

Herramientas básicas para la reproducibilidad en R y RStudio

Jordi Pérez Panadés - SubDG Epidemiología, Dirección General de Salud Pública
5 de junio de 2017

Estructura de la sesión

- **R**
- **RStudio**
- Estructuración del trabajo en R
 - Uso de **proyectos** en Rstudio
 - **packrat**
 - Organización de proyectos en sistemas de **directorios**
 - **Rutas relativas**
- Guía de **estilo R**
- Herramientas para **compartir y acceder** a información
 - **Dropbox**
 - **GoogleSheets**
 - **Figshare**

R

Qué es R?

- Es un **lenguaje** de programación y un **entorno** para computación estadística, análisis de datos y creación de gráficos.
- Fue desarrollado inicialmente por **Robert Gentleman** y **Ross Ihaka** (Auckland, 1993) combinando las fortalezas de dos lenguajes existentes, **S** y **FORTRAN**.
- El resultado se llamó R, en parte al reconocimiento de la influencia de S y en parte porque es la inicial del nombre de los autores (Hornik 2008)
- Se distribuye de forma **libre** bajo la licencia GNU GPL.
- Está disponible para los sistemas operativos Windows, Macintosh, Unix y GNU/Linux.



Qué es R?



Por qué R?

- Diseñado para el **análisis de datos**
- **Productividad**
- **Flexibilidad**
- Cualquier técnica estadística, se puede construir dentro de R... o ya está disponible en algún package de forma libre.
- Forma parte de un **proyecto colaborativo y abierto**. El repositorio oficial de paquetes CRAN (The Comprehensive R Archive Network) tiene actualmente 10563 paquetes y 1380 en Bioconductor (Open source software for bioinformatics) .



Bibliografía: Grant (2016)

Por qué R?

- Excelente **ayuda y documentación**.
- Algunas revistas (Editorial Biostatistics) sólo contrastan reproducibilidad de procesos en R.



R, Lingua Franca of Statistics

Developed by statisticians, for statisticians

Statisticians communicate ideas and
methods for statistical analysis through
R code and packages.



DataCamp
Learn data analysis,
Interactively

Por qué R?

The PYPL PopularitY of Programming Language Index is created by analyzing how often language tutorials are searched on Google.

The more a language tutorial is searched, the more popular the language is assumed to be. It is a leading indicator. The raw data comes from Google Trends.

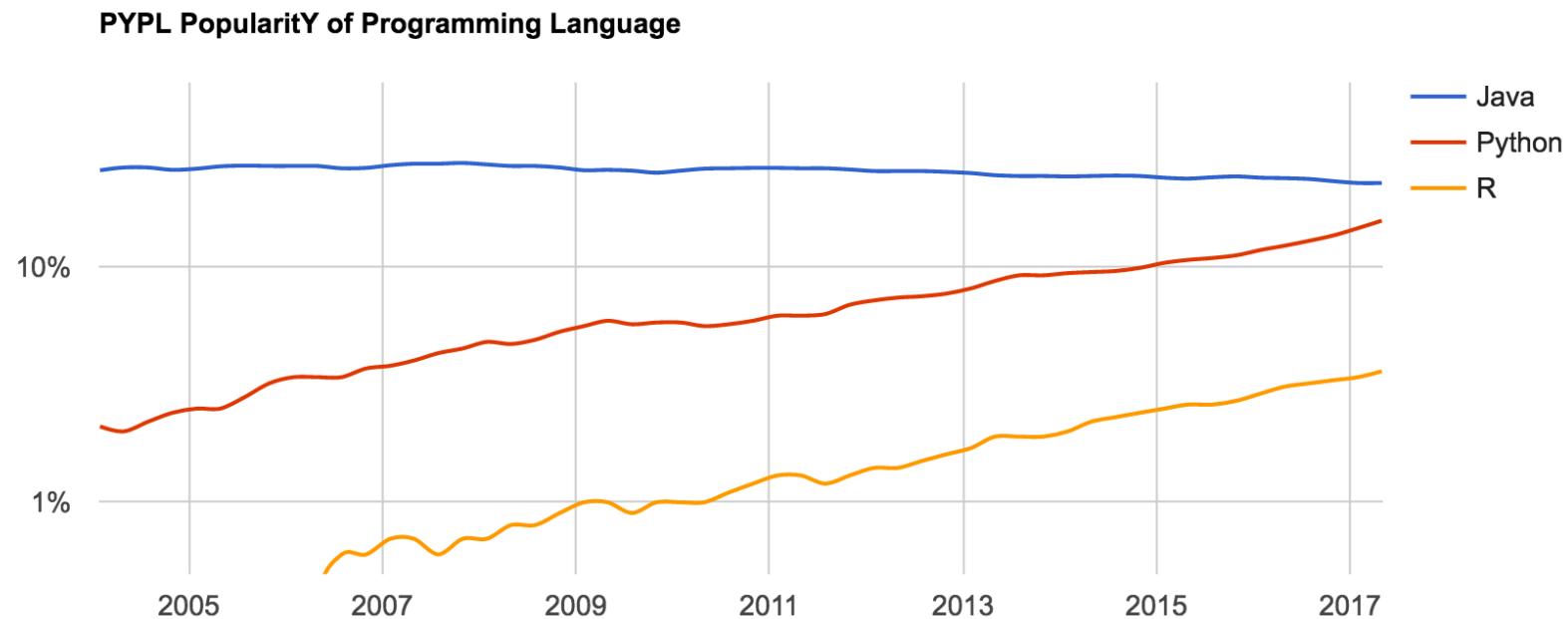
If you believe in collective wisdom, the PYPL Popularity of Programming Language index can help you decide which language to study, or which one to use in a new software project.

Worldwide, May 2017 compared to a year ago:

Rank	Change	Language	Share	Trend
1		Java	22.7 %	-1.3 %
2		Python	15.7 %	+3.5 %
3		PHP	9.3 %	-1.1 %
4		C#	8.3 %	-0.5 %
5		Javascript	7.9 %	+0.5 %
6		C++	6.9 %	-0.2 %
7		C	6.7 %	-0.1 %
8		Objective-C	3.8 %	-0.9 %
9		R	3.6 %	+0.4 %
10		Swift	2.8 %	-0.1 %
11		Matlab	2.6 %	-0.1 %

Por qué R?

Worldwide, Java is the most popular language, Python grew the most in the last 5 years (8.2%) and PHP lost the most (-4.8%)



This chart uses a [logarithmic scale](#). It can show your favorite languages

Java, Python, R



in a country

Worldwide



RStudio

Qué es RStudio ?

- RStudio es la compañía fundada por JJ Allaire (Indiana, 1969) en el año 2009 con el objetivo de crear software de código abierto para el análisis de datos y la computación estadística.
- RStudio es un **entorno de desarrollo integrado (IDE)** para R.



- Parte comercial:
y

<https://rviews.rstudio.com/2016/10/12/interview-with-j-j-allaire/>
[\(https://rviews.rstudio.com/2016/10/12/interview-with-j-j-allaire/\)](https://rviews.rstudio.com/2016/10/12/interview-with-j-j-allaire/)

Por qué RStudio?

- RStudio participa en la creación y desarrollo de paquetes importantes (`ggplot2`, `dplyr`, `tidyverse`, `lubridate`, `stringr`,...) y herramientas como RMarkdown o Shiny.
- RStudio está especialmente bien integrado con herramientas útiles para hacer investigación reproducible: (Gandrud 2015)
 - Un lenguaje estadístico para recoger y analizar datos (**R**)
 - Lenguajes para crear documentos, presentaciones, libros, artículos o webpages (**LaTeX**, **Markdown**, **Shiny**)
 - Herramientas que combinen el código fuente y la presentación de resultados mediante la creación de documentos dinámicos, gestión de archivos y control de versiones (Rpackages **knitr** y **rmarkdown**, **Git**)



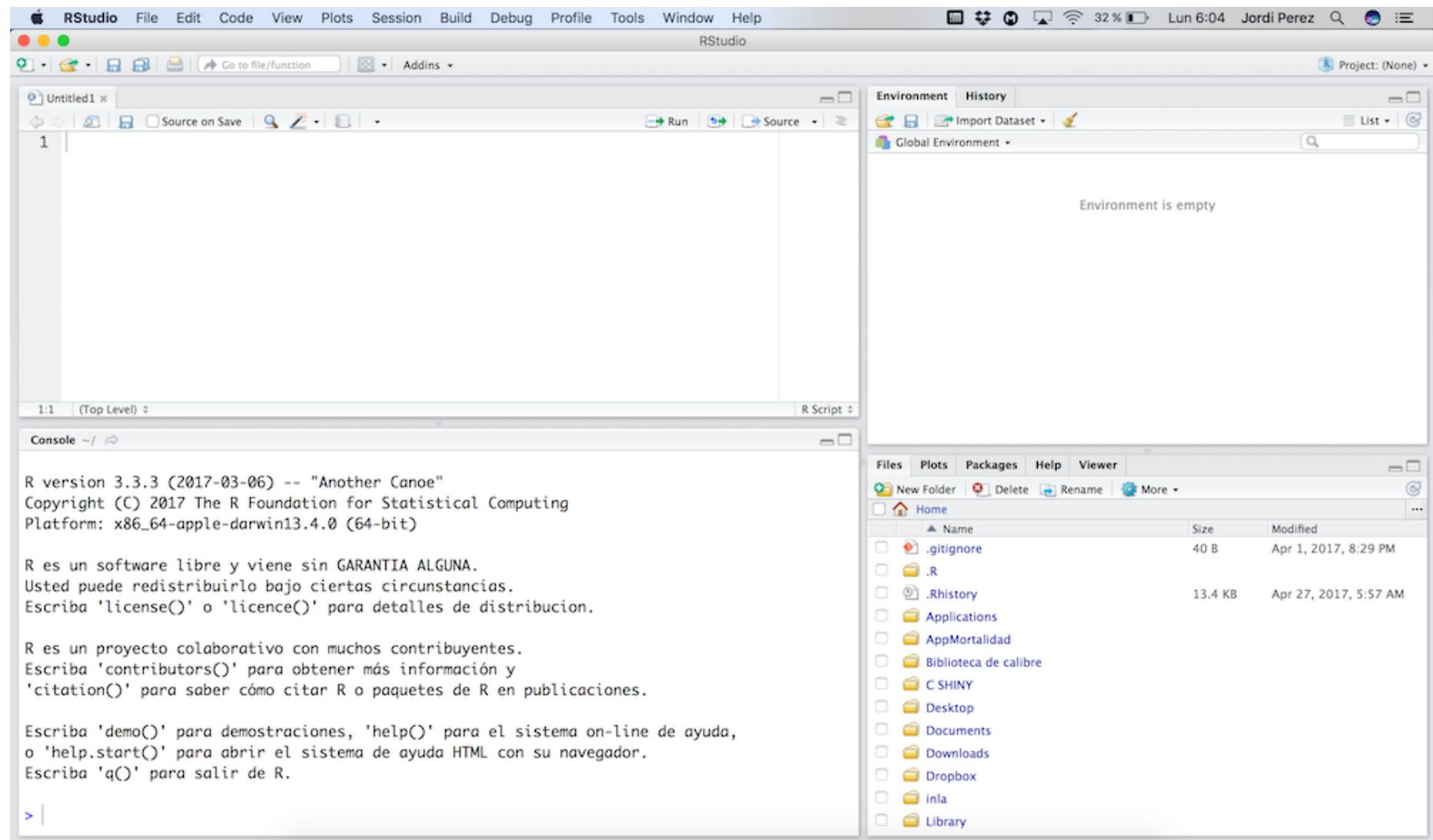
Por qué RStudio?



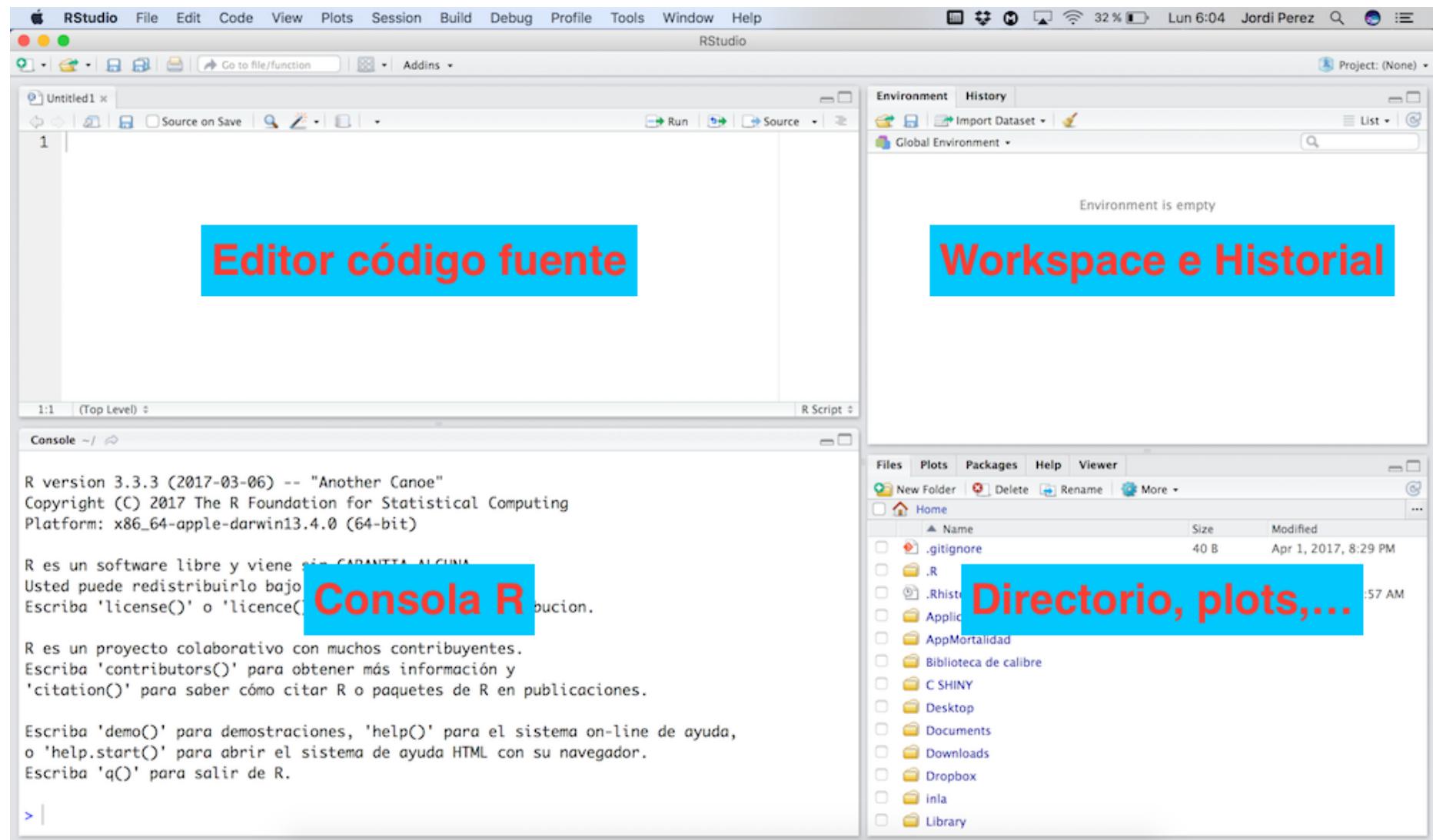
- Proporciona un entorno centralizado y bien organizado donde hacer casi todo lo que quieras con R.

... y es **gratis!**

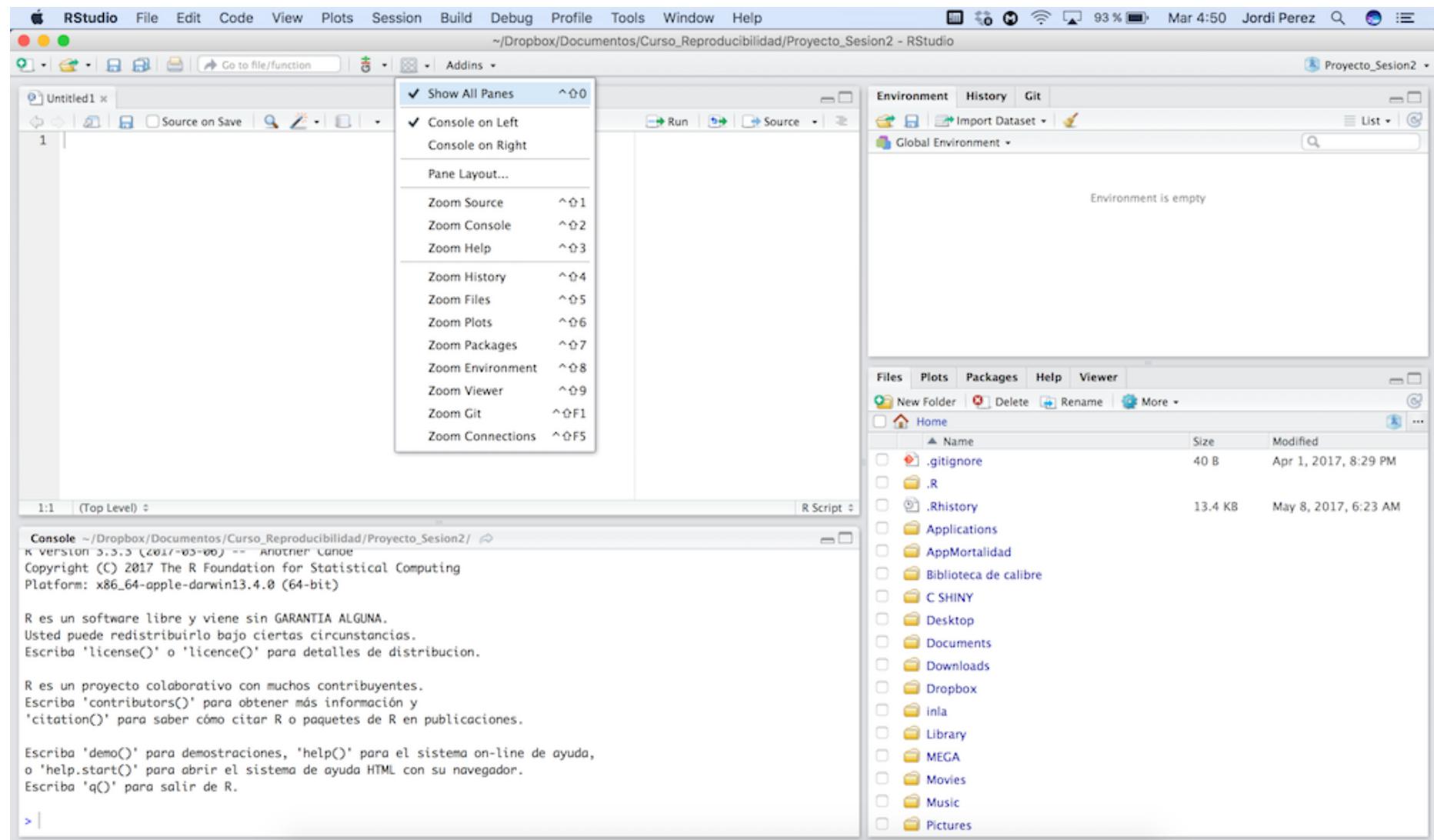
RStudio



RStudio



RStudio



RStudio

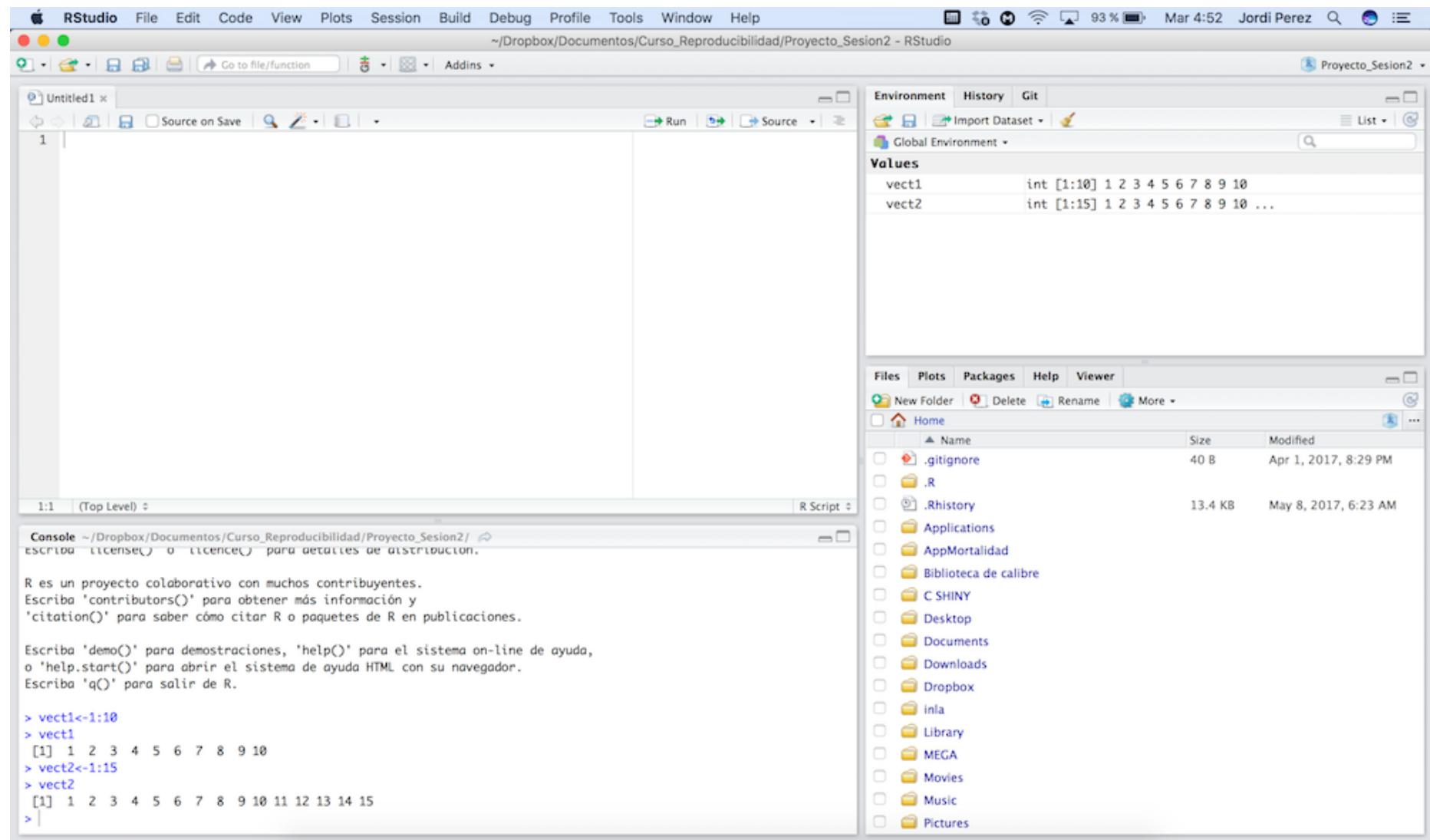
The screenshot shows the RStudio interface with the following components:

- Script Editor:** Displays an R script titled "0 Uteros Completo.R". The code reads a CSV file, performs some data manipulation (like creating a new column "Embarazo" based on "STIG"), and creates a summary table.
- Console:** Shows the R environment setup and the R version information:

 - R version 3.3.3 (2017-03-06) -- "Kangaroo Kid"
 - Copyright (C) 2017 The R Foundation for Statistical Computing
 - Platform: x86_64-apple-darwin13.4.0 (64-bit)

- Environment:** Shows the Global Environment which is currently empty.
- Files:** A file browser showing the project structure:
 - Home > Dropbox > Documentos > Curso_Reproducibilidad > Proyecto_Sesion2
 - Files: .gitignore, .RData, .Rhistory, .Rprofile, datos, figuras, informes, packrat, Proyecto_Sesion2.Rproj, r
 - Size and Modified details for each file.

