The R Software Environment

a (very) short introduction

L. Torgo

ltorgo@dcc.fc.up.pt

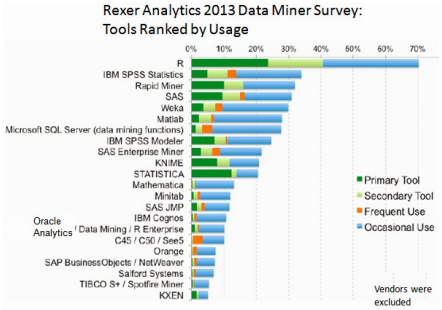
Departamento de Ciência de Computadores
Faculdade de Ciências / Universidade do Porto

Sept, 2014

The R Project (http://www.R-project.org)

What is R?

- R is a programming language and environment for data analysis.
- It is a free and open source software tool.
  - Trustable software; reproducible and adaptable analysis
- One of the most used data analysis tools.
- Strong impact on both academia and industry



The R Project (cont.)

- Available for Windows, several Unix/Linux variants and Mac OS X.
- Command line interface
  - interactive data analysis
  - user writes R commands, hits ENTER and gets back the results
- Base installation comes with an impressive set of functionalities - can be extended by installing extra packages freely available at R site (currently more than 5000!)

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

R Origins

- The S language
  - Developed in the late 70's by John Chambers and collaborators at Bell Labs.
  - Went through several changes having stabilized around the mid 90's.
- Implementations of the S language
  - *S-Plus* - a commercial version of S.
  - *R* - a free and open source version of the S language created at the University of Auckland and currently developped by several researchers across the world.

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

Installation

- Dependent on the host operating system
- Pre-compiled binaries available in R central repository (CRAN)
  - `http://cran.at.r-project.org/`
- Standard installation process similar to any other software

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

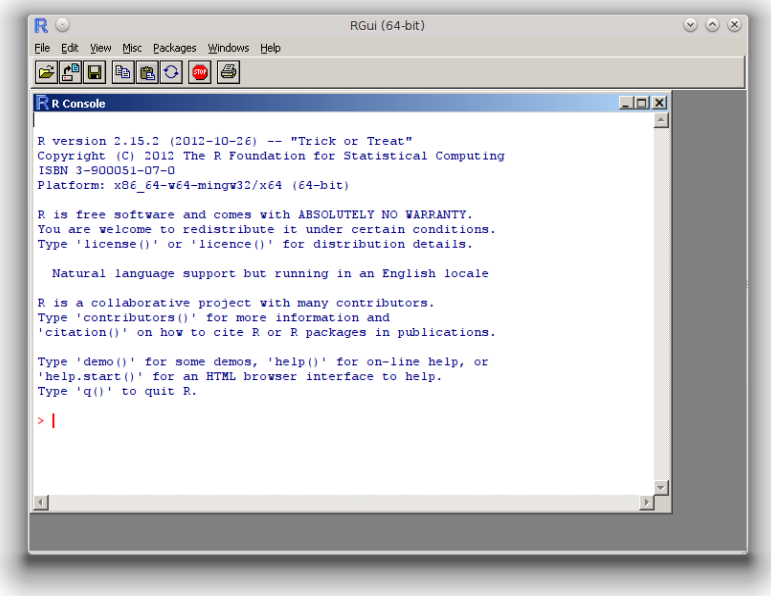
Installing R on Windows

1. Go the R Web Site (`http://www.r-project.org`)
2. Follow the link **CRAN** on the left menu
3. Choose a repository to carry out the *download*
4. Follow the *link Windows* available in the Section *Download and Install R*
5. Follow the *link* **base**
6. Click on *Download R 3.0.2 for Windows (52 megabytes, 32/64 bit)* to *download* of the file **R-3.0.2-win.exe** to some local folder on your computer.
7. Execute the downloaded file on your computer to install R

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

Starting to use R

To execute R on Windows it is enough to click on the desktop icon that is usually available, or then use the respective entry on the Start menu

