

# Data Science (CDA)

## Getting Started with R

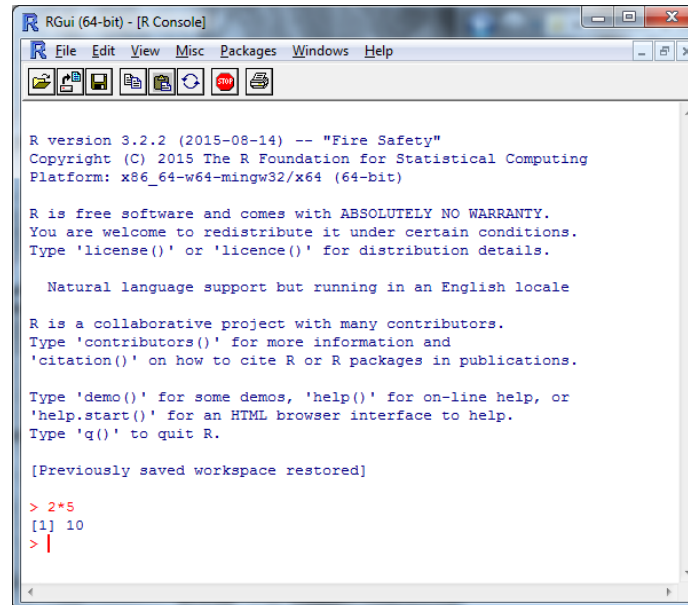
- M. José Ramírez Quintana, DSIC, UPV, [mramirez@dsic.upv.es](mailto:mramirez@dsic.upv.es)
- José Hernández Orallo, DSIC, UPV, [jorallo@dsic.upv.es](mailto:jorallo@dsic.upv.es)



- R is a powerful environment for statistical computing which runs on several platforms (Windows, Mac OS X, and Linux).
- R is available free of charge and is distributed under the terms of the Free Software Foundation's GNU General Public License. You can download the program from the Comprehensive R Archive Network (CRAN).
- R is an interpreted object oriented language.



- To start R in the the Windows environment double click on the R icon.
  - The symbol '>' indicates that R is expecting a command.



```
RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (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.

Natural language support but running in an English locale

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.

[Previously saved workspace restored]



> 2*5
[1] 10
> |
```

- To quit R, close the window or type **q()** followed by enter at the prompt. Note the parentheses after the q: in R you don't type commands but rather call functions.



- The R Console can be used like a calculator:

```
> 3 * 5  
[1] 15  
> sqrt(100)  
[1] 10  
> log(100)  
[1] 4.605
```

- The + prompt is used to indicate that R is waiting for more input. This allows you to break commands over
-  : navigate through your history with arrow keys.
- <ESC>: If things get messed up, press the escape key and try again.
- Ctrl-C: to stop execution or edition.



- The working directory (wd) is where R find all files for reading and writing.

```
>getwd() #get the current wd
```

- We can change the wd via the menu (depending on the platform)
  - *Misc* → *Change Working Directory* (Mac)
  - *File* → *Change Working Directory* (Windows)
- Or using the function

```
>setwd() #set the current wd
```



- To obtain information about a function, use the **help()** function.
- An alternative is to use the equivalent **?** operator, followed by the name of the function.
- Help can be displayed in a help window as a plain-text file, or as an HTML page in a web browser by using the command **help.start()**.
- A novel feature of the R help system is the facility it provides to execute most examples in the help pages via the **example** command. For instance, `example(max)`, `example(plot)`, ...
- Other sources of help: R manuals, frequently-asked-questions (FAQ) lists, forums or mailing lists, ...(<https://stat.ethz.ch/mailman/listinfo/r-help>, <http://tolstoy.newcastle.edu.au/R/> )



- User-defined variables and functions exist in R in a region of memory called the workspace. The R workspace can be saved during the session
  - *Workspace* → *Save Workspace File* (Mac)
  - *File* → *Save Workspace File* (Windows)

or even at the end of a session (when the `quit()` command is executed you will be asked whether you want to save the data from your R session).

- Data which is saved will be available in future R sessions:
  - *Workspace* → *Load Workspace File* (Mac)
  - *File* → *Load Workspace File* (Windows)
- For convention, the saved files have the extension or file type `.txt` or `.RData`.





- The Windows and Mac OS X implementations of R include basic programming or script editors to write our functions.
- You can open a new R script in the Windows *RGui* via the *File* → *New script* menu, or an existing script file via *File* → *Open script*. Similar *New Document* and *Open Document* selections are available under the Mac OS X *R.app File* menu.
- By convention, R script files have names that end with the extension or file type `.R`.
- An R script is executed by using the **source()** command (also available from the menu).





- Much of the power of R comes from the thousands of R packages containing code and data for specialised situations.
- The standard R installation comes with 8 packages.
- Probably, you will have to install the packages available at CRAN: <http://cran.es.r-project.org/>. It can be done
  - from the command line: **install.packages()**
  - from the menu: *Packages&Data* → *Package installer* (first the CRAN mirror is selected and then the package to be installed).
- To access the installed packages use the **library()** command



- Rstudio is an integrated development environment (IDE) for R.
- Rstudio is organised into four panes, some with multiple tabs.
- Some of the important tabs include
  - Console: This is where you can execute R commands interactively.
  - Source Editor: for editing functions and files.
  - History: A record of past commands (can be saved, reloaded, etc.)
  - Workspace: A listing of the objects available in your R session
  - Plots: Where plots show up
  - Help: Where documentation files appear when you ask for them
  - Files: A file manager for locating, loading, moving, renaming, files.
  - Packages: Install and load packages here.
  - Open Files: Open files have a tab labeled with the file name.



1

```

1 #Install the packages we need
2
3 .lib<- c("caret", "rpart", "kernlab","hydroGOF")
4 .inst <- .lib %in% installed.packages()
5 if (length(.lib[!.inst])>0) install.packages(.lib[!.inst])
6 lapply(.lib, require, character.only=TRUE)
7
8 setwd("~/Users/mjose/Desktop/MoreBikes/Material_para_DS2017")
9 directory<-getwd()
10 N=10 ##number of training stations
11 #Load the data in R
12
13 source("Load_dataBike.R")
14
1:1 (Top Level)
  
```

2

Console

```

~/
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.

Error in tools:::httpdPort <= 0L :
  comparación (4) es posible solo para tipos lista y atómico
[Workspace loaded from ~/.RData]
>
  
```

3

Environment

Global Environment

Data

- flights 227496 obs. of 21 variables
- hflights 227496 obs. of 21 variables

Values

- carriers Large character (227496 elements, ...)
- lut
- two

4

Files

Name	Description	Vers...
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-5
<input type="checkbox"/> arules	Mining Association Rules and Frequent Itemsets	1.5-0
<input type="checkbox"/> assertthat	Easy pre and post assertions.	0.1
<input type="checkbox"/> automap	Automatic interpolation package	1.0-14
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.0.5
<input type="checkbox"/> BBmisc	Miscellaneous Helper Functions for B. Bischl	1.11
<input type="checkbox"/> BH	Boost C++ Header Files	1.60.0-1
<input type="checkbox"/> bitops	Bitwise Operations	1.0-6
<input type="checkbox"/> boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18

1.View files and data (editor)

2. Workspace and history

3. Console

4. Files, plots, packages, help



- You can inspect the installed packages in the **Packages tab**.
  - Check marks indicate that the package is loaded (i.e., usable).
  - Click on the install packages icon to search for packages and install them.
- In the **Workspace/Environment tab** you find tools to help you import data.
- The **Files tab** provides a simple interface for finding, opening, moving, renaming files.



There are many, many excellent resources for newcomers to R:

- On-Line Introductions, e-Books and Tutorials
  - <http://cran.r-project.org/doc/manuals/R-intro.pdf>
  - <http://cran.r-project.org/doc/contrib/usingR.pdf>
  - <http://www.burns-stat.com/>
  - [http://en.wikibooks.org/wiki/R\\_Programming](http://en.wikibooks.org/wiki/R_Programming)
- Books
  - Adler. R in a Nutshell. O'Really 2012.
  - L. Torgo. Data Mining with R. CRC Press 2011.
  - Paul Teetor. 25 recipes for getting started with R. O'Really 2011.
- More references in Quick-R  
(<http://www.statmethods.net/about/books.html>)



- More references:
  - Randall Pruim. Getting Started with Rstudio.  
<http://www.calvin.edu/~rpruim/talks/Rminis/RStartingUp.pdf>
  - John Verzani. Getting Started with Rstudio. O'Really 2011.
  - D.V. Conesa Guillén. Curso Introducción R: Sesión 1.
  - [https://en.wikibooks.org/wiki/R\\_Programming](https://en.wikibooks.org/wiki/R_Programming)





- Interactive on-line courses
  - *swirl* is a software package that turns the R console into an interactive learning environment. Follow the instructions at <http://swirlstats.com/students.html> (Learn menu item) to install the swirl package and the interactive course of R.
  - Another introductory course of R can be found at <https://www.datacamp.com/courses/free-introduction-to-r>.

Through any of these courses (or both of them), you will become familiar with the R syntax, the different kinds of data (types) supported by R and how R programs are executed.