RStudio



RStudio

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

- Environment pane:** Displays the global environment with objects like `vect1<-1:10`, `vect1`, `vect2<-1:15`, `vect2`, and `View(iris)`.
- Files pane:** Shows the project directory structure under "Home".
- Console pane:** Displays R startup information and a command history.
- Data frame viewer:** Shows the first 15 rows of the `iris` dataset.

Environment pane content:

```
vect1<-1:10
vect1
vect2<-1:15
vect2
View(iris)
```

Files pane content:

Name	Size	Modified
.gitignore	40 B	Apr 1, 2017, 8:29 PM
.R	13.4 KB	May 8, 2017, 6:23 AM
.Rhistory		
Applications		
AppMortalidad		
Biblioteca de calibre		
C SHINY		
Desktop		
Documents		
Downloads		
Dropbox		
inla		
Library		
MEGA		
Movies		
Music		
Pictures		

Console pane content:

```
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.

> vect1<-1:10
> vect1
[1] 1 2 3 4 5 6 7 8 9 10
> vect2<-1:15
> vect2
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> View(iris)
> |
```

Data frame viewer content:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa

Showing 1 to 15 of 150 entries

RStudio

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa

Showing 1 to 15 of 150 entries

Console ~/Dropbox/Documentos/Curso_Reproducibilidad/Proyecto_Sesion2/ 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.

```
> vect1<-1:10
> vect1
[1] 1 2 3 4 5 6 7 8 9 10
> vect2<-1:15
> vect2
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> View(iris)
> plot(iris)
```

Environment History Git

vect1<-1:10
vect1
vect2<-1:15
vect2
View(iris)
plot(iris)

Files Plots Packages Help Viewer

Zoom Export Publish

2.0 3.0 4.0 0.5 1.5 2.5 4.5 6.5
Sepal.Length
2.0 3.5 0.5 1.5 2.5 4.5 6.5
Sepal.Width
2.0 3.5 0.5 1.5 2.5 4.5 6.5
Petal.Length
2.0 3.5 0.5 1.5 2.5 4.5 6.5
Petal.Width
1.0 2.0 3.0 4.5 6.0 7.5
Species

RStudio

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa

Showing 1 to 15 of 150 entries

Console ~/Dropbox/Documentos/Curso_Reproducibilidad/Proyecto_Sesion2/ ↵
citation() para saber como 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.

```
> vect1<-1:10
> vect1
[1] 1 2 3 4 5 6 7 8 9 10
> vect2<-1:15
> vect2
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> View(iris)
> plot(iris)
> library(dygraphs)
> dygraph(mdeaths)
> |
```

Environment History Git

vect1<-1:10
vect1
vect2<-1:15
vect2
View(iris)
plot(iris)
library(dygraphs)
dygraph(mdeaths)

Files Plots Packages Help Viewer

Dec, 1976: V1: 2059

91% Mar 4:57 Jordi Perez

RStudio

The screenshot shows the RStudio interface with the following components:

- Data Viewer:** Displays the first 15 rows of the Iris dataset (Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, Species).
- Console:** Shows the R session history with commands like `vect1<-1:10`, `vect1`, `vect2<-1:15`, `vect2`, `View(iris)`, `plot(iris)`, `library(dygraphs)`, `dygraph(mdeaths)`, and `?lm`.
- Environment:** Shows the current environment with objects like `vect1`, `vect2`, `View(iris)`, `plot(iris)`, `library(dygraphs)`, `dygraph(mdeaths)`, and `?lm`.
- Help Documentation:** The `?lm` help page is open, titled "Fitting Linear Models". It includes sections for "Description", "Usage", and "Arguments".

RStudio

The screenshot shows the RStudio IDE interface. The top menu bar includes RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Window, and Help. The top right corner displays system status (91% battery, Mar 4:58, Jordi Perez). The left sidebar shows two tabs: Untitled1 (active) and iris. The main workspace displays the iris dataset as a table:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa

Showing 1 to 15 of 150 entries

The Console tab shows the following R session:

```
> vect1<-1:10
> vect1
[1] 1 2 3 4 5 6 7 8 9 10
> vect2<-1:15
> vect2
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> View(iris)
> plot(iris)
> library(dygraphs)
> dygraph(mdeaths)
> ?lm
>
```

The Environment tab shows the following objects:

- vect1<-1:10
- vect1
- vect2<-1:15
- vect2
- View(iris)
- plot(iris)
- library(dygraphs)
- dygraph(mdeaths)
- ?lm

The Help section provides links to various R resources and documentation, including:

- R Resources
 - Learning R Online
 - CRAN Task Views
 - R on StackOverflow
 - Getting Help with R
- RStudio
 - RStudio IDE Support
 - RStudio Cheat Sheets
 - RStudio Tip of the Day
 - RStudio Packages
 - RStudio Products
- Manuals
 - An Introduction to R
 - Writing R Extensions
 - R Data Import/Export
- Reference
 - The R Language Definition
 - R Installation and Administration
 - R Internals
- Packages
- Search Engine & Keywords
- Miscellaneous Material

RStudio

The screenshot shows the RStudio IDE interface. In the top-left corner, there's a preview of the 'iris' dataset as a table. Below it is the 'Console' window displaying R code and its output. The 'Help' menu is open, specifically the 'Cheatsheets' section, which is highlighted with a blue border. This section contains links to various R-related documentation and tools. The right side of the screen features the 'Source' editor with a file named 'Proyecto_Sesion2.R'. At the bottom, there are several navigation tabs: Files, Plots, Packages, Help, and Viewer.

Help

- Search
- R Help
- About RStudio
- Check for Updates
- RStudio Docs
- RStudio Support
- Cheatsheets
- Keyboard Shortcuts Help
- Markdown Quick Reference
- Roxygen Quick Reference
- Diagnostics

RStudio IDE Cheat Sheet

Data Manipulation with dplyr, tidyr

Data Visualization with ggplot2

R Markdown Cheat Sheet

R Markdown Reference Guide

Shiny Web Applications

Package Development with devtools

Source

Untitled1 x iris x

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa

Showing 1 to 15 of 150 entries

Console ~/Dropbox/Documentos/Curso_Reproducibilidad/Proyecto_Sesion2/

```
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.

> vect1<-1:10
> vect1
[1] 1 2 3 4 5 6 7 8 9 10
> vect2<-1:15
> vect2
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> View(iris)
> plot(iris)
> library(dygraphs)
> dygraph(mdeaths)
> ?lm
>
```

R Resources

- Learning R Online
- CRAN Task Views
- R on StackOverflow
- Getting Help with R

RStudio

- RStudio IDE Support
- RStudio Cheat Sheets
- RStudio Tip of the Day
- RStudio Packages
- RStudio Products

Manuals

- An Introduction to R
- Writing R Extensions
- R Data Import/Export

The R Language Definition

R Installation and Administration

R Internals

Reference

Packages

Search Engine & Keywords

Miscellaneous Material

RStudio

The RStudio IDE is an Integrated Development Environment in R that comes in three versions

The screenshot shows the RStudio IDE Cheat Sheet page on a Mac OS X desktop. The browser window title is "rstudio-IDE-cheatsheet". The page content includes a header with the RStudio logo and a "Documents and Apps" section showing icons for Shiny, R Markdown, and Beamer files. It also features sections for "Write Code", "R Support", and "RStudio Pro Features". Each section contains a screenshot of the RStudio interface with callouts explaining specific features like code completion, file navigation, and project management.

Desktop IDE
A local version of the IDE for your desktop

Open Source Server
for larger compute resources and remote access

Professional Server
for teams that share large compute resources, large data, and uniform environments for collaboration

Download all at www.rstudio.com. Each provides the same useful interface:

RStudio IDE Cheat Sheet

learn more at www.rstudio.com

Documents and Apps

Open Shiny, R Markdown, knitr, Sweave, LaTeX, .Rd files and more in Source Pane

Check spelling
Render output
Choose output format
Choose output location
Insert code chunk
Jump to previous chunk
Jump to next chunk
Run selected lines
Publish to server
Show file outline
Access markdown guide at Help > Markdown Quick Reference
Jump to chunk
Set knitr chunk options
Run this and all previous code chunks
Run this code chunk
Multiple cursors/column selection with Alt + mouse drag
Code diagnostics that appear in the margin. Hover over diagnostic symbols for details.
Syntax highlighting based on your file's extension
Tab completion to finish function names, file paths, arguments, and more.
Multi-language code snippets to quickly use common blocks of code.
Jump to function in file
Change file type
Path to displayed directory
Working Directory
Press ↑ to see
Maximize, minimize panes
Drag pane
RStudio recognizes that files named app.R, server.R, ui.R, and global.R belong to a shiny app

Run app
Choose location to view app
Publish to shinyapps.io or server
Manage publish accounts

Write Code

Navigate tabs
Open in new window
Save
Find and replace
Compile as notebook
Run selected code
Cursors of shared users
Re-run previous code
Source with or without Echo
Show file outline
Load workspace
Save workspace
Delete all saved objects
Search inside environment
Choose environment to display from list of parent environments
Display saved objects by type with short description
View in data viewer
View function source code

R Support

Import data file with wizard
History of past commands to run/add to source
Display .RPres slideshows
File > New File > R Presentation

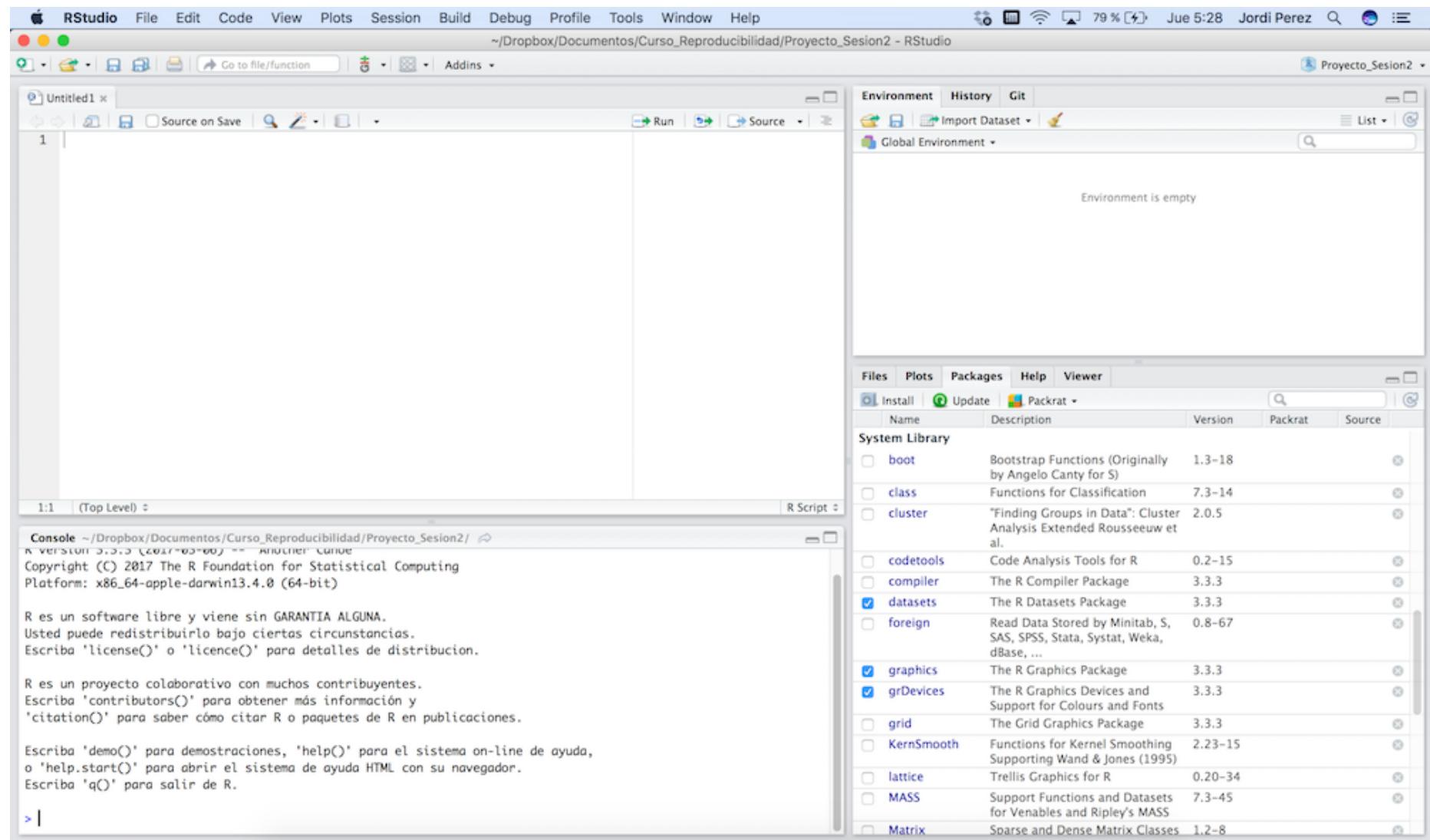
RStudio Pro Features

Share Project with Collaborators
Active shared collaborators
Start new R Session in current project
Close R Session in project
Select R Version
New Project...
Open Project...
Close Project
Share Project...
RStudio-Essentials
shiny-examples
Clear Project List
Project Options...
Name of current project
File > New Project
RStudio saves the call history, workspace, and working directory associated with a project. It reloads each when you re-open a project.

Project System

RStudio opens plots in a dedicated Plots pane
Navigate recent plots
Open in window
Export plot
Delete plot
Delete all plots
GUI Package manager lists every installed package
Install Packages
Update Packages
Create reproducible package library for your project
Click to load package with Package
Package
Delete

RStudio



RStudio

The screenshot shows the RStudio IDE interface. The top menu bar includes: RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Window, Help. The title bar indicates the current project is "Proyecto_Sesion2".

Environment pane: Shows a list of R objects: vect1<-1:10, vect1, vect2<-1:15, vect2, View(iris), plot(iris), library(dygraphs), dygraph(mdeaths), ?lm.

Console pane: Displays the R command `vect1<-1:10` and its output.

Data pane: Displays the "iris" dataset as a table with columns: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, Species. Rows 1 through 15 are shown, with a note at the bottom indicating 150 entries in total.

Help pane: The "write" function is currently selected in the search results. The help text for "write" states: "The data (usually a matrix) x are written to file file. If x is a two-dimensional matrix you need to transpose it to get the columns in file the same as those in the internal representation." A tooltip also appears, stating: "Press F1 for additional help". Other help topics listed include: write.csv, write.csv2, write.dcf, write.ftable, write.socket, write.table, wr, write, write.base, write.util, write.csv.util, write.csv2.util, write.dcf.util, write.ftable.util, write.socket.util, write.table.util, wr.util, and write.util.

Bottom right sidebar: Contains links to R Resources, RStudio, Manuals, Reference, Packages, and Miscellaneous Material.

RStudio

The screenshot shows the RStudio interface with the following components:

- Environment:** Shows a list of objects: vect1<-1:10, vect1, vect2<-1:15, vect2, View(iris), plot(iris), library(dygraphs), dygraph(mdeaths), ?lm.
- Console:** Displays R code and its output. A tooltip is shown over the output of the command `View(iris)`:

```
[1] 1 2 3 4 5 6 7 8 9 10
[1] 1 2 3 4 5 6 7 Sepal.Length 14 15
[1] 1 2 3 4 5 6 7 Sepal.Width 14 15
[1] 1 2 3 4 5 6 7 Petal.Length 14 15
[1] 1 2 3 4 5 6 7 Petal.Width 14 15
[1] 1 2 3 4 5 6 7 Species 14 15
```
- Help:** The right sidebar contains links to R Resources, RStudio, Manuals, Reference, Packages, and Miscellaneous Material.

Below the sidebar, there is a message: "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."

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa

RStudio

The screenshot shows the RStudio interface with the following components:

- Environment pane:** Displays the global environment with objects like vect1, vect2, View(iris), plot(iris), library(dygraphs), dygraph(mdeaths), ?lm.
- Console pane:** Shows R code and its output. The code includes: > vect1<-1:10, > vect1, [1] 1 2 3 4 5 6 7 8 9 10, > vect2<-1:15, > vect2, [1] 1 2 3 4 5 6 7 8 9 10 11 12, > View(iris), > plot(iris), > library(dygraphs), > dygraph(mdeaths), > ?lm, > write.csv(x=iris[1:100,"Sepal.Length"],). The output for vect1 and vect2 is shown as vectors of length 10 and 15 respectively.
- Help pane (bottom right):** A tooltip for the 'file' argument is displayed, stating: "either a character string naming a file or a connection open for writing. "" indicates output to the console." It also includes links to R Resources, RStudio, Manuals, Reference, Packages, and Miscellaneous Material.
- File Explorer pane:** Shows the project structure under 'Proyecto_Sesion2'.
- Search bar:** Located at the top center of the interface.

RStudio

The screenshot shows the RStudio interface with the following components:

- Top Bar:** RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Window, Help.
- Left Panel:** A data frame titled "iris" with columns Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species. The data shows 15 entries of setosa species.
- Console:** Displays R code and its output. The code includes creating vectors vect1 and vect2, viewing the iris dataset, plotting it, and writing a CSV file.
- Tools Menu (Open):** Shows options like Install Packages..., Check for Package Updates..., Version Control, Shell..., Addins, Keyboard Shortcuts Help (selected), Modify Keyboard Shortcuts..., Project Options..., and Global Options... .
- Environment Tab:** Shows the current session environment with objects like vect1, vect2, View(iris), plot(iris), library(dygraphs), dygraph(mdeaths), and ?lm.
- Help Tab:** Provides links to R Resources (Learning R Online, CRAN Task Views, R on StackOverflow, Getting Help with R), RStudio Support (IDE Support, Cheat Sheets, Tip of the Day, Packages, Products), Manuals (Introduction to R, Writing R Extensions, Data Import/Export), Reference (R Language Definition, Installation and Administration, Internals), Packages, and Search Engine & Keywords.
- Bottom:** Navigation icons and a search bar.

RStudio

The screenshot shows the RStudio interface with the following components:

- Script Editor (Left Panel):** Displays an R script titled "Untitled1" with code related to neighborhood relationships and WinBUGS preparation. The code includes comments in Spanish and some English, such as "Environment is empty". Lines 117 and 136 are collapsed.
- Environment Browser (Top Right Panel):** Shows the "Global Environment" tab with the message "Environment is empty".
- Help Sidebar (Bottom Right Panel):** Contains links to R Resources, RStudio support, manuals, reference, packages, and miscellaneous material.

```
99 #2013/01/24 RELACIONES DE VECINDAD DEL RINCON DE ADEMUS
100 #####
101 #Hemos visto en resultados de Diabetes y Pulmón que el Rincón de Ademuz va por libre.
102 #Hablo con Migue y me dice que le ponga alguna relación de "vecino" con alguien cercano.
103 #Tras ver el mapa, los suyo serían:
104 # 317 - 46041 - Aras de Alpuente "vecino de"
105 # 364 - 46088 - Casas Bajas
106 # 477 - 46201 - Puebla de San Miguel "vecino de"
107 # 312 - 46036 - Alpuente
108 #Para identificar los municipios en la lista de relaciones de vecindades creamos kk
109 kk<-cbind(1:542,as.numeric(carto$INE_MUN), as.character(carto$INE_MUN)), as.character(carto$NOMBRE))
110 kk[match(c("317","364","477","312"), kk[,2]),]
111 #[,1] [,2] [,3] [,4]
112 #[1,] "176" "317" "46041" "Aras de Alpuente" 110 en smoothASR
113 #[2,] "223" "364" "46088" "Casas Bajas" 98
114 #[3,] "336" "477" "46201" "Puebla de San Miguel" 76
115 #[4,] "171" "312" "46036" "Alpuente" 108
116 #
117 #
118 #Añadimos relaciones de vecindad:
119 Veci[[317]]<-as.integer(c(Veci[[317]],364))
120 Veci[[364]]<-as.integer(c(Veci[[364]],317))
121 Veci[[477]]<-as.integer(c(Veci[[477]],312))
122 Veci[[312]]<-as.integer(c(Veci[[312]],477))
123 #####
124
125
126 #Transformamos la estructura de vecindades para pasarl a WinBUGS
127 VeciWB<-nb2WB(Veci)
128 #VeciWB tiene tres objetos: adj, weights y num (argumentos para
129 # la función CAR (espacial) de WINbugs)
130
131 ngredad<-11 #nperiods (el último grupo de edad ahora sí lo consideraremos porque quizás mejore la estimación del grupo 19 y
132 # para tener las estimaciones de las tasas específicas)
133 nreg<-542 # Número de regiones
134
135 #Estructura de vecindades grupos de edad (8 grupos)
136 VeciT<-list()
137 VeciT$num<-c(1.rep(2,ngredad-2),1)
138 (Top Level) :
```

RStudio

The screenshot shows the RStudio interface with the following components:

- Top Bar:** RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Window, Help.
- Header Bar:** ~/Dropbox/Documentos/Curso_Reproducibilidad/Proyecto_Sesion2 - RStudio, Vie 13:06 Jordi Perez.
- Left Panel:** An R script editor titled "Untitled1" containing R code. The code is related to preparing data for WinBUGS, including creating VeciWB objects, summing weights, and defining VeciT variables. It also includes a function "inits" for generating initial values for parameters like inter, mediainter, sd.inter, etc.
- Right Panel:** A "Global Environment" pane showing that the environment is empty. Below it is a "Viewer" pane displaying RStudio resources and manuals.
- Bottom Panel:** A "Console" tab.
- Help Section:** A large sidebar on the right side of the interface, titled "RStudio". It contains links to "R Resources", "Manuals", "Reference", "Packages", and "Miscellaneous Material".

```
99
100 #2013/01/24 RELACIONES DE VECINDAD DEL RINCON DE ADEMUZ
101 #####
102 #####
103 #####
104 #####
105 #####
106 #####
107 #####
108 #####
109 #####
110 #####
111 #####
112 #####
113 #####
114 #####
115 #####
116 #####
117 #####
118 #####
119 #####
120 #####
121 #####
122 #####
123 #####
124 #####
125 #####
126 #Transformamos la estructura de vecindades para pasarl a WinBUGS
127 VeciWB<-nb2WB(Veci)
128 #VeciWB tiene tres objetos: adj, weights y num (argumentos para
129 # la funci?n CAR (espacial) de WINbugs)
130
131 ngredad<-11 #nperiods (el ?ltimo grupo de edad ahora s? lo consideramos porque quiz? mejor? la estima del grupo 19 y
132 #para tener las estimaciones de las tasas espec?ficas)
133 nreg<-542 # Numero de regiones
134
135 #Estructura de vecindades grupos de edad (8 grupos)
136 VeciT<-list()
137 VeciT$num<-c(1,rep(2,ngredad-2),1)
138 VeciT$weights<-rep(1,sum(VeciT$num))
139 VecinosTemp<-2
140 for (i in 2:(ngredad-1)){
141   VecinosTemp<-c(VecinosTemp,i-1,i+1)
142 }
143 VecinosTemp<-c(VecinosTemp,(ngredad-1))
144 VeciT$adj<-VecinosTemp
145
146
147 #Ya estamos en situaci?n de poder correr el modelo para estimar las tasas especificadas por edad suavizadas de cada
148 #municipio.
149 inits<-function(){list(inter=rnorm(nreg, -6,1),
150                         mediainter=rnorm(1,0,0.1),
151                         sd.inter=rnorm(1,0,0.5),
152                         sd.het=rnorm(1,0,0.2),
153                         sd.spot=rnorm(1,0,0.2),
154                         ro=rnorm(1,0.3,0.6),
155                         S=matrix(rnorm(ngredad*nreg,0,1),nrow=ngredad),
156                         H=matrix(rnorm(nreg*nreg,0,1),nrow=nreg))}
```

RStudio

The screenshot shows the RStudio IDE interface. The top menu bar includes RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Window, and Help. The status bar at the bottom right shows the date and time as Vie 13:07 Jordi Perez.

The code editor on the left displays an R script with several functions defined:

```
10 # 20170516.
11 # Esta función lee una comarca en literal y devuelve su CodComa
12 DameCodComa2<-function(coma){
13   tabla <- utils::read.csv2(file="S:/SSCC/EPI/EES/SISP/Jordi/POBLACIONES/Poblaciones Referenciadas Comarcas2017
/cartounisCV.csv",
                           colClasses = c('numeric','character','character','numeric'
                           ,'character'), stringsAsFactors = F)
14   return(tabla[match(toupper(coma),toupper(tabla$NomComa)),"CodComa"])
15 }
16
17
18 # 20170516.
19 # Esta función lee una comarca en literal y devuelve su CodComa
20 DameNomComa<-function(codcomo){
21   tabla <- utils::read.csv2(file="S:/SSCC/EPI/EES/SISP/Jordi/POBLACIONES/Poblaciones Referenciadas Comarcas2017
/cartounisCV.csv",
                           colClasses = c('numeric','character','character','numeric'
                           ,'character'), stringsAsFactors = F)
22   return(tabla[match(codcomo,tabla$CodComa),"NomComa"])
23 }
24
25
26 # 20160209
27 # Lista de DS15SP en el orden geográfico habitual:
28 ds15sp <- c("Vinaròs","Castelló","LaPlana","Sagunt","ClinicMr","ArnaulLiria","Manises","Requena",
            "VGral","Peset","LaFafe","LaRibera","Gandia","Dénia","XàtivaOnt","Alcoi",
            "MarinaB","SantJoan","Elda","Elx","AGral","Orihuela","Torrevieja","València","Alacant")
29 pcv<- read.table("S:/SSCC/EPI/EES/SISP/Jordi/SISP15/SISP15 Plan Códigos Variables SinA.csv", sep=";", header=T,
colClasses='character', stringsAsFactors=F)
30 codmuni <- pcv[pcv$VbleNom=="Municipios", "NivelesCod" ]
31
32 # 20150518 Revisada.
33 #Esta función lee un CodMuni y devuelve su DS15SP
34 DameDS<-function(CodMuni){
35   tabla<-read.table("S:/SSCC/EPI/EES/SISP/Jordi/POBLACIONES/Poblaciones Referenciadas DS15SP/Tabla_CodMuni_DS15SP
.csv", sep=";", header=T, colClasses="character")
36   return(tabla[match(CodMuni,tabla$CodMuni),"DS15SP"])]
37 }
38
39 # 20150518 Revisada.
40 #Esta función lee un municipio en literal y devuelve su CodMuni
```

The Environment panel on the right shows the Global Environment is empty.

The Resources panel on the right contains links to R Resources, RStudio, Manuals, Reference, Packages, and Miscellaneous Material.

RStudio

The screenshot shows the RStudio IDE interface. On the left, the script editor displays a file named 'FuncionesRJordi 2015.r' containing R code. The code includes several functions such as DameCodComa2, DameNomComa, DameDS, DameCodMuni, and AgrupaDe0aGrEdad85. The right side of the interface features the Environment browser, which shows the 'Global Environment' tab with the message 'Environment is empty'. Below the browser is a navigation bar with tabs for 'Files', 'Plots', 'Packages', 'Help', and 'Viewer'. The bottom right corner contains a 'Help' section titled 'RStudio Resources' with links to various R-related documentation and support resources.

```
10 # 20170516.
11 # Esta función lee una comarca en literal y devuelve su CodComa
12 DameCodComa2<-function(coma){}
13
14 # 20170516.
15 # Esta función lee una comarca en literal y devuelve su CodComa
16 DameNomComa<-function(codcoma){}
17
18 # 20160209
19 # Lista de DS15SP en el orden geográfico habitual:
20 ds15sp <- c("Vinaròs","Castelló","La Plana","Sagunt","Clínic Mr","Arnaulliria","Manises","Requena",
21 "VGral","Peset","La Fe","La Ribera","Gandia","Dénia","XàtivaOnt","Alcoi",
22 "MarinaB","SantJoan","Elda","Elx","AGral","Orihuela","Torrevieja","València","Alacant")
23 pcv<- read.table("S:/SSCC/EPI/EES/SISP/Jordi/SISP15/SISP15 Plan Códigos Variables SimA.csv", sep=",", header=T,
24 colClasses='character', stringsAsFactors=F)
25 codmunicipis <- pcv[pcv$VbleNom=="Municipios", "NivelesCod" ]
26
27 # 20150518 Revisada.
28 #Esta función lee un CodMuni y devuelve su DS15SP
29 DameDS<-function(CodMuni){}
30
31 # 20150518 Revisada.
32 #Esta función lee un municipio en literal y devuelve su CodMuni
33 DameCodMuni<-function(muni){}
34
35 # 20150518 Revisada.
36 #Esta función lee un CodMuni y devuelve su municipio en literal
37 # (OOJO que puede devolver más de uno porque metí
38 # pedanías, pueblos y diferentes formas de llamar al municipio)
39 DameMuni<-function(CodMuni){}
40
41 # 20150521 Creada.
42 # Esta función lee un vector con edades de 0 a 90ymas y lo agrupa en 19 niveles ( GrEdad85):
43 AgrupaDe0aGrEdad85 <- function(v){}
44
45 # 20150522 Creada.
46 # Esta función lee un vector con edades de 0 a 100ymas y lo agrupa en 19 niveles ( GrEdad85):
47 AgrupaDe0aGrEdad100 <- function(v){}
```

Environment History Git
Global Environment
Environment is empty

Files Plots Packages Help Viewer
Home Find In Topic

R Resources RStudio
Learning R Online RStudio IDE Support
CRAN Task Views RStudio Cheat Sheets
R on StackOverflow RStudio Tip of the Day
Getting Help with R RStudio Packages
RStudio Products

Manuals
An Introduction to R The R Language Definition
Writing R Extensions R Installation and Administration
R Data Import/Export R Internals

Reference
Packages Search Engine & Keywords

Miscellaneous Material

Estructuración del trabajo en R. Proyectos

Estructuración del trabajo

La investigación reproducible proporciona al investigador:

- mejores hábitos de trabajo (**proyectos**, evitar duplicidades, **organización**, depuración, ...)
- mejor trabajo en equipo, más fácil y útil para tus colaboradores (**guía de estilo**, **sintaxis**, **código claro**, ...)
- modificaciones sencillas (documentos dinámicos)
- mayor impacto en la investigación

(Gandrud 2015)

RStudio Projects

El uso de proyectos en RStudio tiene 5 **ventajas fundamentales**:

recuperación configuración proyecto y última sesión

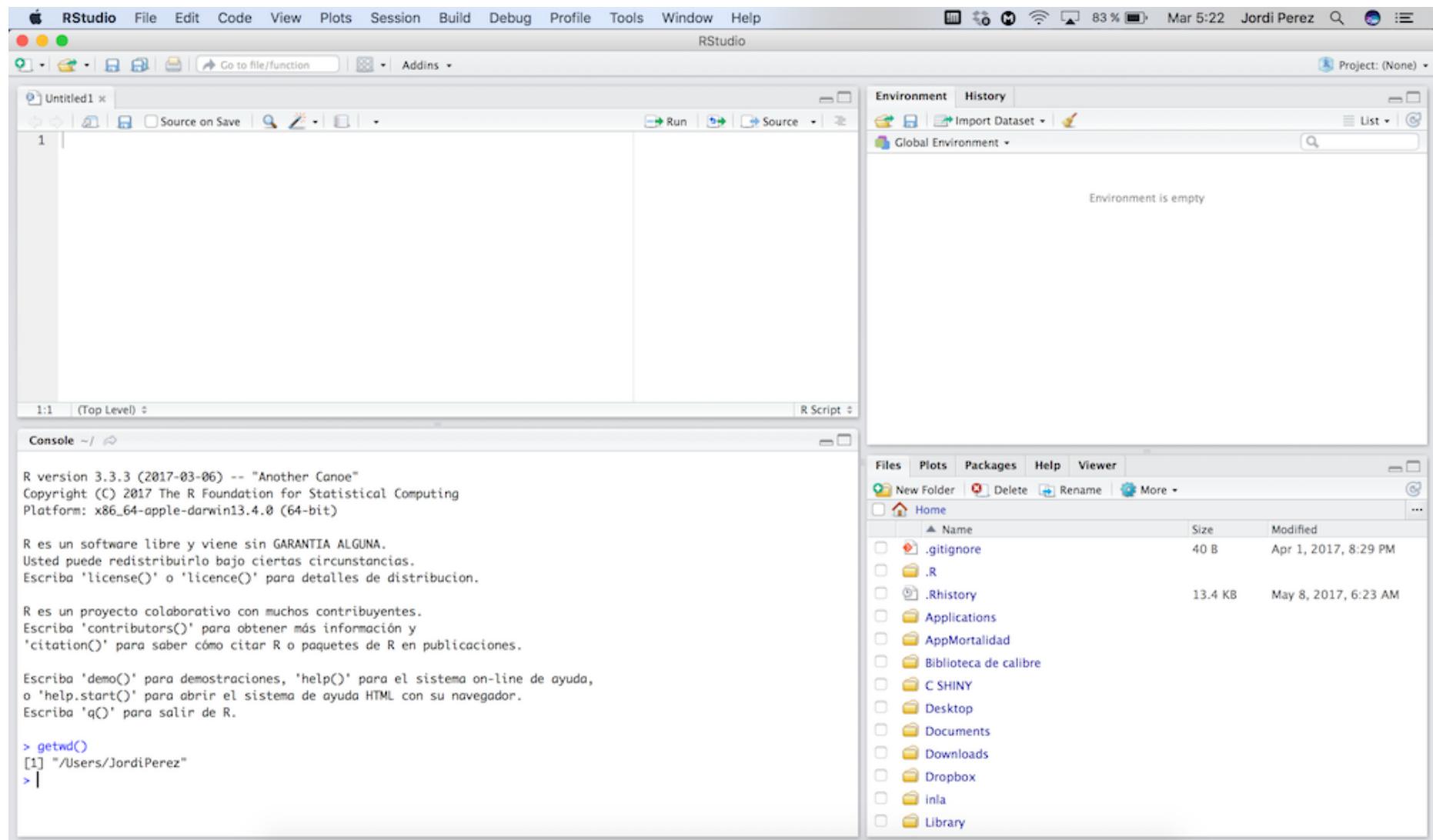
working directory

packrat y github

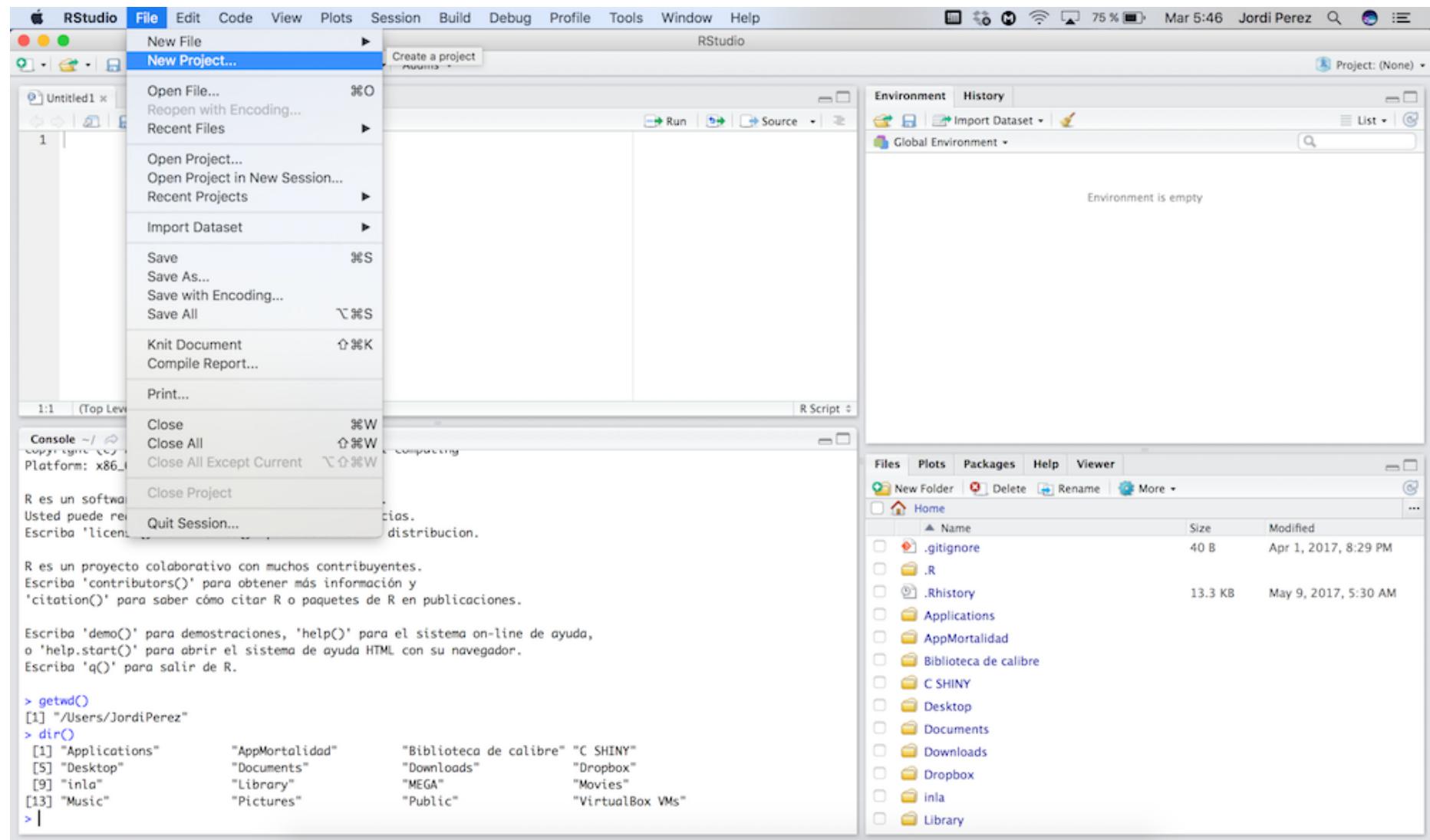
estructura de directorio

rutas relativas

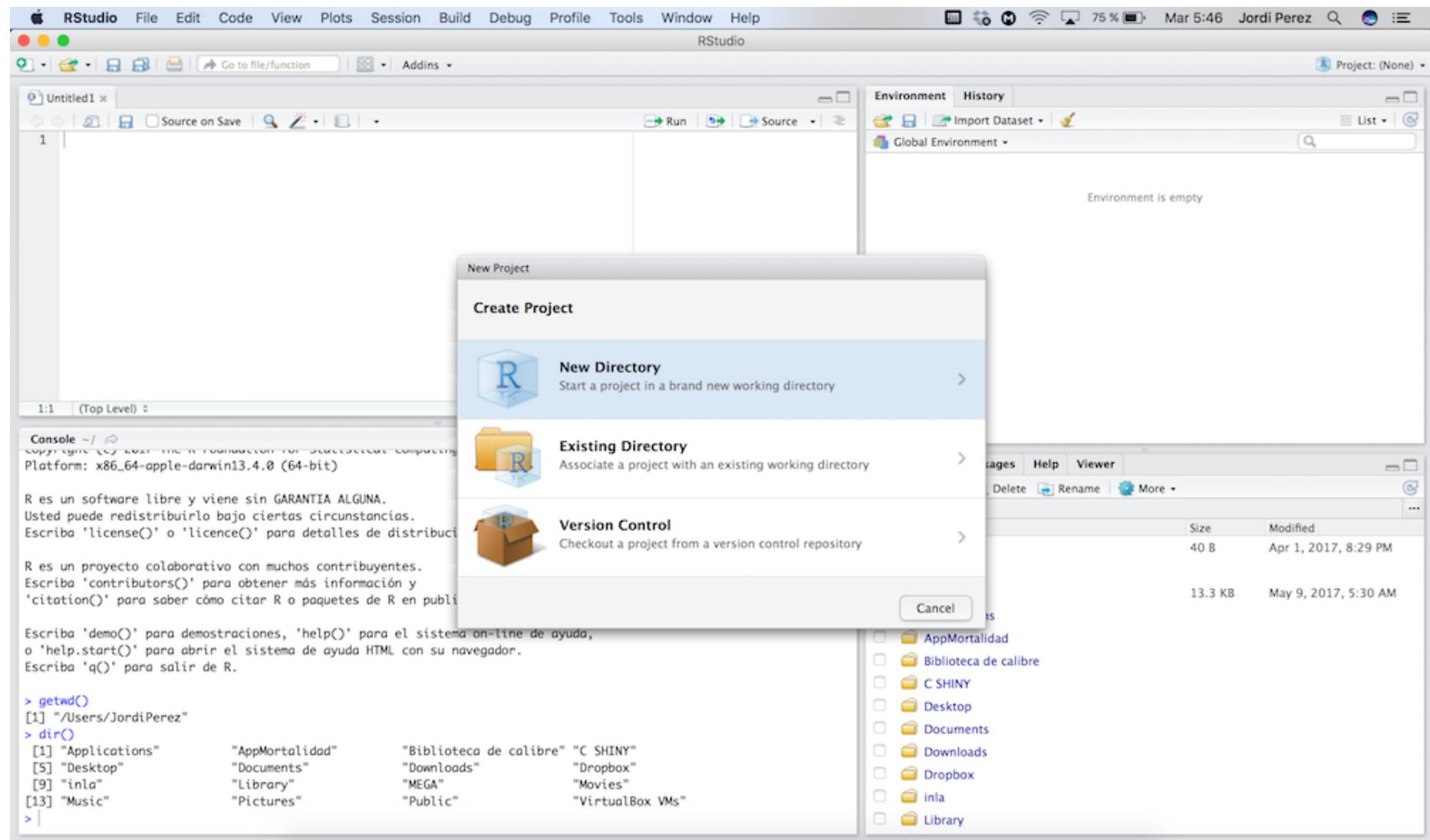
RStudio NewProject



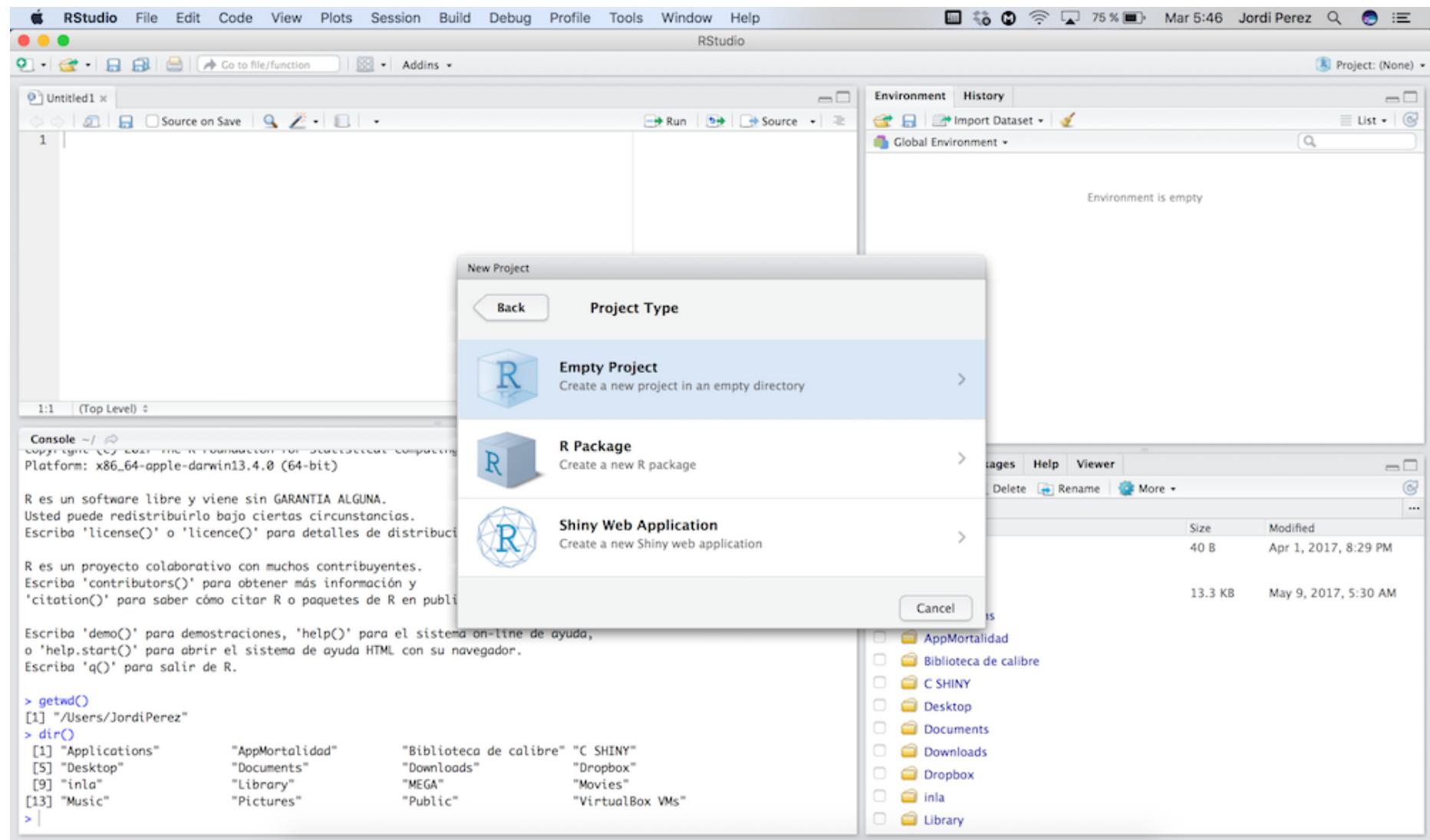
RStudio NewProject



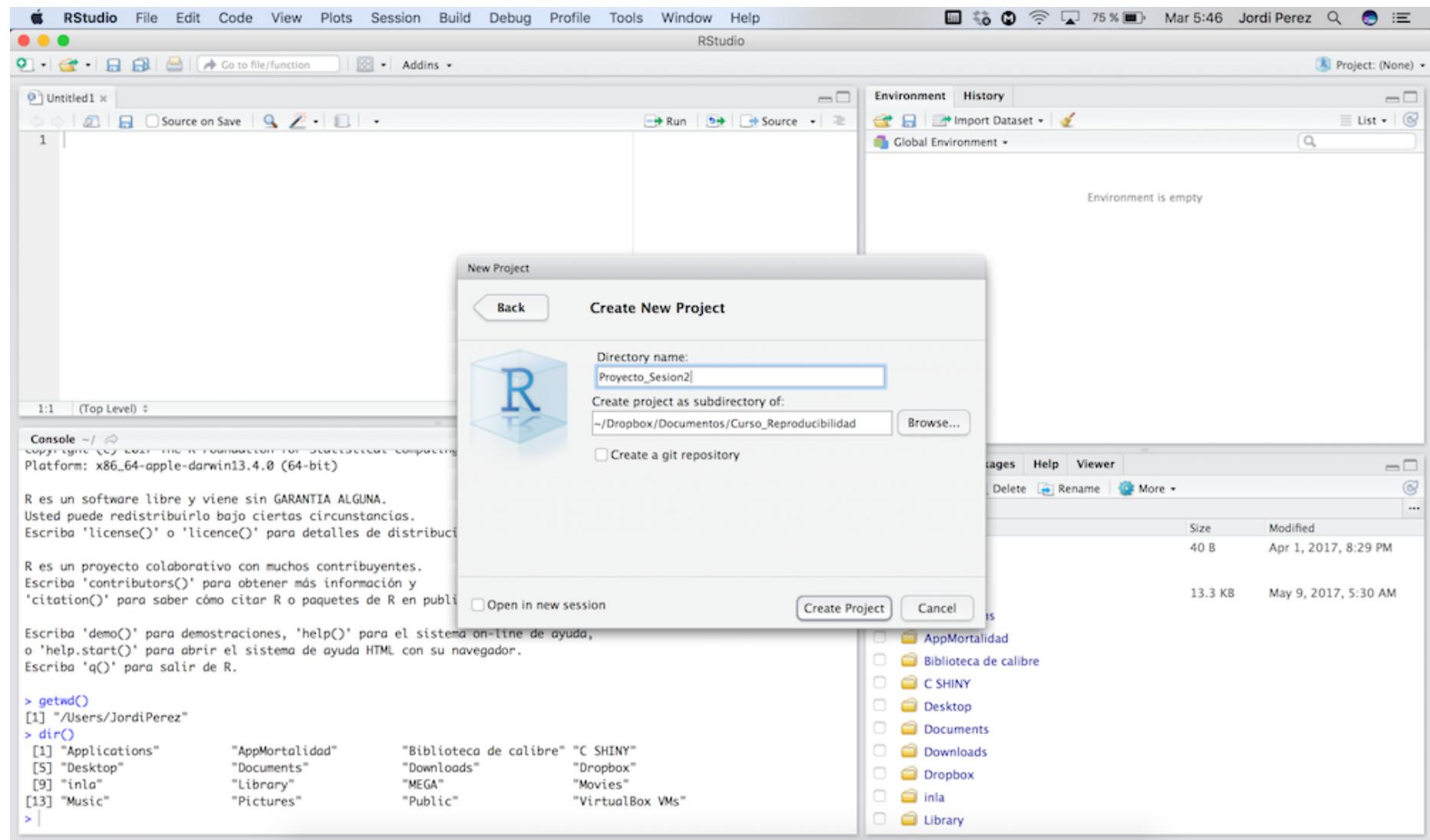
RStudio NewProject



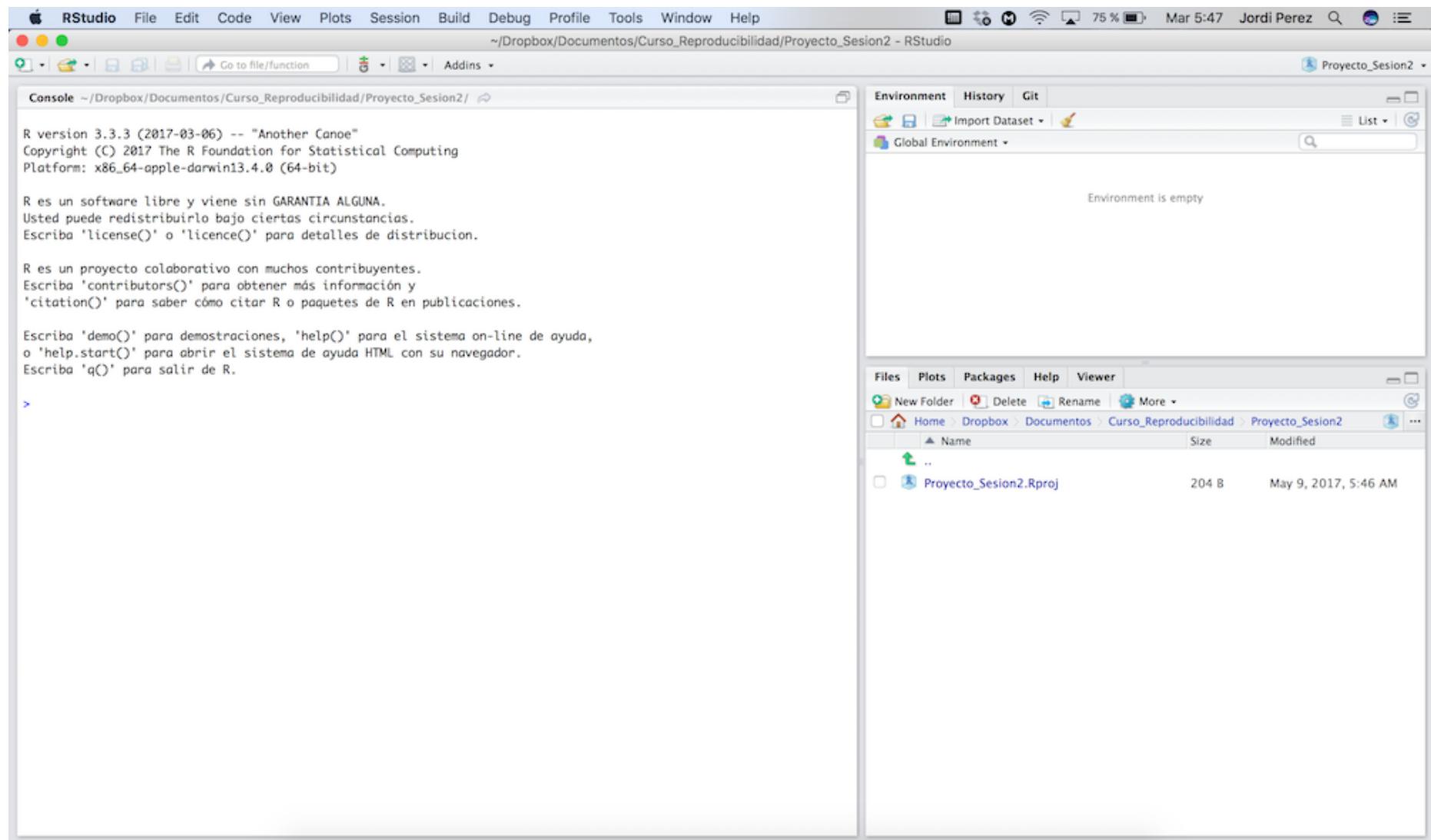
RStudio NewProject



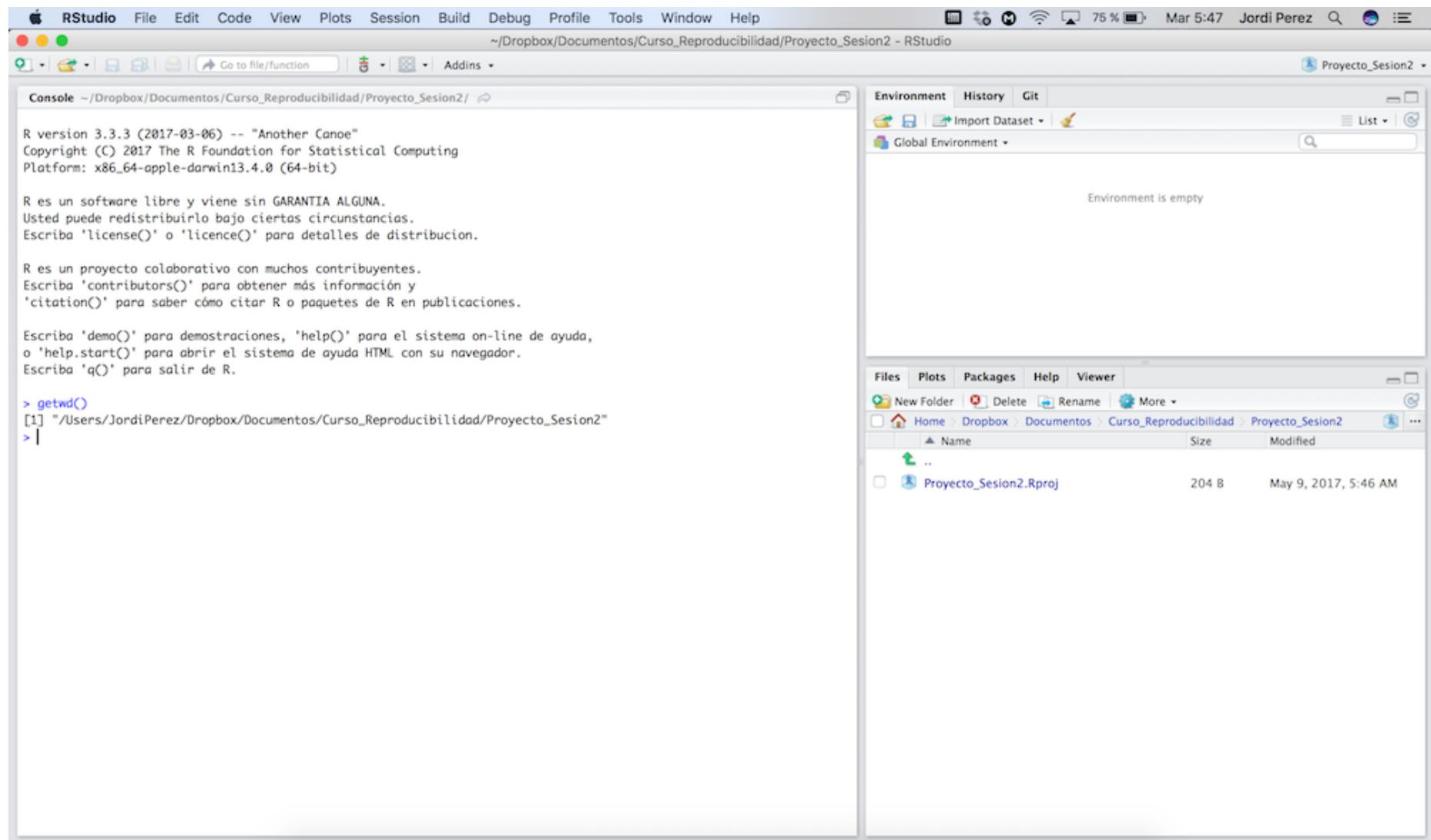
RStudio NewProject



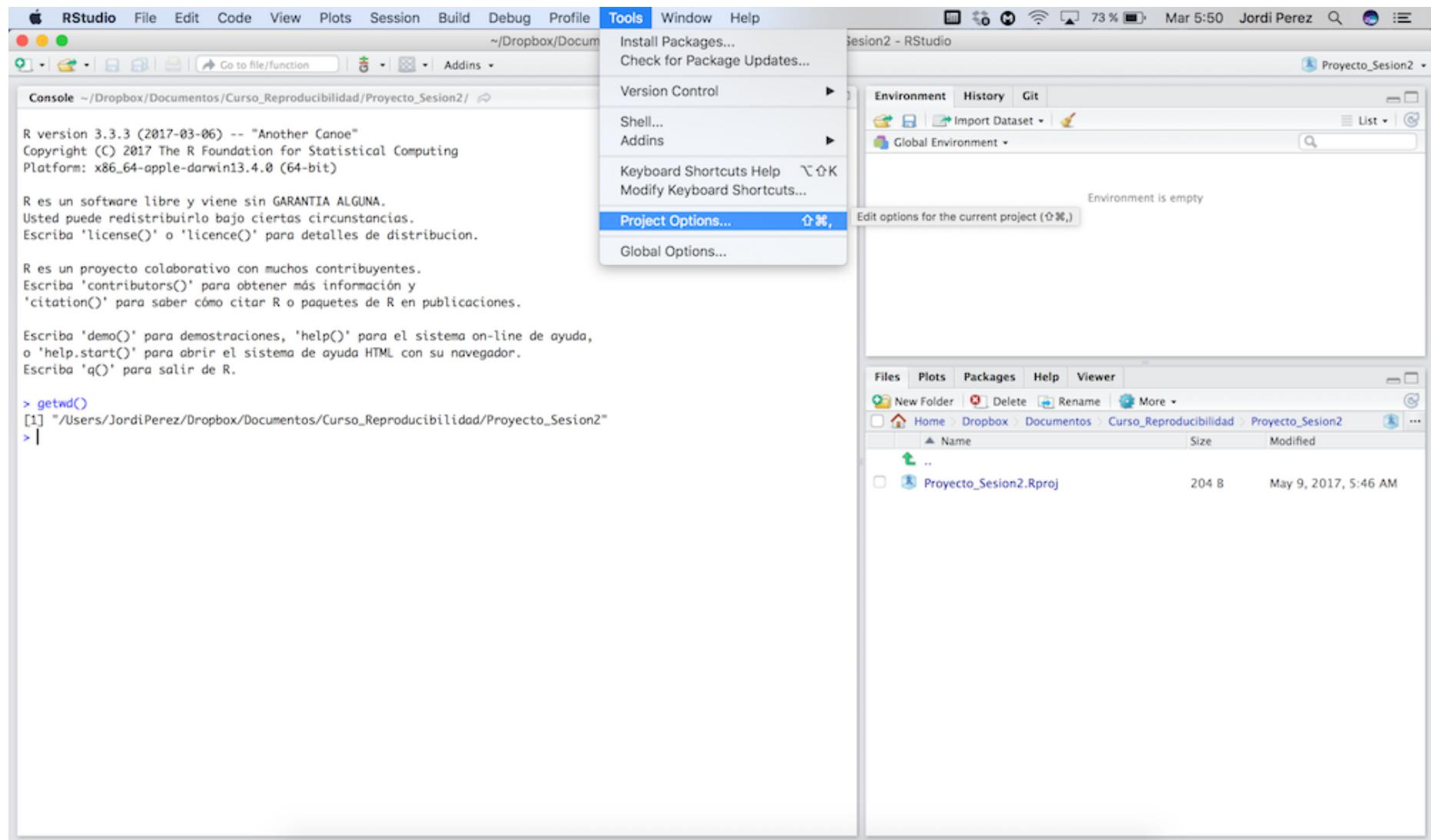
RStudio NewProject



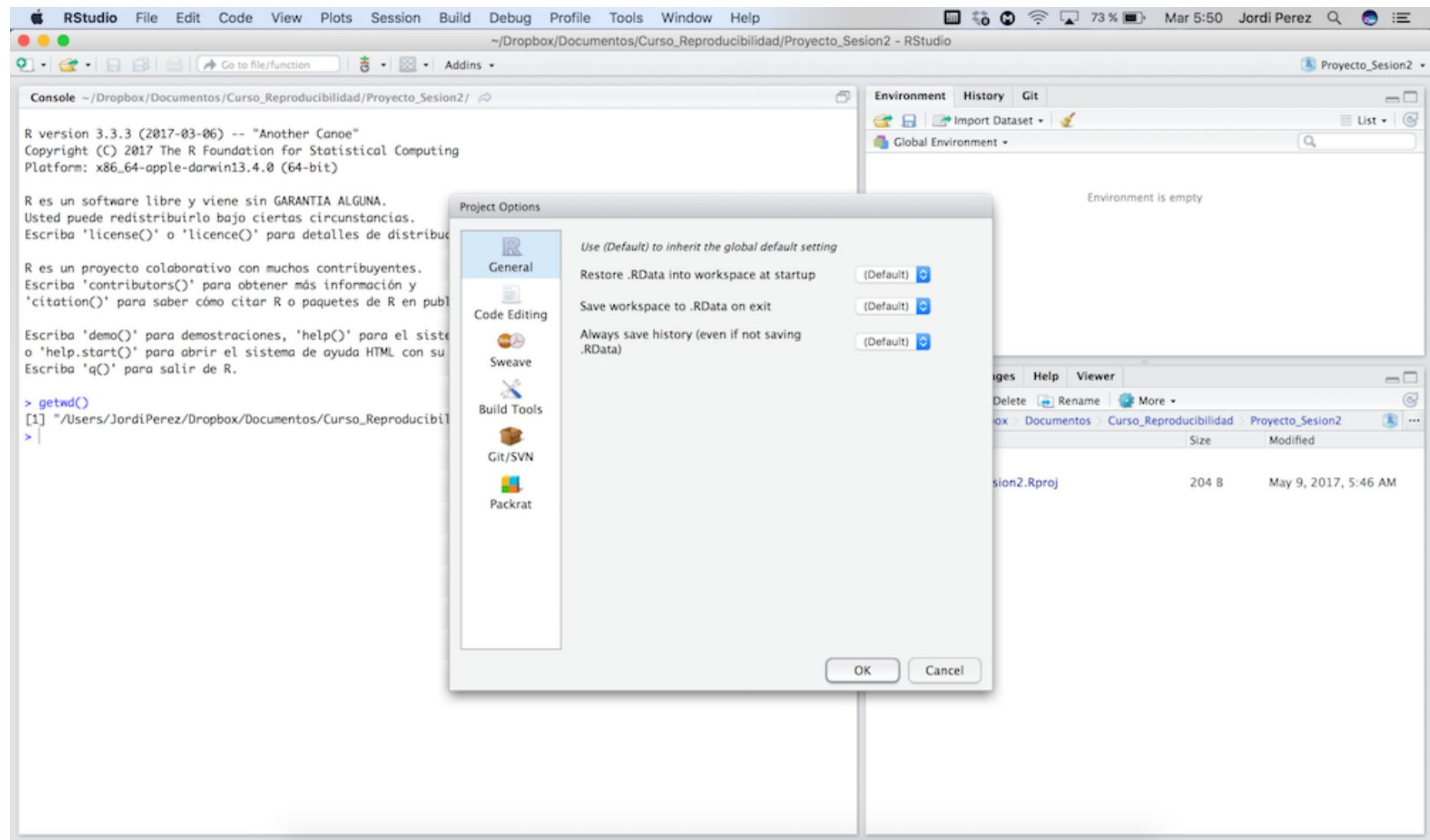
RStudio NewProject



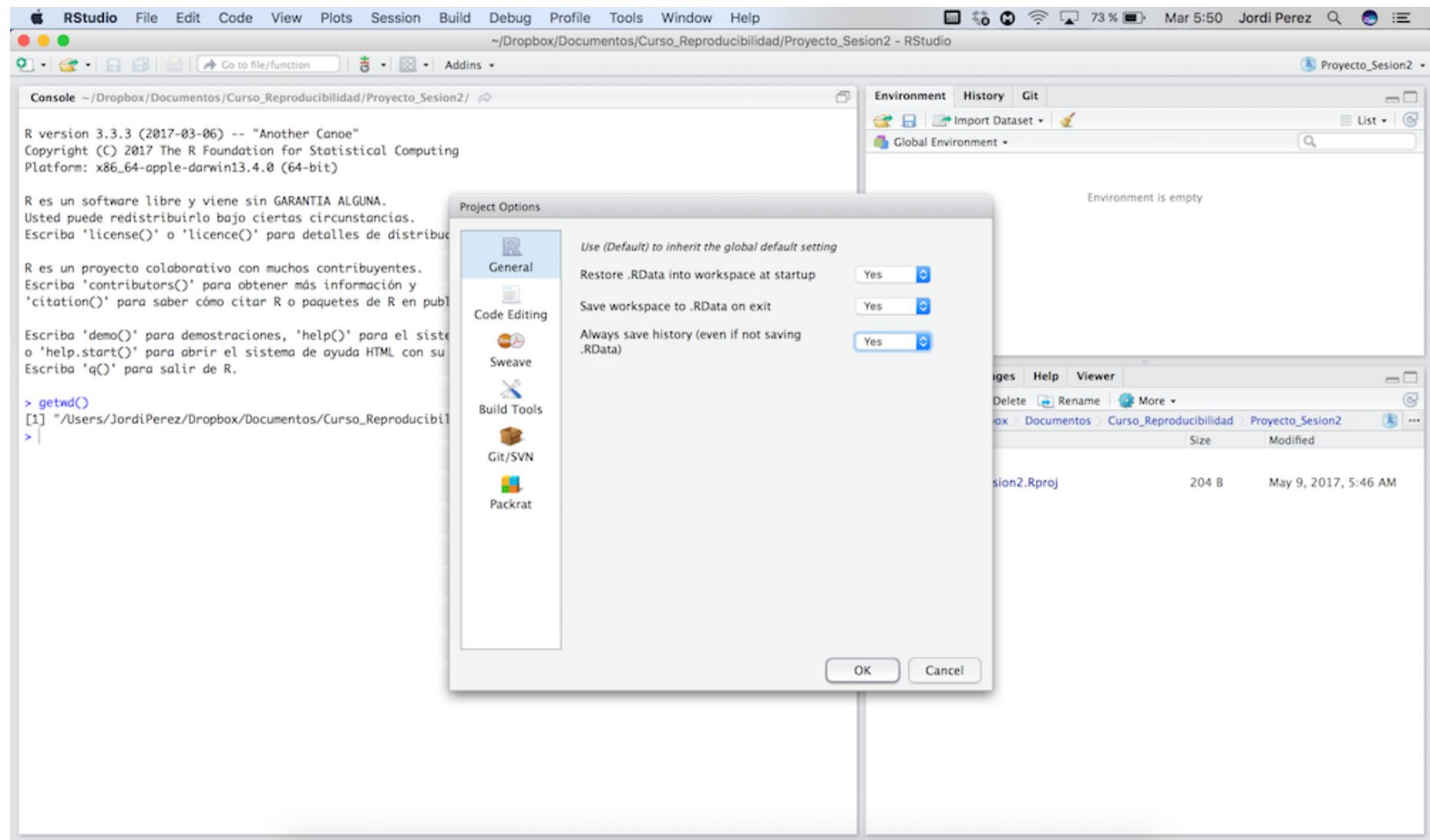
RStudio NewProject



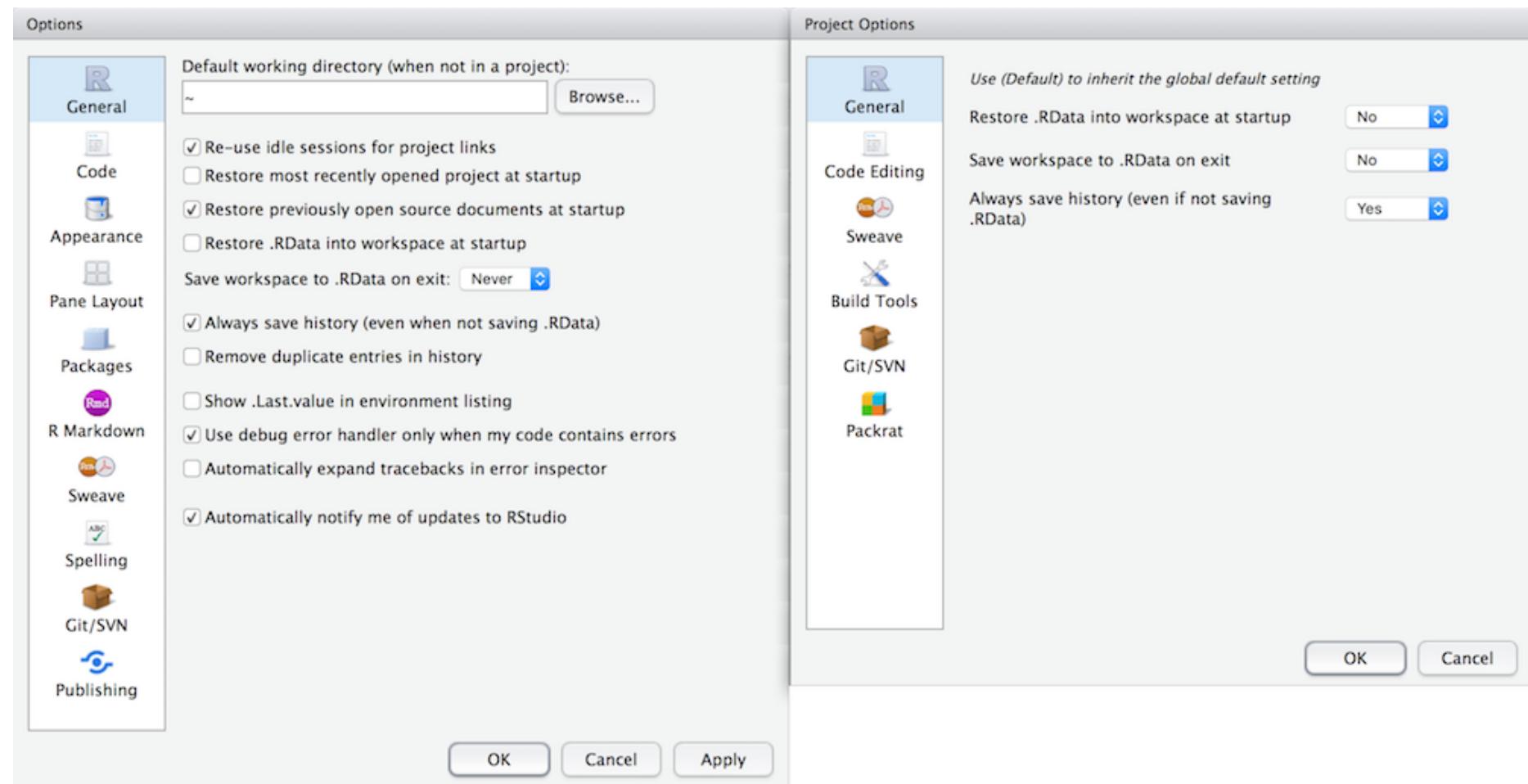
RStudio NewProject



RStudio NewProject



RStudio NewProject



Estructuración del trabajo en R. Packrat.

Packrat

Packrat es un sistema de gestión de dependencias

Packrat trata de resolver problemas con paquetes, dependencias y versiones, haciendo los proyectos:

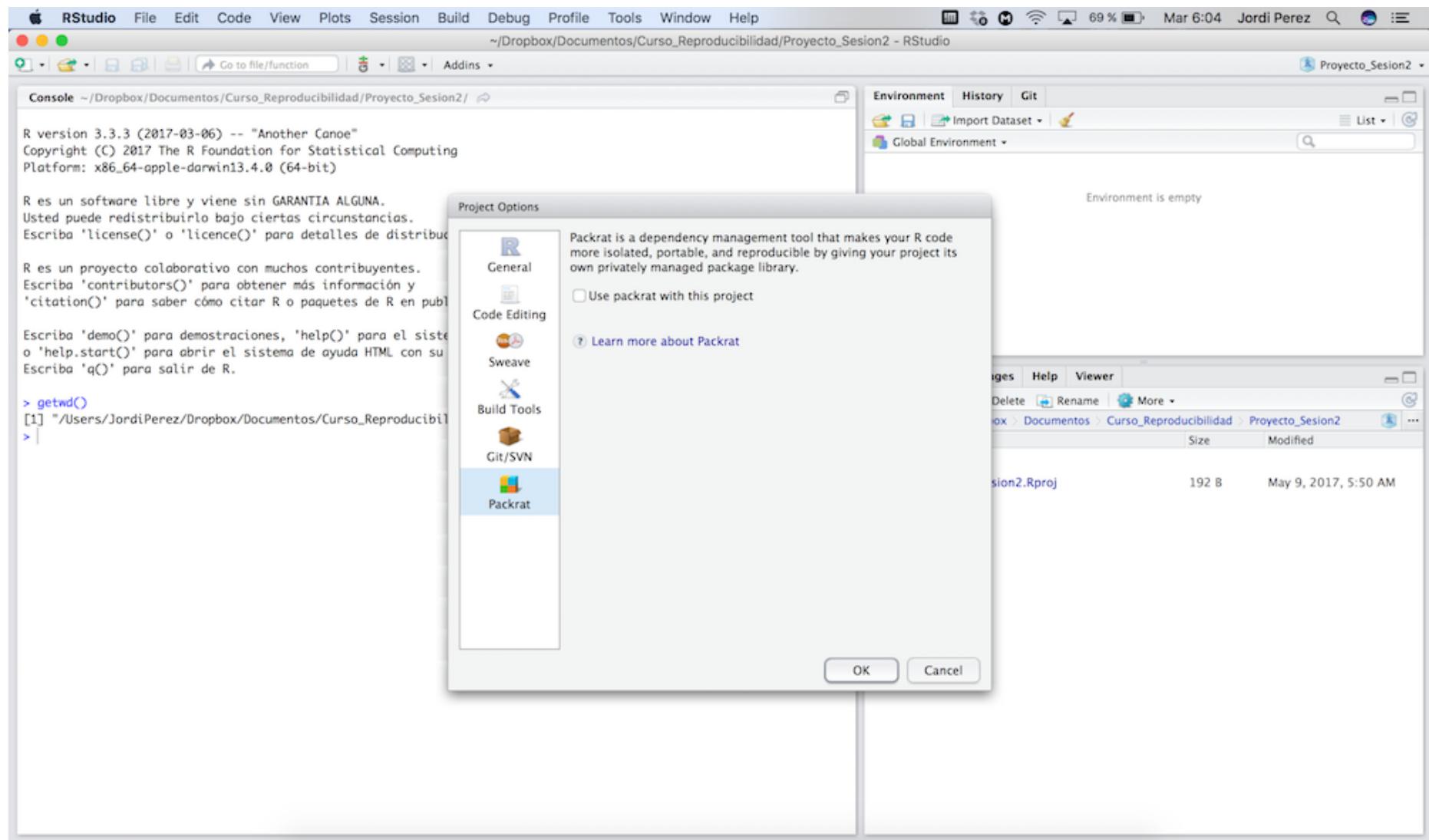
- **Aislados**: crea en cada proyecto su propia librería de paquetes.
- **Portables**: puedes llevarte el proyecto a otro ordenador incluso entre diferentes plataformas. Packrat facilita la instalación de paquetes de los que depende el proyecto.
- **Reproducibles**: guarda las versiones exactas de los paquetes que utiliza el proyecto y asegura que esas versiones exactas serán las que se instalen a donde vayas.

Packrat

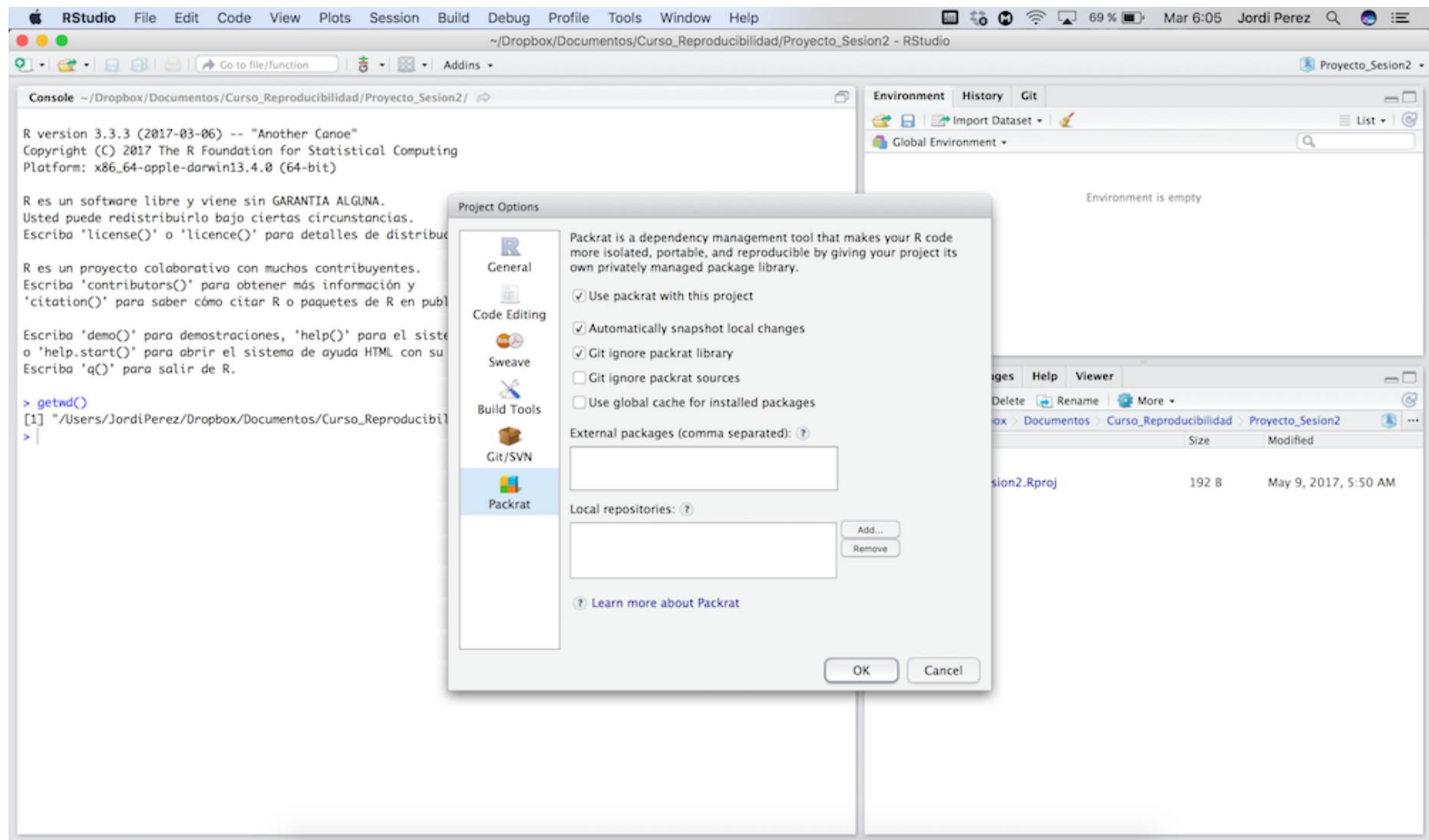
Cómo trabaja Packrat?

- Crea una **cabecera** en la carpeta del proyecto
- Cuando abres una sesión de un proyecto con **packrat**, R mira sólo a esa carpeta. Cada vez que instalas o borras un paquete, packrat hace esos cambios en la **cabecera** (**aislamiento**).
- Desafortunadamente, las **bibliotecas** (como todas) no viajan bien porque sus contenidos son compilados según la arquitectura de la máquina, sistema operativo y la versión de R. Así que packrat hace **copias** del estado de la librería y de cualquier cambio, y se asegura de tener toda la información necesaria para poder restaurar esa biblioteca en otra máquina (**portabilidad y reproducibilidad**).
- Packrat también te informa de paquetes que tienes que cargados y que no son necesarios para correr tu código. Así mantiene limpia tu biblioteca.

Packrat



Packrat

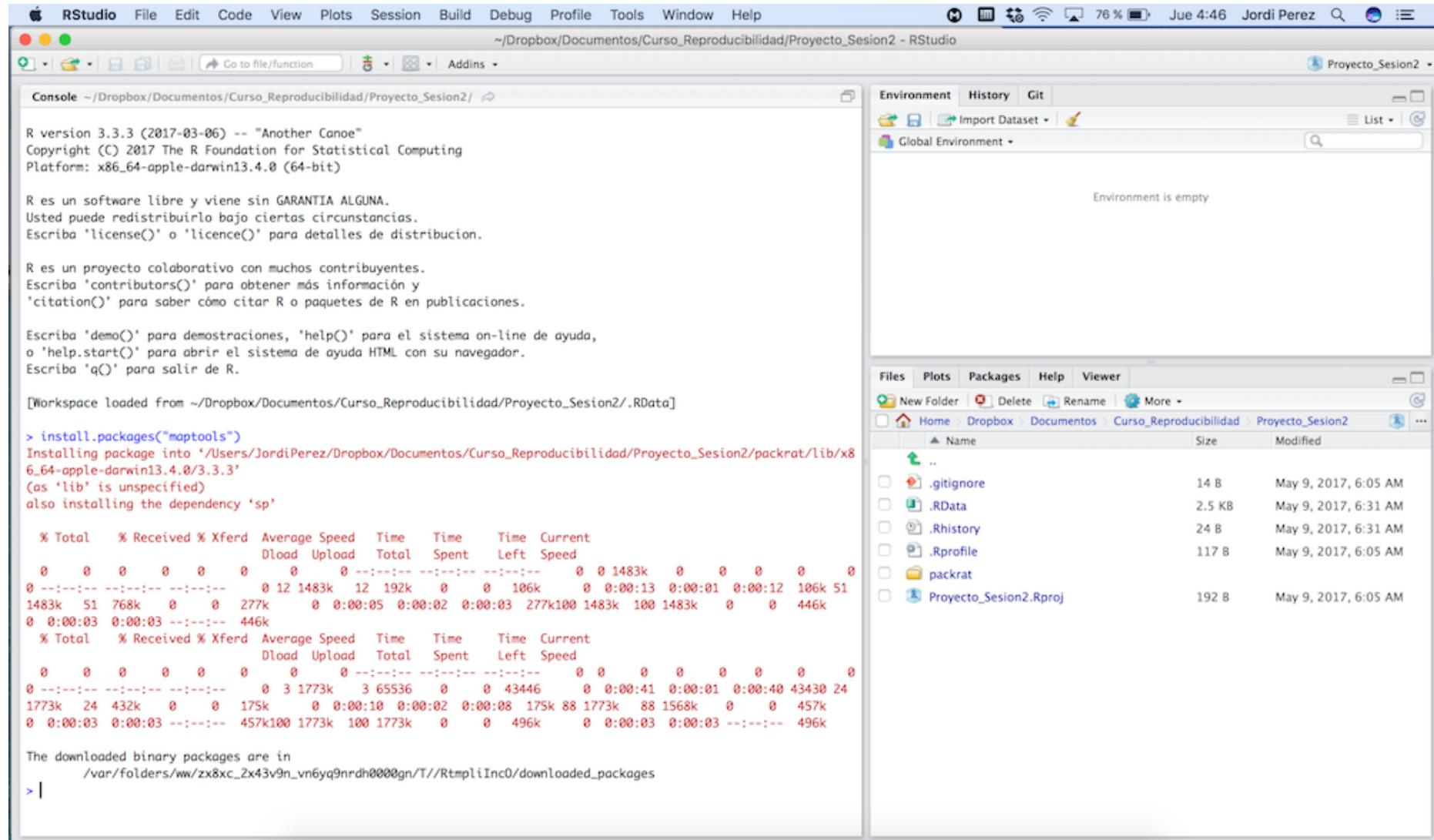


Packrat

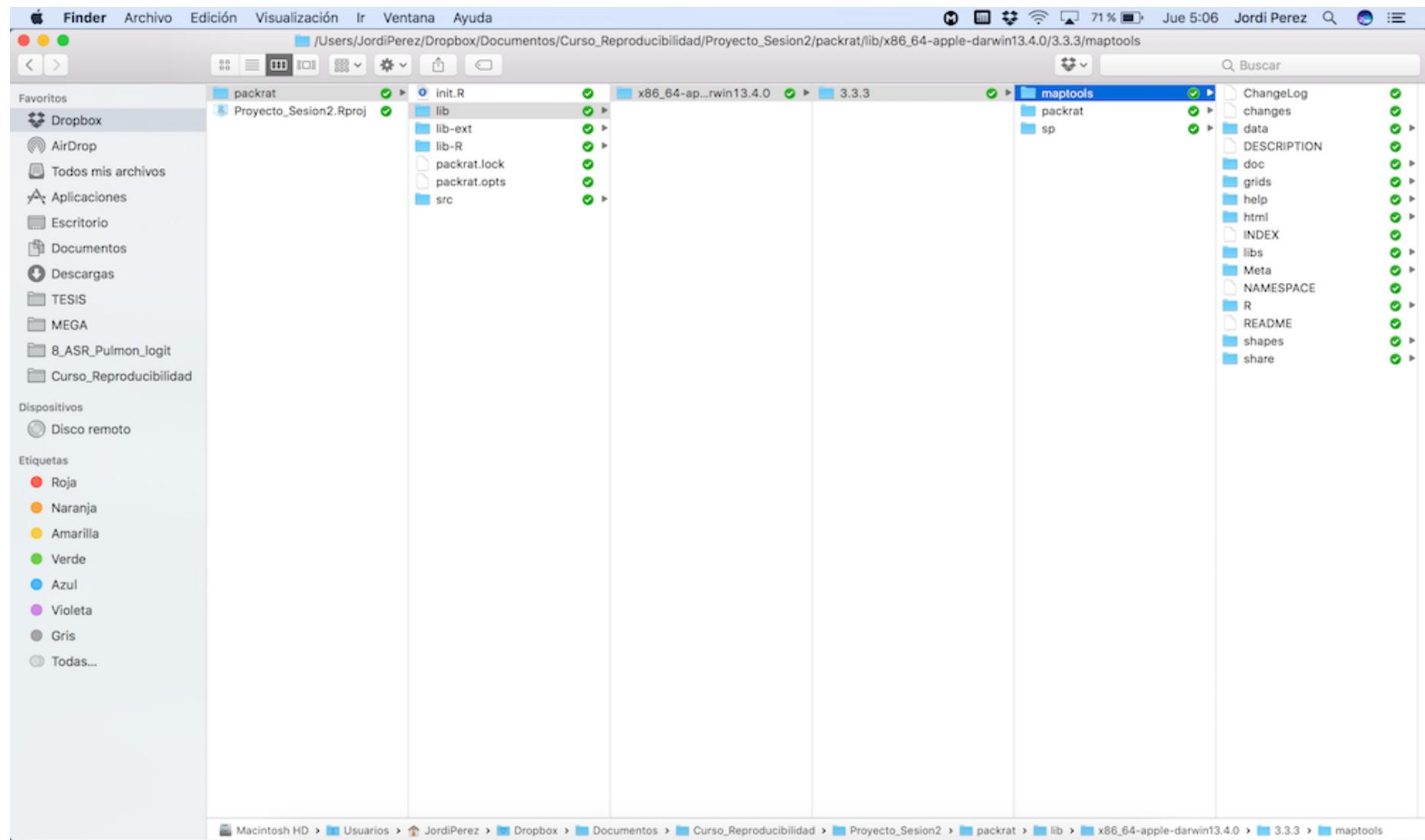
The screenshot shows the RStudio interface with the following details:

- Console:** Displays the R session output. It starts with the standard R startup message, followed by a note about being a free software. It then describes R as a collaborative project and provides help documentation links. The user runs `getwd()` to check the current working directory, which is set to a Dropbox folder. They then run `packrat::init()` to initialize a packrat project in that directory. The console shows the creation of a `packrat` folder and its subfiles: `Projecto_Sesion2.Rproj`, `.Rprofile`, and `.gitignore`. A message indicates that sources were fetched from CRAN. The packrat package is installed, and initialization is complete. The session is then restarted.
- Environment:** Shows the Global Environment pane, which is currently empty.
- Files:** Shows the file tree for the initialized packrat project. The structure is: Home > Dropbox > Documentos > Curso_Reproducibilidad > Projecto_Sesion2. Inside this folder are the files: `Projecto_Sesion2.Rproj` (size 192 B, modified May 9, 2017, 6:05 AM), `.Rprofile` (size 117 B, modified May 9, 2017, 6:05 AM), `.gitignore` (size 14 B, modified May 9, 2017, 6:05 AM), and the folder `packrat`.

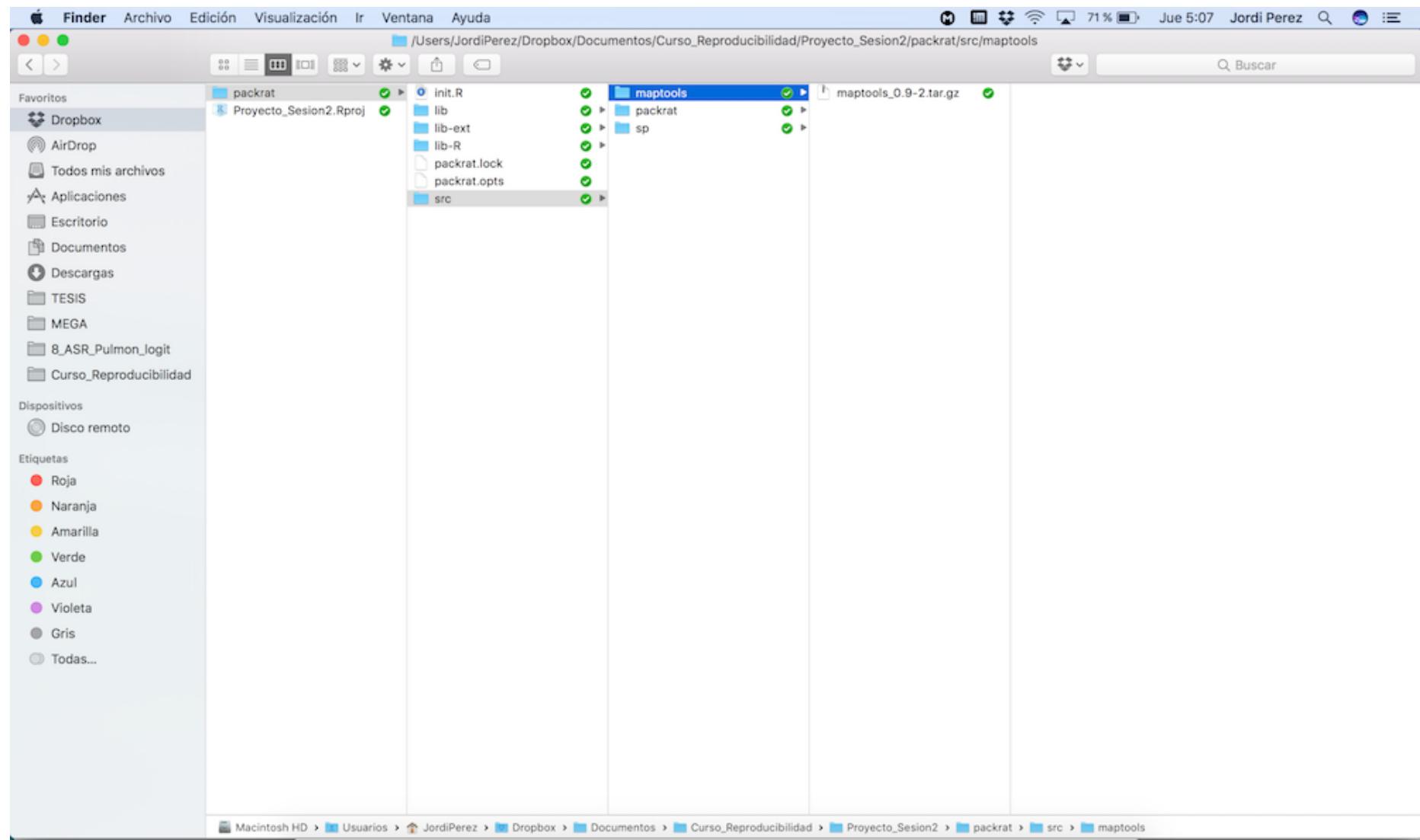
Packrat



Packrat



Packrat



Estructuración del trabajo en R. Organización.

Organización.

El manejo cuidadoso de archivos es crucial para la investigación reproducible.

Hay que tener un plan para organizar, almacenar y compartir los archivos.

Gandrud (2015)

La falta de organización en el espacio de trabajo es una fuente potencial de errores. Por tanto, un factor a evitar en nuestro trabajo.

... la organización lleva su tiempo.

Organización.



Organización.



Organización.

Principios básicos:

- Sé consistente (estructura de directorio, nombres)
- Coloca todo en un único directorio
- Separa código de los datos
- Separa datos brutos de los datos procesados
- Resultados
- Figuras
- Informes

Organización.

- Debemos ser capaces de organizarlo todo (K Broman 2016)
 - Nuestras cosas

0 SmoothASR	✓	8_ASR_Pul...n_logit.Rproj	✓	ASRpaper.aux	✓
1 EVmunicipal	✓	Articulo	✓	ASRpaper.bbl	✓
2 Ciudades	✓	Datos	✓	ASRpaper.log	✓
3 Ciudades aFondo	✓	Figuras	✓	ASRpaper.pdf	✓
4 SEE fijos	✓	R	✓	ASRpaper.synctex.gz	✓
5 SEE aleat	✓	Resultados	✓	ASRpaper.tex	✓
6 ASR	✓			Bibliografia	✓
7 ASR Pulmon logit	✓			Bibliografia...perASR.bib	✓
8_ASR_Pulmon_logit	✓			Enviado	✓
9_EV_IMV	✓			Figuras	✓
20170126...easEV.docx	✓			imsart.cls	✓
BugsParallel	✓			imsart.sty	✓
Jordi-Migue-Paloma	✓			R	✓
Pbugs-working	✓				
Proyecto_IMV	✓				

Organización.

- Debemos ser capaces de organizarlo todo (K Broman 2016)
 - Nuestros proyectos

0 SmoothASR	✓	8_ASR_Pul...n_logit.Rproj	✓	ASRpaper.aux	✓
1 EVmunicipal	✓	Articulo	✓	ASRpaper.bbl	✓
2 Ciudades	✓	Datos	✓	ASRpaper.log	✓
3 Ciudades aFondo	✓	Figuras	✓	ASRpaper.pdf	✓
4 SEE fijos	✓	R	✓	ASRpaper.synctex.gz	✓
5 SEE aleat	✓	Resultados	✓	ASRpaper.tex	✓
6 ASR	✓			Bibliografia	✓
7 ASR Pulmon logit	✓			Bibliografia...perASR.bib	✓
8_ASR_Pulmon_logit	✓			Enviado	✓
9_EV_IMV	✓			Figuras	✓
20170126...easEV.docx	✓			imsart.cls	✓
BugsParallel	✓			imsart.sty	✓
Jordi-Migue-Paloma	✓			R	✓
Pbugs-working	✓				
Proyecto_IMV	✓				

Organización.

- Debemos ser capaces de organizarlo todo (K Broman 2016)
 - Nuestros artículos

0 SmoothASR	✓	8_ASR_Pul...n_logit.Rproj	✓	ASRpaper.aux	✓
1 EVmunicipal	✓	Articulo	✓	ASRpaper.bbl	✓
2 Ciudades	✓	Datos	✓	ASRpaper.log	✓
3 Ciudades aFondo	✓	Figuras	✓	ASRpaper.pdf	✓
4 SEE fijos	✓	R	✓	ASRpaper.synctex.gz	✓
5 SEE aleat	✓	Resultados	✓	ASRpaper.tex	✓
6 ASR	✓			Bibliografía	✓
7 ASR Pulmon logit	✓			Bibliografia...perASR.bib	✓
8_ASR_Pulmon_logit	✓			Enviado	✓
9_EV_IMV	✓			Figuras	✓
20170126...easEV.docx	✓			imsart.cls	✓
BugsParallel	✓			imsart.sty	✓
Jordi-Migue-Paloma	✓			R	✓
Pbugs-working	✓				
Proyecto_IMV	✓				

Organización.

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS *THE* CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13

20130227 2013.02.27 27.02.13 27-02-13

27.2.13 2013.II.27. 27½-13 2013.158904109

MMXIII-II-XXVII MMXIII $\frac{LVII}{CCCLXV}$ 1330300800

$((3+3)\times(111+1)-1)\times3/3-1/3^3$ 2013 Mississ 10/11011/1101 02/27/20/13 01237 22713



Organización.

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS **THE** CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13
20130227 2013.02.27 27.02.13 27-02-13
27.2.13 2013.II.27. 2⁷/₂-13 2013.158904109
MMXIII-II-XXVII MMXIII^{LVII}/_{CCCLXV} 1330300800
 $((3+3)\times(111+1)-1)\times3/3-1/3^3$ 2013.158904109
10/11011/1101 02/27/20/13 01237 22/13 Mississipi

English ▾

ISO International Organization for Standardization

Great things happen when the world agrees



Standards > Popular standards > ISO 8601 Date and time format

Date and time format - ISO 8601

What can ISO 8601 do for me?

ISO 8601 can be used by anyone who wants to use a standardized way of presenting dates and times. It helps cut out the uncertainty and confusion when communicating internationally.

Organización.

- Código de este tipo puede crear la estructura de directorios adecuada para ejecutar la aplicación.

```
if(!dir.exists("datos")){
    dir.create("datos")
}
```

- ... con estructura jerárquica:

```
dir.create("datos/brutos")
dir.create("datos/procesados")
```

Organización.

The screenshot shows the RStudio interface with the following components:

- Script Editor:** Displays the following R code:

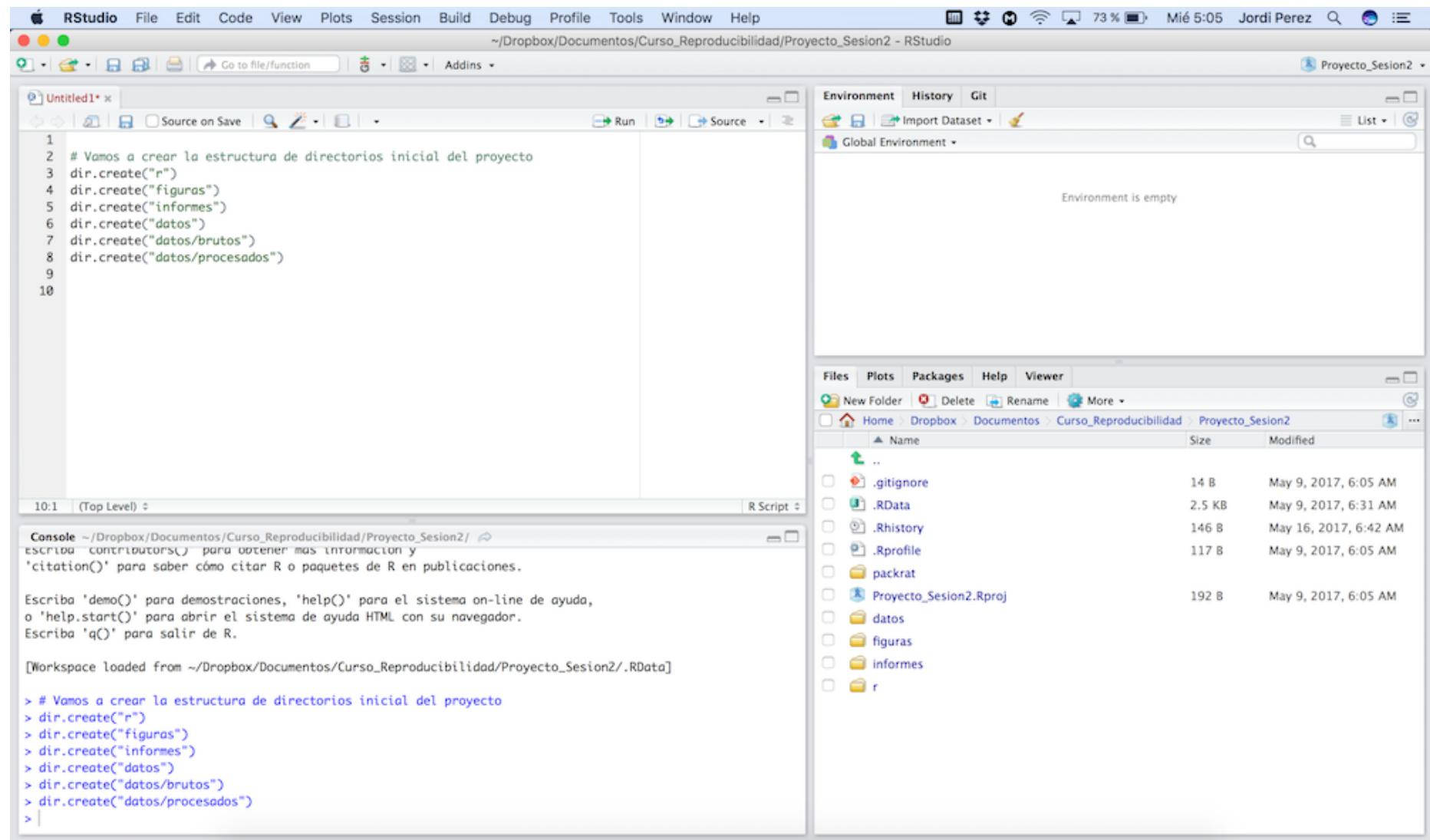
```
1 # Vamos a crear la estructura de directorios inicial del proyecto
2 dir.create("r")
3 dir.create("figuras")
4 dir.create("informes")
5 dir.create("datos")
6 dir.create("datos/brutos")
7 dir.create("datos/procesados")
```