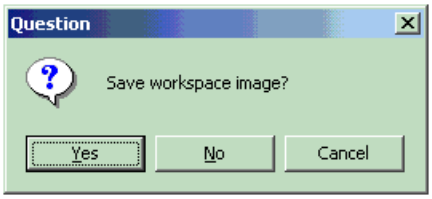


Quit from R

To quit from R it is enough to execute the command,

```
> q()
```

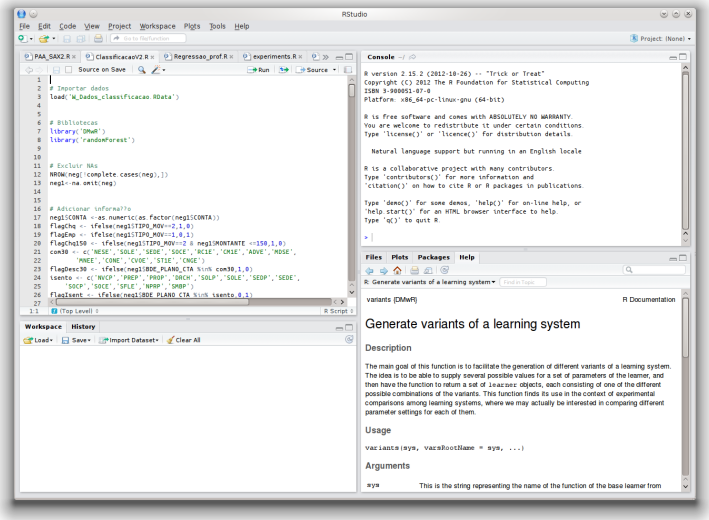
or else click on the "X" of the R window. Either way you will get the following dialog box,



If you answer *Yes*, R will store the information currently in memory on a file in the current directory, so that when we start R again in this directory it will recover this stored information allowing us to continue our work from where we left it.

RStudio

An excellent way of using R is through the integrated development environment provided by RStudio which is also freely available for Windows, OS X e Linux. This is probably the most user-friendly way of using R. RStudio can be obtained at `http://www.rstudio.org`



R Packages System

- Any base installation of R already comes with a series of installed packages.
- A package is a set of functions and data that someone has made available to the community in a free and open source way.
- Anyone can create a package and submit it to the R web site. A series of tests need to be overcome for the package to be officially posted in the site.
- On top of the packages included in a R base installation you may install many other extra packages
- After this installation (done once), whenever you wish to use these extra packages you must tell that to R, which is done by *loading* the package.

R Packages System (cont.)

- Installation of new packages is done through the function *install.packages* as illustrated in the following example:

```
install.packages("DMwR")
```

- To start using the functionalities of a package which is *already installed in our computer* we may use the function `library`, e.g.

```
library("DMwR")
```

- In some environments menu entries may exist to help in these tasks

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

Obtaining Help in R

- At the command line:
  - Typing `? commandName` at the command line
  - Typing `help(commandName)` at the command line
  - Typing `help.search(subject)` at the command line
  - Typing `RSiteSearch(subject)` at the command line
- Using the HTML help system coming with R
  - You start it typing `help.start()`
  - In RStudio this is embedded in the integrated environment
- Reading some of the freely available books available at the R web site
- Asking a question in the R-help mailing list

Important: Do read the posting guide before asking questions
`http://www.r-project.org/posting-guide.html`

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO

Some Books about R

- Adler, J. (2010): R in a nutshell. O'Reilly.
- Chambers, J. (2008): Software for Data Analysis, programming in R. Springer.
- Crawley, M. (2007): The R Book. Wiley.
- Dalgaard, P. (2008): Introductory Statistics with R. Springer.
- Torgo, L. (2010): Data Mining with R, learning with case studies. CRC Press.
- Williams, G. (2011): Data Mining with Rattle and R. Springer.

- Book Series on R from major publishers
  - *Use R!* - Springer
  - *The R Series* - CRC press

FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO