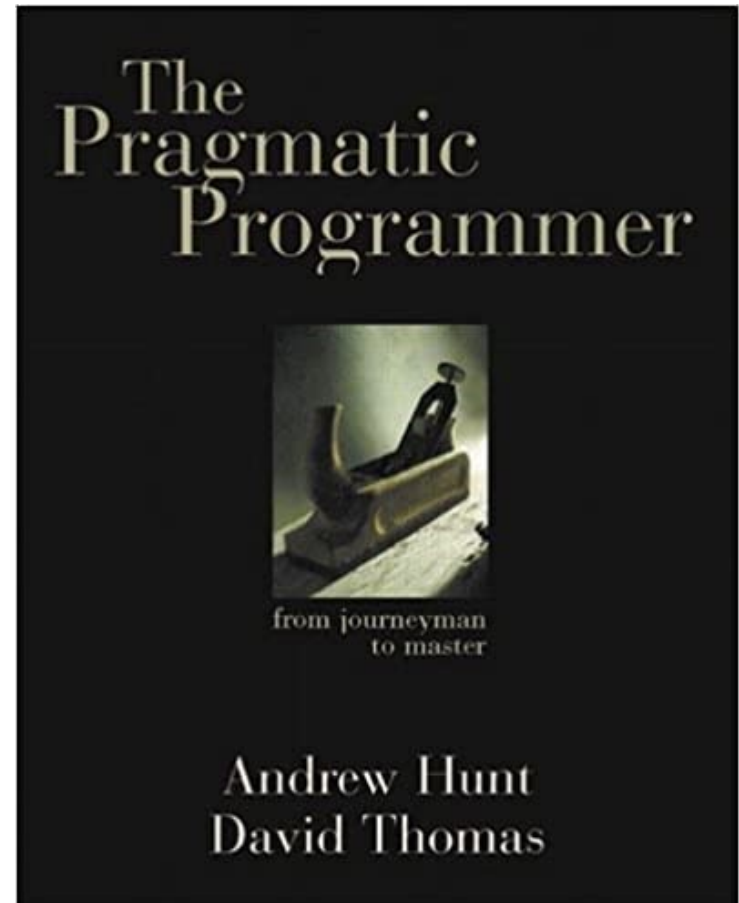


# Bibliografía recomendada



Para comprender diferencias entre R y otros lenguajes de programación

Ser mejor programador





# RStudio Cheatsheets

The cheatsheets below make it easy to use some of our favorite packages. From time to time, we will add new cheatsheets. If you'd like us to drop you an email when we do, click the button below.

[SUBSCRIBE TO CHEATSHEET UPDATES](#)

---

 [CONTRIBUTED CHEATSHEETS](#)

---

 [TRANSLATIONS](#)

---

 [HOW TO CONTRIBUTE](#)

---

# RStudio Cheat Sheets

The cheat sheets below make it easy to use some of our favorite packages. From time to time, we will add new cheat sheets. If you'd like us to drop you an email when we do, click the button below.

[SUBSCRIBE TO CHEAT SHEET UPDATES](#)
[CONTRIBUTED CHEAT SHEETS](#)
[TRANSLATIONS](#)
[+ HOW TO CONTRIBUTE](#)

## Base R Cheat Sheet

### Getting Help

#### Accessing the help files

##### ?mean

Get help of a particular function.

**help.search('weighted mean')**

Search the help files for a word or phrase.

**help(package = 'dplyr')**

Find help for a package.

#### More about an object

##### str(iris)

Get a summary of an object's structure.

##### class(iris)

Find the class an object belongs to.

### Vectors

#### Creating Vectors

c(2, 4, 6)	2 4 6	Join elements into a vector
2:6	2 3 4 5 6	An integer sequence
seq(2, 3, by=0.5)	2.0 2.5 3.0	A complex sequence
rep(1:2, times=3)	1 2 1 2 1 2	Repeat a vector
rep(1:2, each=3)	1 1 1 2 2 2	Repeat elements of a vector

#### Vector Functions

<b>sort(x)</b>	<b>rev(x)</b>
Return x sorted.	Return x reversed.
<b>table(x)</b>	<b>unique(x)</b>
See counts of values.	See unique values.

#### Selecting Vector Elements

### Programming

#### For Loop

```
for (variable in sequence){
  Do something
}
```

#### Example

```
for (i in 1:4){
  j <- 1 + 10
  print(j)
}
```

#### While Loop

```
while (condition){
  Do something
}
```

#### Example

```
while (1 < 5){
  print(1)
  1 <- 1 + 1
}
```

#### If Statements

```
if (condition){
  Do something
} else {
  Do something different
}
```

#### Functions

```
function_name <- function(var){
  Do something
  return(new_variable)
}
```

# Algunas webs útiles

Stack Overflow: <https://stackoverflow.com/>

Quick R: <https://www.statmethods.net/>

Analysis of Environmental Data: <http://finzi.psych.upenn.edu/views/Environmetrics.html>

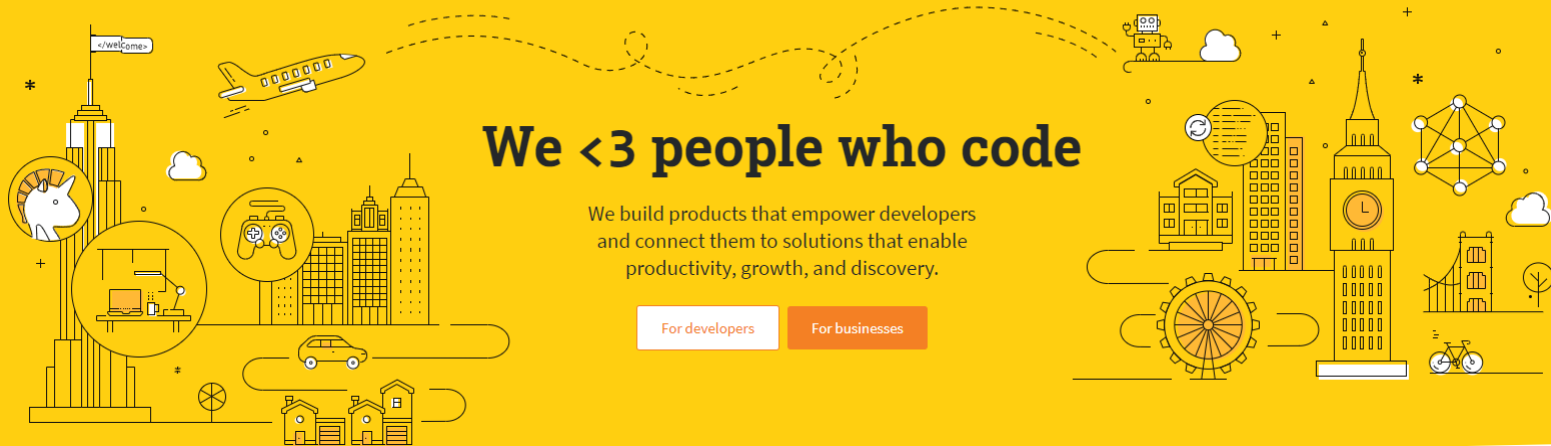
Luis Cayuela Blog: <https://luiscayuela.blogspot.com>

Data Camp: <https://www.datacamp.com/courses/tech:r>

Spatial Data Science with R: <https://www.rspatial.org/>

Advanced R: <http://adv-r.had.co.nz/>

stackoverflow Products Customers Use cases Search... Log in Sign up



We <3 people who code

We build products that empower developers  
and connect them to solutions that enable  
productivity, growth, and discovery.

For developers

For businesses

For developers, by developers

# ¿Que es R?

*R es un lenguaje de programación para análisis estadístico*

¿Cuántos de vosotros  
estáis de acuerdo?





**Trevor Branch**  
@TrevorABranch



My rule of thumb: every analysis you do on a dataset will have to be redone 10–15 times before publication. Plan accordingly. [#Rstats](#)

[Traducir Tweet](#)

12:28 a. m. · 30 sept. 2015 · Twitter Web Client

304 Retweets 320 Me gusta



**Dr. David Shiffman** @WhySharksMatter · 30 sept. 2015



En respuesta a [@TrevorABranch](#)

[@TrevorABranch](#) And then again 3-5 more times during review, grumble grumble



1

4



**Matthew J. Smith** @JungleTeuch · 14 oct. 2015



En respuesta a [@TrevorABranch](#)

[@TrevorABranch](#) Gross underestimate!



1



**Nash Turley** @NashTurley · 30 sept. 2015



En respuesta a [@TrevorABranch](#)

[@TrevorABranch](#) [@LauraLadwig](#) agreed



1







# Start small!

**You don't need to master everything at once.**

# Flujo de trabajo

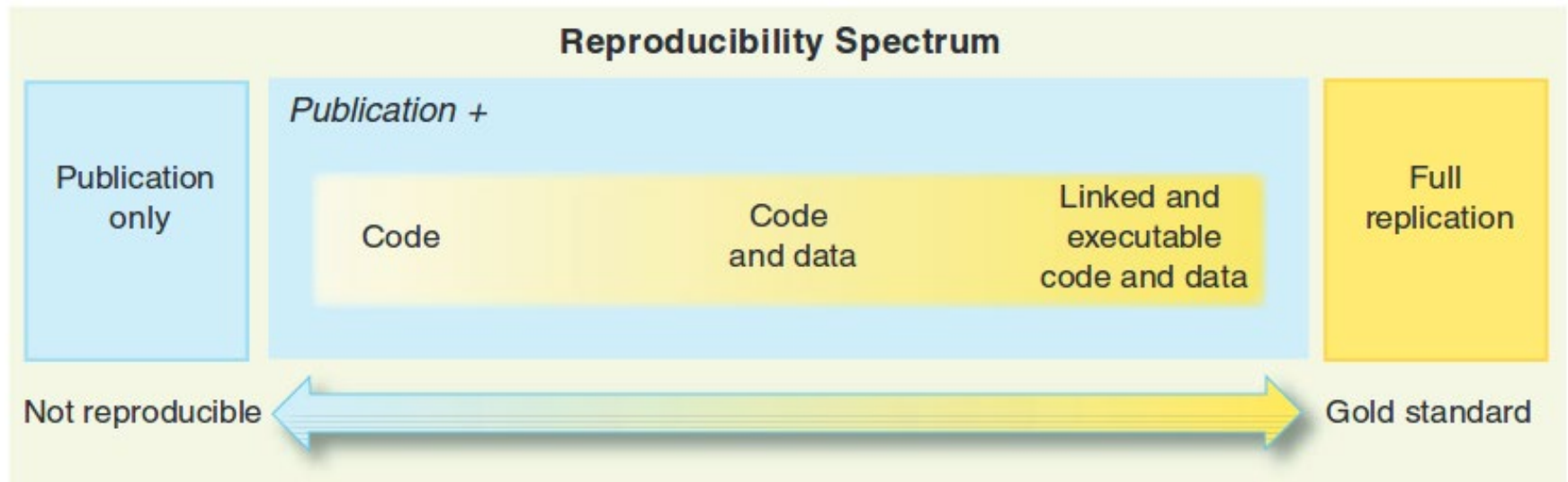


¿Qué software usáis?

# Flujo de trabajo



Peng, Roger D. 2011. "Reproducible Research in Computational Science." *Science* 334 (6060). American Association for the Advancement of Science: 1226–7.



R Open Sci: <https://ropensci.org/>

Tools for reproducible Research: <https://kbroman.org/Tools4RR/>

Good programming skills: <http://adv-r.had.co.nz/Style.html>

Un flujo de trabajo reproducible  
por Ignacio Bartomeus y Francisco  
Rodríguez Sánchez

La importancia de ciencia reproducible  
por Keith Baggerly

<https://www.youtube.com/watch?v=7gYIs7uYbMo&feature=youtu.be>

A dark, grainy, black and white photograph of a cloudy sky. A bright light source, possibly the sun or moon, is breaking through the clouds in the lower right quadrant, creating a strong lens flare and illuminating the surrounding clouds. The overall mood is dramatic and atmospheric.

# A NON REPRODUCIBLE WORKFLOW

# Esenciales cuando trabajamos en Ciencia

- ¿puedes repetir los análisis con 5 datos más y rehacer las figuras?
- ¿Consideraste el subgrupo “X” al hacer la Figura 2?
- Hay un error en los datos ¿puedes repetir TODOS los análisis y figuras?

# ¿Por qué debemos preocuparnos sobre la reproducibilidad?

## **Investigador / Académico:**

- Verificar los resultados
- Colaborar
- Avanzar la ciencia

## **Empresa**

- Código de producción
- Fiabilidad
- Reutilización
- Regulación



# ¿Cómo hacer ciencia reproducible?

1. **Organización de archivos**
2. **Gestión de datos** (buenas prácticas en el manejo de datos)
3. **Análisis de datos** basado en código: Rmarkdown.
4. **Control de versiones y escritura colaborativa.**

A E E T

ASOCIACIÓN ESPAÑOLA  
DE ECOLOGÍA TERRESTRE

Ecosistemas 25(2): 83-92 [Mayo-Agosto 2016]  
Doi.: 10.7818/ECOS.2016.25-2.11

Artículo publicado en Open Access bajo los términos  
de Creative Commons attribution Non Comercial License 3.0.

REVISIONES

ecosistemas

REVISTA CIENTÍFICA DE ECOLOGÍA Y MEDIO AMBIENTE

ISSN 1697-2473 / Open access  
disponible en [www.revistaecosistemas.net](http://www.revistaecosistemas.net)

# Ciencia reproducible: qué, por qué, cómo

F. Rodríguez-Sánchez<sup>1,\*</sup>, A.J. Pérez-Luque<sup>2,\*\*</sup>, I. Bartomeus<sup>1,\*\*</sup>, S. Varela<sup>3,\*\*</sup>

- (1) Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD-CSIC), Consejo Superior de Investigaciones Científicas, Avda. Américo Vespucio s/n, E-41092 Sevilla, España.
- (2) Laboratorio de Ecología (iEcolab), Instituto Interuniversitario Sistema Tierra (CEAMA), Universidad de Granada, Avda. del Mediterráneo s/n, Granada 18006, España.
- (3) Departamento de Ciencias de la Vida, Facultad de Biología, Ciencias Ambientales y Química, Universidad de Alcalá, Campus Universitario. Ctra. Madrid-Barcelona, Km. 33,600, 28805 Alcalá de Henares, Madrid, España.

\* Autor de correspondencia: F. Rodríguez-Sánchez [[frodriguez.work@gmail.com](mailto:frodriguez.work@gmail.com)]

\*\* Estos autores contribuyeron de manera equivalente y el orden se determinó ejecutando en R: `sample(c("AJPL", "IB", "SV"))`.

> Recibido el 08 de marzo de 2016 - Aceptado el 12 de junio de 2016

# Contesta a las siguientes preguntas – Práctica 1

- (1) ¿Qué es la Ciencia Reproducible? ¿En qué casos se consigue un 100% de la reproducibilidad?
- (2) ¿Cuáles de los beneficios que se sugieren son para ti los más importantes? Justifica tu respuesta
- (3) ¿Por qué no hace todo el mundo Ciencia Reproducible?

*use* **R**!