- Environment:** Shows "Environment is empty".
- File Browser:** Shows the directory structure:

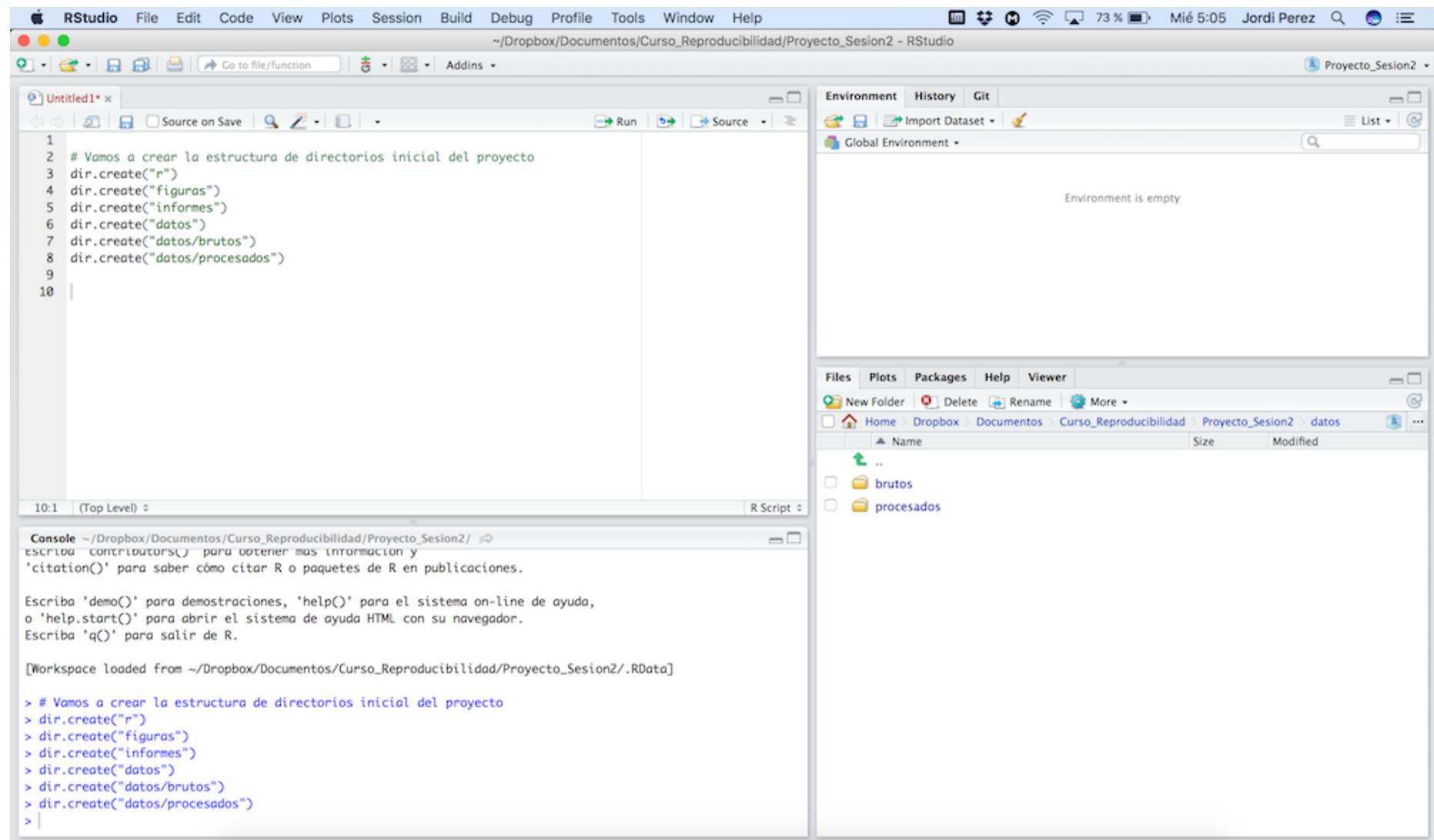
Name	Size	Modified
.gitignore	14 B	May 9, 2017, 6:05 AM
.RData	2.5 KB	May 9, 2017, 6:31 AM
.Rhistory	146 B	May 16, 2017, 6:42 AM
.Rprofile	117 B	May 9, 2017, 6:05 AM
packrat		
Projeto_Sesion2.Rproj	192 B	May 9, 2017, 6:05 AM

- Console:** Displays the R startup message and workspace loading information.

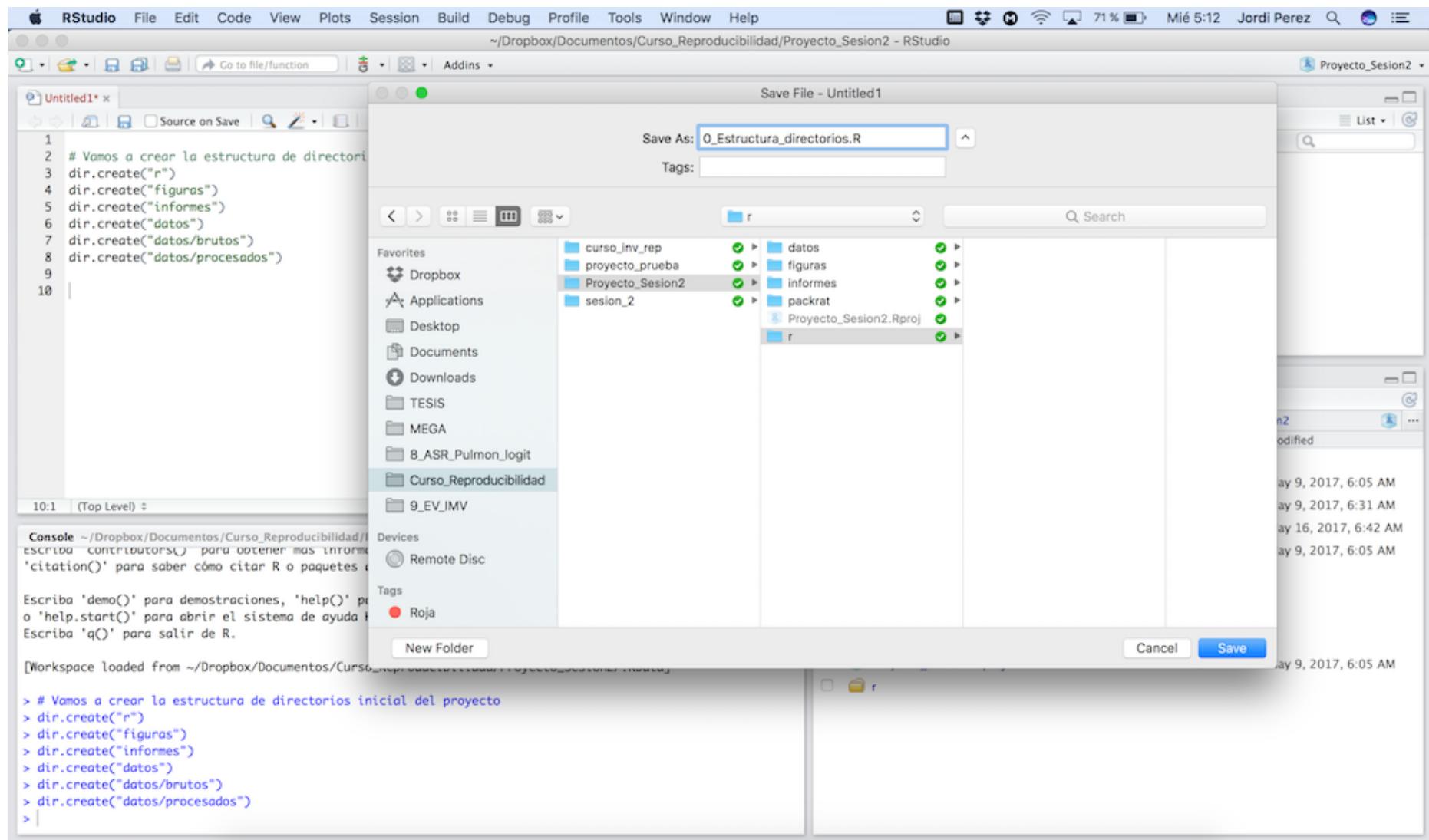
Organización.



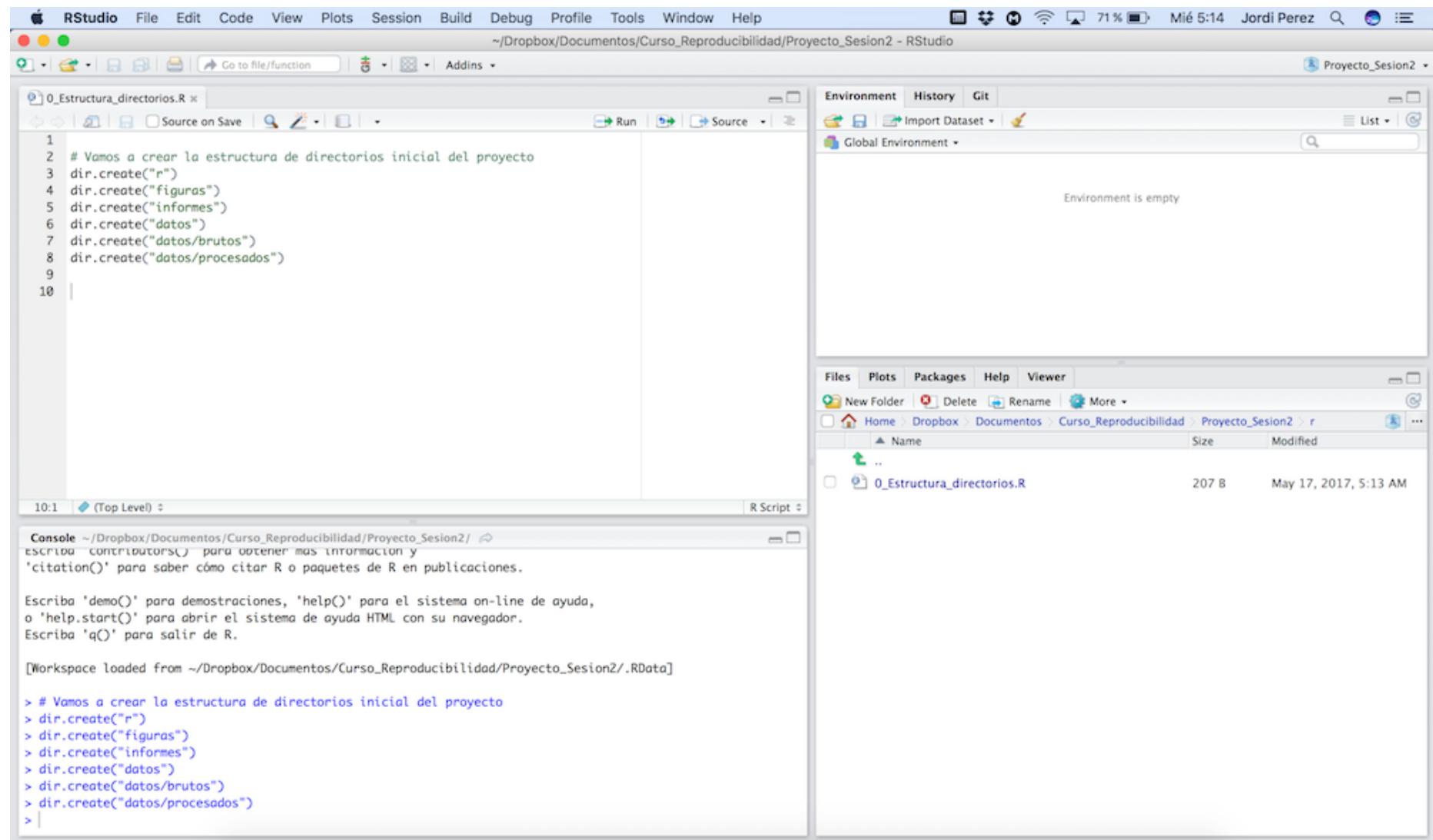
Organización.



Organización.



Organización.



Estructuración del trabajo en R. Rutas relativas.

Rutas relativas

Existen **dos estilos de rutas**, Mac/Linux y Windows, cuyas **diferencias** básicamente en:

- Mac y Linux utilizan (p.e. datos/datos_brutos) y Windows utiliza (p.e. datos\datos_brutos).

R puede trabajar con los dos pero, desafortunadamente, tiene un significado especial para R y hay que duplicarlos (p.e. datos\\datos_brutos).

Recomendación: siempre .

Grolemund and Wickham (2016)

Rutas relativas

- Las **rutas absolutas** (rutas que apuntan al mismo lugar independientemente del directorio de trabajo) se ven diferentes. En Windows empiezan con la letra del disco (p.e. C:) y en Mac/Linux empiezan con una (p.e. /Users/JordiPerez).

Recomendación: deberíamos utilizar rutas absolutas porque dificulta el intercambio de proyectos.

- La última diferencia menor es al lugar que señala ~. En Mac/Linux apunta a tu . En Windows, apunta a .

Grolemund and Wickham (2016)

Rutas relativas.

Si se distribuye código con directorios referenciados de forma relativa
funcionará siempre.

.. sube un nivel

. directorio actual

~ directorio home

```
read.table("../datos/brutos/datos_perinatal_2015.txt", sep=";", header=T)
setwd("./informes/rmarkdown")
setwd("~/Dropbox")
```

Jaffe and Leek (2016)

Rutas relativas.

Funciones útiles:

```
getwd()
```

```
# [1] "/home/carlos/Documentos/GitHub/curso_inv_rep/sesion_02_rstudio"
```

```
path.expand("~/")
```

```
# [1] "/home/carlos"
```

```
file.path("~/", "Dropbox")
```

```
# [1] "~/Dropbox"
```

Rutas relativas.

Funciones útiles:

```
path.expand( file.path( "~", "Dropbox") )
```

```
# [1] "/home/carlos/Dropbox"
```

En el servidor Linux, el resultado es:

```
# [1] "/home/jordi/Dropbox"
```

Y en Windows:

```
# [1] "C:/Users/Jordi/Documentos/Dropbox"
```

Guía de estilo de R

Guía de estilo.

Un buen estilo de escribir código es como utilizar correctamente los signos de puntuación. Puedes escribir sin ellos pero facilita la compresión.

Wickham (2014)

Cuando se colabora en un proyecto, un código tiene muchos autores y lectores. En estos casos es muy útil tener un estilo claro y bien definido, y a ser posible, común.

Guía de estilo. Código claro.

Guía de estilo. Código claro.

Principios básicos:

- Código que funcione (sin errores, la eficiencia es secundario)
- Inteligible (código simple y claro)
- Reutilizable (modular)
- Reproducible (re-runnable)
- No tengas prisa. Piensa antes de escribir (se minimizan errores y re-escrituras)
- Aprende de otros (R, R packages)

(Karl Broman 2016; Wilson et al. 2014)

Guía de estilo. Código claro.

Funciones

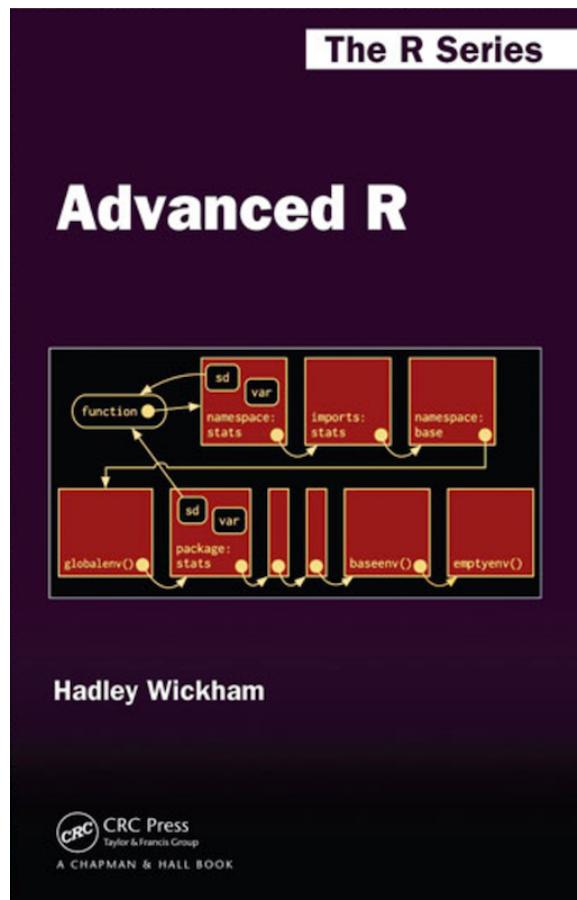
- Crea funciones y llámalas repetidamente
- Es más sencillo de mantener (!)
- Mira en librerías y paquetes (!)
- Rompe el código en pequeñas funciones
- Parte una gran función en pequeñas funciones
- Si escribes el mismo código más de una vez, haz una función
- Si una línea o bloque de código es complicado, haz una función
- Puedes crearte tu propio R package! Con tus funciones!

Guía de estilo. Código claro.

Comentarios

- Incluye un breve comentario al principio del script
- Comenta los puntos críticos/difíciles y las secciones
- No contradigas el código (!)
- Comenta el código cuando lo escribas (o antes)
- Planea gastar 1/4 del tiempo comentando

Guía de estilo.



Propuesta extraída de
Wickham.

de Hadley

Disponible en Amazon (por 78.5€) o libre en su web (Wickham 2014)

La propuesta está basada en la Guía de estilo de R elaborada por usuarios de Google (Google 2014)



Chief Scientist at RStudio, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University.

Guía de estilo. Nombres.

Nombres de **archivos**

- Utiliza nombres útiles
- Define nombres descriptivos pero concisos **NO ES FÁCIL!**
- Evita los acentos.
- Evita nombres del estilo:
`temp1`
`temp2`
`fv`
`nms`

Guía de estilo. Nombres.

Nombres de **archivos**

- Sustituye espacios por guiones.

funciones-proyecto.R
modelos-ajustados.R

cosas.r

informe.r

- Si hay que ejecutar los archivos secuencialmente, ordénalos.

0-datos.R.

1-descriptivo.R

2-analisis.R

Guía de estilo. Nombres.

Nombres de **objetos**

- Utiliza nombres útiles
- Define nombres descriptivos pero concisos **NO ES FÁCIL!**
- Funciones con verbos, objetos con nombres
- En minúsculas y separados por guión bajo _
periodo_1
periodo_uno

primer_periodo_del_ensayo

PeriodoUno

periodouno

ppde

Guía de estilo. Nombres.

Nombres de **objetos**

- Evita la confusión.
- No utilices nombres de variables o funciones

```
T <- FALSE  
c <- 10  
mean <- function(x) sum(x)
```

- No utilices a la vez: **total** y **totals**, **Mat** y **mat**, **g** y **gg**

Guía de estilo. Sintaxis.

Espaciado

Dejar siempre un espacio alrededor de todos los operadores (= + - <- ...) y después de una coma (y nunca, antes).

```
# Bien
media <- mean(feet / 12 + inches, na.rm = TRUE)

# Mal
media<-mean(feet/12+inches,na.rm=TRUE)
```

Con una excepción, no dejar espacios alrededor de : y ::

```
# Bien
x <- 1:10
maptools::kmlPolygon()

# Mal
x <- 1 : 10
maptools :: kmlPolygon()
```

Guía de estilo. Sintaxis.

Espaciado

Dejar un espacio a la izquierda de los paréntesis excepto en llamadas de funciones

```
# Bien
if (debug) do(x)
plot(x, y)

# Mal
if(debug)do(x)
plot (x, y)
```

Espacios extra si ayuda a alinear signos = o asignaciones <-

```
list(
  total = a + b + c,
  mean  = (a + b + c) / n
)
```

Guía de estilo. Sintaxis.

Espaciado

No dejar espacios alrededor de código entre paréntesis o corchetes (a menos que haya una coma)

```
# Bien
if (debug) do(x)
diamonds[5, ]

# Mal
if ( debug ) do(x) # No necesita espacios alrededor de debug
x[1,] # Necesita un espacio después de la coma
x[1 ,] # El espacio va después de la coma, no antes.
```

Guía de estilo. Sintaxis.

Llaves

Una llave nunca se abre en su propia línea y debe ir siempre seguida de una nueva línea. Una llave se cierra siempre en una nueva línea, a menos que vaya seguida de un `else`. Siempre sangrar el código entre llaves.

```
# Bien
if (y == 0) {
    log(x)
} else {
    y ^ x
}

# Mal
if (y == 0) {
    log(x)
}
else {
    y ^ x
}
```

Guía de estilo. Sintaxis.

Logitud de las líneas

Limitar el código a 80 caracteres por línea. Esta anchura se adapta cómodamente a una página impresa con una fuente de tamaño razonable.

Guía de estilo. Sintaxis.

Sangría

La sangría se limita a dos espacios. Nunca utilizar tabuladores.

La única excepción es en la definición de una función cuando ocupa varias líneas. En ese caso debe añadirse espacios la segunda línea hasta donde empieza la definición de argumentos:

```
funcion_con_nombre_largo <- function(a = "un primer argumento largo",
                                         b = "otro argumento largo",
                                         c = "y otro más")
  # el código en general sangrado 2 espacios.
}
```

Guía de estilo. Sintaxis.

