Short Introduction to R

Master of Science in Signal Theory and Communications

TRACK: Signal Processing and Machine Learning for Big Data

Instr.: **J.L. Blanco-Murillo**

jl.blanco@upm.es, Room C-329

Departamento de Señales, Sistemas y Radiocomunicaciones

E.T.S. Ingenieros de Telecomunicación

Universidad Politécnica de Madrid

What is R?

- High-level language to perform calculations
- Interactive programming environment
- Focus on data analysis
  - clean, organize, store and recover data
  - analyze
  - program
    - You can write your own functions in R language, and there are interfaces for OS commands, or functions written in other languages (C, Fortran…)
  - create plots (visualizations) for inspection.
    Different ways of “looking” at the data
  - compute batches and/or data streams

Where does it come from? Where does it work?

- R was at first known as “GNU S”
- Similar to the S language, that was developed in Bell Labs by John Chambers and others for internal use, and was the base of the S-Plus (comercial product)
- www.r-project.org
- Compiled versions for Windows, macOS and Linux (some variants)
- Source code to compile for any platform
- Frequent updates (latest, july 2019)

What is it intended for?

- Data inspection and evaluation
- Experimentation / testings
- Descriptive statistics
- Also possible to run big-data analysis (with specific packages)

What does it have?

- A myriad of numerical and graphical statistical tools / implemented techniques
  - Linear and non-linear modelling, temporal series analysis, classification, clustering…
- It is extensible
  - Base R environment
  - About 25 additional packages in the R distribution (standard and recommended)
  - Thousands available
- We are going to follow a basic tutorial

Tips & trics handling R

- Workspace management (e.g., saving and recovering).
- Recommened: one folder per task.
- Almost all computations carried in memory (i.e., RAM demanding).
- Use GUI for inspection, experimentation and beta testing (e.g., R-studio).
- Use Command line (pure scripting) & packages for demanding tasks (big-data).
- Particularly useful to use dataframes (optimized memory handling and speed for big data).

Benchmarking w/ other languages

Concept	R	Others
Memory management	All in memory	Optimized access (as Spark)
Processing time	Relatively low (actually low-level routines tend to written in optimized languages)	Others tend to be faster (but presentation is always costly)
Data access and storage	Run time: RAM Storage: full connectiviy (ODBC, SQL, spark, etc.)	Mixed approaches Promote interconnectivity.
Operating Programming paradigm	Object-oriented (as Java, C++, python)	Non object oriented (C, Fortran, Pascal...)
Execution, debugging and monitoring	Specific functions (Rstudio may be useful)	Various approaches. (Highly connected to IDE)
Language learning curve	Said to be slow	Python is said to be relatively quick...
Data structures	Vectorized (as MATLAB)	Coefficients – elementwise (as C)

Some texts that may be useful

E. Paradis, R for beginners

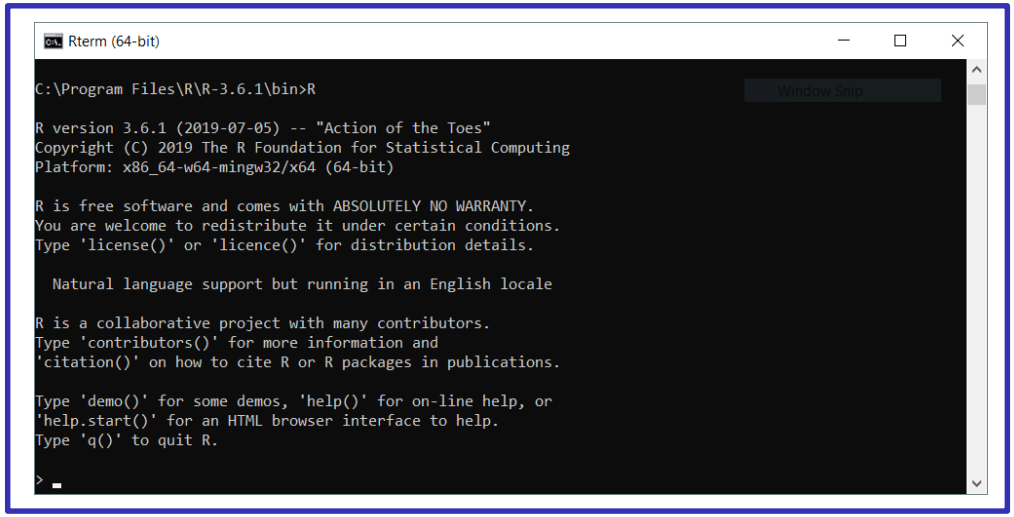
W.N. Venables, D. M. Smith and the R Core Team, An introduction to R

Installation

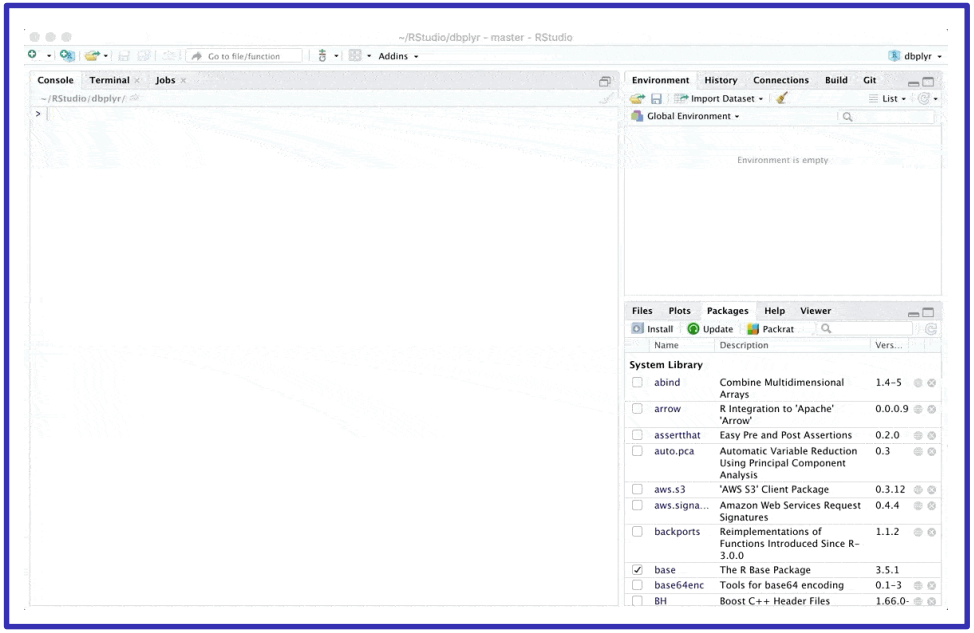
- Installation is simple
  - Go to https://cran.r-project.org
  - Choose your platform (Windows, Linux, MacOS)
  - Download corresponding file
  - Run file / installation script
- For this course, we recomment that you use RStudio, as a GUI that integrates some useful tools.
  - Go to https://www.rstudio.com
  - Choose your platform (Windows, Linux, MacOS)
  - Download corresponding file
  - Run file / installation script

Ready to work

Scripting language.
No need for more than a plain text file and a command line.



Graphical interface,
including packages installation tool, data import tools, etc.



Test the tool

Scripting language.
No need for more than a plain text file and a command line.

On Rterm (on system terminal launch R),

Write: 3*2
Press enter
Check result (no errors expected)

Graphical interface, including packages installation tool, data import tools, etc.

On Rstudio Console área,

Write: 3*2
Press enter
Check result (no errors expected)

Next steps

Assignment (for today / tomorrow):

1. Install R on your laptop
2. Install RStudio on your laptop
3. Check R and RStudio work
4. Bring laptop to coming sessions

Coming sessions:

- Wednesday **11/09/2019**, 17:00 to 19:00
  17:00 to 17:30 solve installation issues
  Short Introduction to language
- Wednesday **18/09/2019**, 17:00 to 20:00
  Statistical basics