Herramientas para compartir y acceder a información

Compartir y acceder a información.

read.table()

Lee un fichero en formato tabla y crea un data.frame a partir de él. Los casos corresponden a las filas y las variables a las columnas.

```
read.table( file, # filename  
header = FALSE, # are there column names?  
sep = "", # what separates columns?  
as.is = !stringsAsFactors, # do you want character strings as factors or characters?  
quote = "\"\"", dec = ".", row.names, col.names,  
na.strings = "NA", nrow = -1,  
skip = 0, check.names = TRUE, fill = !blank.lines.skip,  
strip.white = FALSE, blank.lines.skip = TRUE, comment.char = "#",  
stringsAsFactors = default.stringsAsFactors())
```

- Por ejemplo:

```
read.table("file.txt", header = TRUE, sep="\t", as.is=TRUE)`
```

Notemos que el filename puede ser una ruta a un fichero en una web (e.g. 'www.algunaurl.com/tabla1.txt') (Jaffe 2016)

Compartir y acceder a información

readr

Read Rectangular Text Data

xlsx, readxl

Read/write xls/xlsx. Import excel files into R

foreign

Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...

Compartir y acceder a información

haven

Import and Export 'SPSS', 'Stata' and 'SAS' Files

RODBC

ODBC Database Access

pxR

PC-Axis with R. Provides a set of functions for reading and writing PCAxis files, used by different statistical organizations around the globe for data dissemination.

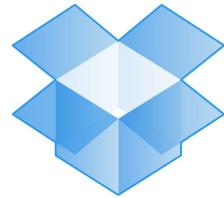
Compartir y acceder a información

download.file

Descargas de datos de la web. Ejemplo: Banco de España. Series temporales cuadros de Cuentas Financieras:

```
fileUrl <- "http://www.bde.es/webbde/es/estadis/ccff/csvs/cf.zip"
download.file(fileUrl,
destfile="./data/cf.zip")
list.files("./data")
dateDownloaded <- date()
dateDownloaded
```

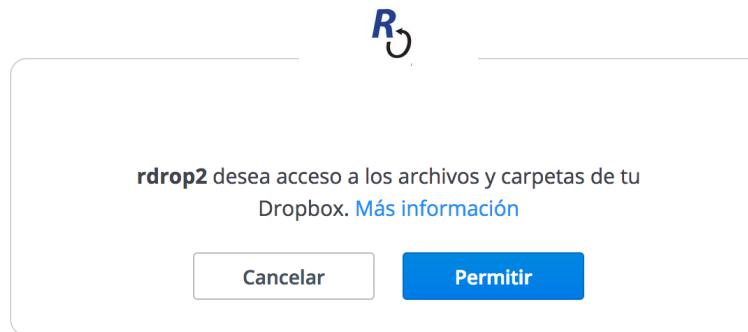
Dropbox



Dropbox

El package `rdrop2` de Karthik Ram reemplaza a `rDrop`.

En primer lugar, debes autenticarte:



tras permitir:

Authentication complete. Please close this page and return to R.

Dropbox

Tras identificarse, las funciones deseables:

- Listar archivos y directorios.

```
> drop_dir('Documentos')
No encoding supplied: defaulting to UTF-8.
# A tibble: 11 × 5
*   path      is_dir    root bytes modified
  <chr>     <lgl>    <chr> <int>   <chr>
1 /Documentos/BEINSTITUTE TRUE  dropbox     0 Fri, 04 Mar 2016 12:18:36 +0000
2 /Documentos/Bibliografía TRUE  dropbox     0 Wed, 20 Jul 2016 11:52:25 +0000
3 /Documentos/Curso BayesianSurvivalAnalysis-UPC TRUE  dropbox     0 Fri, 10 Jun 2016 11:16:58 +0000
4           /Documentos/Curso HMM TRUE  dropbox     0 Wed, 30 Nov 2016 14:15:55 +0000
5           /Documentos/Curso INLA-SPDE TRUE  dropbox     0 Fri, 10 Jun 2016 11:17:16 +0000
6           /Documentos/Curso_Reproducibilidad TRUE  dropbox     0 Thu, 16 Feb 2017 05:21:57 +0000
7           /Documentos/Curso_Supervivencia_EASP TRUE  dropbox     0 Mon, 27 Mar 2017 08:44:18 +0000
8           /Documentos/DOCTORADO TRUE  dropbox     0 Thu, 18 Feb 2016 16:48:23 +0000
9           /Documentos/EPI TRUE  dropbox     0 Mon, 27 Apr 2015 21:17:07 +0000
10          /Documentos/PROYECTOS TRUE  dropbox     0 Mon, 11 Apr 2016 08:52:36 +0000
11          /Documentos/TESIS  TRUE  dropbox     0 Wed, 16 Mar 2016 09:04:12 +0000
```

Dropbox

- Crear una carpeta en Dropbox: `drop_create('drop_test')`
- Subir un fichero a Dropbox: `drop_upload('mtcars.csv')`
- Descargar un fichero a Dropbox: `drop_get('mtcars.csv')`
- Borrar un fichero de Dropbox: `drop_delete('mtcars.csv')`
- Mover un fichero en Dropbox: `drop_move("mtcars.csv", "new_folder/mtcars.csv")`
- Copiar un fichero en Dropbox: `drop_copy("new_folder/mtcars.csv", "new_folder2/mtcars.csv")`
- Buscar un fichero en Dropbox: `drop_search('gif')`

googlesheets

Google

Almacenamiento



Drive

Colaboración



Documentos



Hojas de Cálculo



Formularios

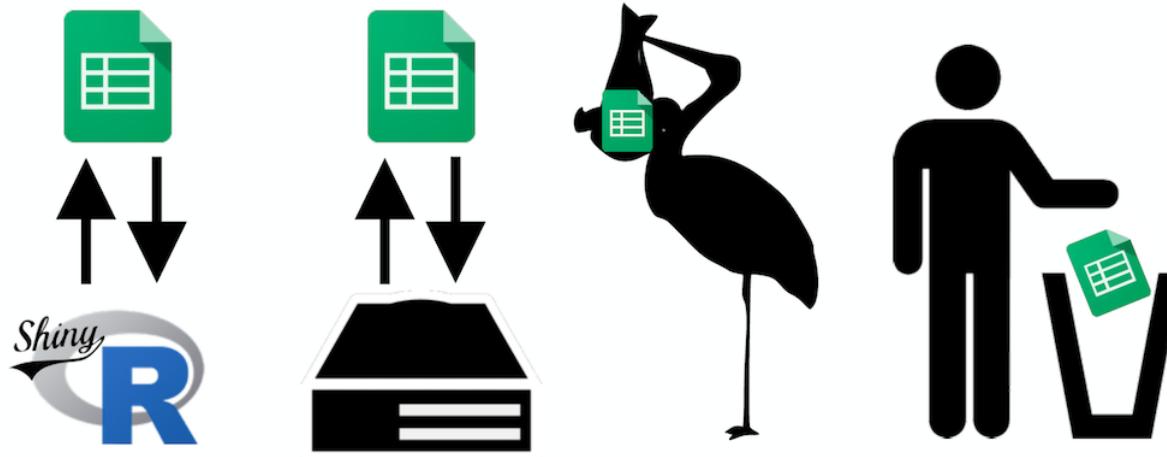


Presentaciones



Sites

googlesheets



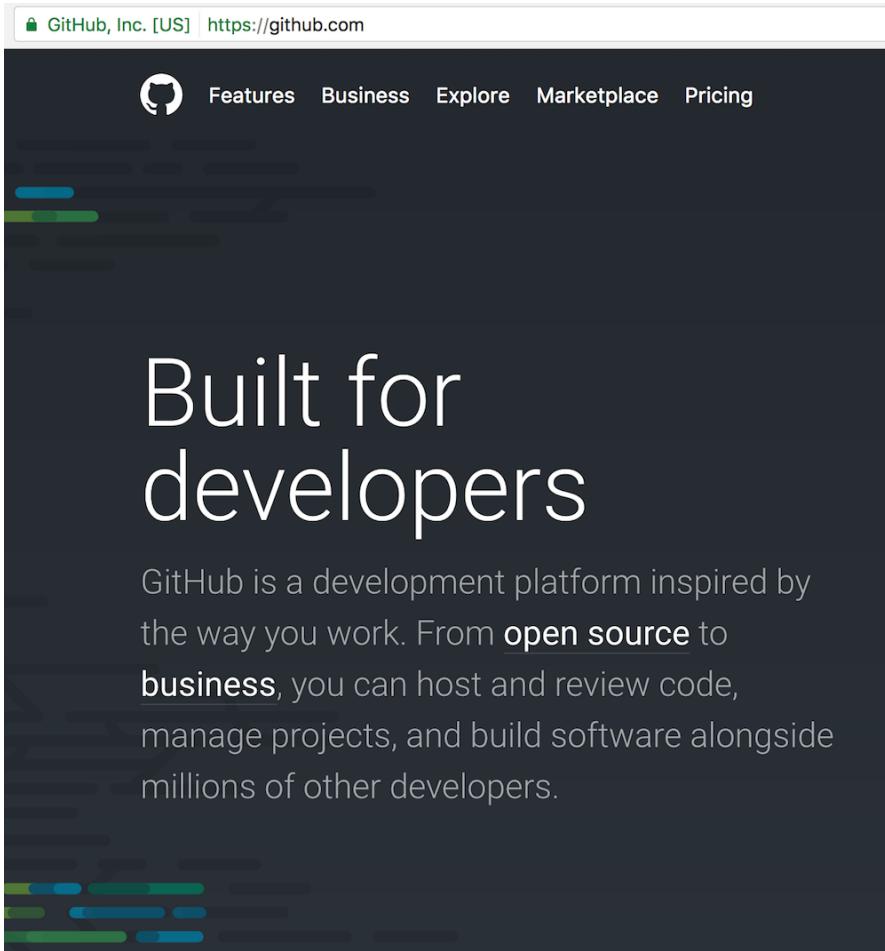
El package `googlesheets` de Jennifer Bryan y Joanna Zhao permite importar y exportar hojas de cálculo desde una app de Shiny o desde nuestro pc, crearlas o eliminarlas (completa o parcialmente).

googlesheets.

Por qué utilizar googlesheets? (Bryan 2015)

- Colaboración: no adjentes, pasa un link!
- Aligerar el almacenamiento de datos en Shiny
- Trabajar con GoogleSheets públicas o privadas desde R
- GoogleSheets tiene funciones incorporadas para importar datos web:
`importxml()`, `importhtml()`, `importfeed()`
- Servicios web: ifttt.com

github



<https://github.com/es/Octocat/blob/master/scrapelimgs.sh>

figshare



figshare es un repositorio abierto donde los usuarios pueden hacer que todos sus resultados de investigación estén disponibles de una manera **citable, compatible y comprobable**.



	Figure Figures, images
	Media Videos, audio
	Dataset Tables, statistics
	Fileset Multiple associated files
	Poster Illustrations, diagrams
	Paper Publication document
	Presentation Slide(s)
	Thesis Essays or dissertations
	Code Scripts, binaries

Se pueden almacenar archivos de todo tipo.



figshare

My data search on figshare Browse Upload Jordi Perez Pan... ▾

My data Projects Collections Activity

Drag files to upload
You can add keywords and other details later
Browse for files

+ Create a new item

0 kB 20 GB search my data...

Actions STATUS TYPE CREATED SIZE



Es seguro | <https://figshare.com/account/home>

datos_practicas.RData 266.73 kB

[Download](#)

needed to publish & get DOI

Title
datos_practicas.RData

Authors
 Jordi Perez Panades
Add co-authors by name, full email or ORCID

Tips
you can still drag more file(s) on the page or [browse](#)

Categories
Statistics

Item type (what's this?)
Dataset

Keyword(s)
 curso_reproducibilidad Add keywords for easy discovery

Description
privacion_valencia
Contiene 16 indicadores sociales para las 598 secciones censales de la ciudad de Valencia. Todas las variables han sido medidas como el porcentaje de población (en cada sección) que pertenece a estos colectivos o que dice observar dichos problemas. Todos estos indicadores podrían ser reflejo de la privación económica de cada una de las secciones censales de la ciudad de Valencia.

B I U X₂ X²



Es seguro | <https://figshare.com/account/home>

References

Funding

Licence ([what's this?](#))

Tips
you can still drag
more file(s) on the page or
[browse](#)

[Preview item \(private\)](#)

This item is a draft (metadata required for publication missing)

[Apply embargo](#)

[Make file\(s\) confidential](#)

[Generate private link](#) [x](#)
<https://figshare.com/s/c66244c233dbc9399db4>
This link can be used by non figshare users also.

DOI Digital Object Identifier

[Delete item](#)

[Cancel](#) [DOI](#) [🔗](#) [🔗](#) [Publish](#) [Publish item](#)

Es seguro | https://figshare.com/articles/datos_practicas_RData/5016575

[Download \(266.73 kB\)](#) [Share](#) [Cite](#) [Embed](#) + Collect (you need to log in first)

datos_practicas.RData

18.05.2017, 05:32 by Jordi Perez Panades

privacion_valencia

Contiene 16 indicadores sociales para las 598 secciones censales de la ciudad de Valencia. Todas las variables han sido medidas como el porcentaje de población (en cada sección) que pertenece a estos colectivos o que dice observar dichos problemas. Todos estos indicadores podrían ser reflejo de la privación económica de cada una de las secciones censales de la ciudad de Valencia.

carto_valencia

Cartografía de la ciudad de Valencia con 598 secciones censales.

0 Comments figshare [Login](#)

[Recommend](#) [Share](#) Sort by Newest

 Start the discussion...

Be the first to comment.

0 views | 0 downloads | 0 citations

CATEGORIES • [Statistics](#)

KEYWORD(S) [curso_reproducibilidad](#)

LICENCE  CC BY

EXPORT [RefWorks](#) [BibTeX](#) [Ref. manager](#) [Mendeley](#) [Endnote](#) [DataCite](#) [NLM](#) [DC](#)



Es seguro | https://figshare.com/articles/datos_practicas_RData/5016575

[Download \(266.73 kB\)](#) [Share](#) [Cite](#) [Embed](#) + Collect (you need to log in first)

Perez Panades, Jordi (2017): datos_practicas.RData. figshare.
<https://doi.org/10.6084/m9.figshare.5016575.v1>
Retrieved: 03 49, May 18, 2017 (GMT)

Place your mouse over the citation text to select it

datos_practicas.RData

18.05.2017, 05:32 by [Jordi Perez Panades](#)

privacion_valencia

Contiene 16 indicadores sociales para las 598 secciones censales de la ciudad de Valencia. Todas las variables han sido medidas como el porcentaje de población (en cada sección) que pertenece a estos colectivos o que dice observar dichos problemas. Todos estos indicadores podrían ser reflejo de la privación económica de cada una de las secciones censales de la ciudad de Valencia.

carto_valencia

Cartografía de la ciudad de Valencia con 598 secciones censales.

0 Comments [figshare](#) [Login](#) [Sort by Newest](#)

[Recommend](#) [Share](#) [Start the discussion...](#)

CATEGORIES
• [Statistics](#)

KEYWORD(S)
[curso_reproducibilidad](#)

LICENCE
 CC BY

EXPORT
[RefWorks](#)
[BibTeX](#)
[Ref. manager](#)
[Mendeley](#)
[Endnote](#)
[DataCite](#)
[NLM](#)



① <https://bmccancer.biomedcentral.com/articles/10.1186/s12885-016-2725-z>

Supplemental data set

High sensitivity isoelectric focusing to establish a signaling biomarker for the diagnosis of human colorectal cancer

Narendra Padhan¹⁾, Torbjörn E. M. Nordling^{1,2,§}, Magnus Sundström¹⁾, Peter Åkerud³⁾, Helgi Birgisson³⁾, Peter Nygren¹⁾, Sven Nelander¹⁾ and Lena Claesson-Welsh^{1)*}

 figshare

Share Download

① journals.plos.org/plosone/article?id=10.1371/journal.pone.0052374

RAD51 and Breast Cancer Susceptibility: No Evidence for Rare Variant Association in the Breast Cancer Family Registry
Florence Le Calvez-Kelm, Javier Oliver, Francesca Damiola, Nathalie Forey, Nivonirina Robinot, Geoffroy Durand, Catherine Voegele, ...

PLOS

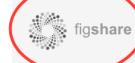
Abstract
Introduction
Results
Discussion
Materials and Methods
Supporting Information
Acknowledgments
Author Contributions
References

1

Le Calvez-Kelm et al. RAD51 case-control mutation screening study

1 | European | 805 / 913 | 0.063 / 0.062 | 0.86 | 1.05 [0.79, 1.40] | 0.75

1 | East Asian | 194 / 70 | 0.106 / 0.186 | 0.015 | 0.50 [0.28, 0.87] | 0.015

 figshare

download

Stratified analyses of RAD51 non-coding SNPs on breast cancer risk in the BCFR.

Table S1.

figshare

- En la primera práctica descargaremos esos datos:

```
fileUrl <- "https://ndownloader.figshare.com/files/8455721"  
download.file(fileUrl, destfile="./data/datos_practicas.RData")  
dateDownloaded <- date()
```

download.file mode="wb"

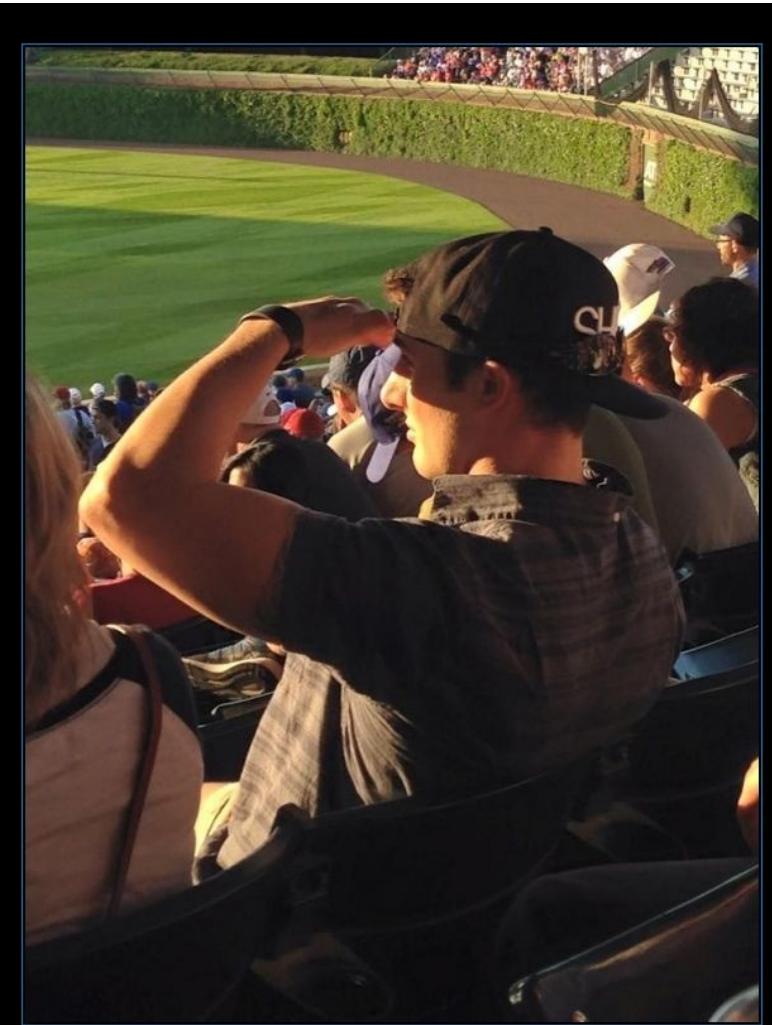
Resumen



Dropbox



Google Sheets



TENER TRES OPCIONES

Y elegir la peor

más en cuantarazon.com

Referencias bibliográficas

Referencias

- Broman, K. 2016.
<http://kbroman.org/Tools4RR/> (<http://kbroman.org/Tools4RR/>).
- Broman, Karl. 2016. <http://kbroman.org/Tools4RR/>
(<http://kbroman.org/Tools4RR/>).
- Bryan, Jennifer. 2015. . UseR! 2015.
<https://speakerdeck.com/jennybc/googlesheets-talk-at-user2015>
(<https://speakerdeck.com/jennybc/googlesheets-talk-at-user2015>).
- Gandrud, Christopher. 2015. . Chapman; Hall/CRC.
- Google, R user community at. 2014. .
<https://google.github.io/styleguide/Rguide.xml> (<https://google.github.io/styleguide/Rguide.xml>).
- Grant, Barry. 2016. . https://bioboot.github.io/web-2016/class-material/slides_day3-1_R.pdf
(https://bioboot.github.io/web-2016/class-material/slides_day3-1_R.pdf).
- Grolemund, Garrett, and Hadley Wickham. 2016. . O'Reilly. <http://r4ds.had.co.nz/>
(<http://r4ds.had.co.nz/>).
- Hornik, Kurt. 2008. https://cran.r-project.org/doc/FAQ/R-FAQ.html#Why-is-R-named-R_003f (https://cran.r-project.org/doc/FAQ/R-FAQ.html#Why-is-R-named-R_003f).

