

Reproducible Research: The R Ecosystem, Github, Etc.

Departamento de Automática
Universidad de Alcalá



`/gso>`

Table of Contents

1 Introduction

- The practice of research
- Reproducible research
- Components of reproducible research
- Literate Programming
- Tex and Latex

2 R Ecosystem

- The R language
- R Studio as Open IDE for R
- Markdown
- knitr
- Bookdown

3 Collaborating and sharing your research

- Git

4 Conclusions

Introduction

The practice of research (I)

What is research?

- In short, writting papers

Many processes involved

- Code development
- Dataset collection
- Scripts for everything
- Figures creation
- Statistics and analysis

The review comes, after 6 months

- ... you must repeat an experiment
- ... or change a figure or ...

Introduction

The practice of research (I)

What is research?

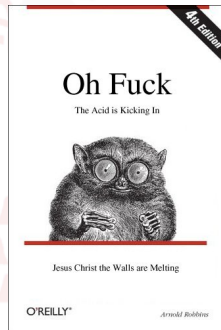
- In short, writting papers

Many processes involved

- Code development
- Dataset collection
- Scripts for everything
- Figures creation
- Statistics and analysis

The review comes, after 6 months

- ... you must repeat an experiment
- ... or change a figure or ...



Introduction

The practice of research (II)

Even worse: Imagine you need to reproduce other's experiments

- Or any one else reproduce your experiments ... **in teams**

Need of many assets

- Experiment run
- Data analysis
- Experiment documentation
- Paper authoring
- **On-line publication**

Need of procedures and tools to handle this

Our objective

Formalize the whole publishing *collaborative* process

Introduction

Reproducible research

Two types of experiments

- Repetible
- Reproducible

“Reproducible research is the idea that data analyses, and more generally, scientific claims, are published with their data and software code so that others may verify the findings and build upon them.”

<http://reproducibleresearch.net/>

Reproducible research is a cornerstone of research

- Science **must** be reproducible

Introduction

Components of reproducible research

Task	Tool
Collaborative code development	Git, GitHub
Experiment run	Python
Statistical analysis	R
Process documentation	Markdown
Results writting	Bookdown, Knitr, Swave
Assets publication	GitHub

Literate Programming

Literate programming is an approach to programming introduced by Donald Knuth in which a program is given as an explanation of the program logic in a natural language, such as English, interspersed with snippets of macros and traditional source code, from which a compilable source code can be generated.

- It has been around for many years...



https://en.wikipedia.org/wiki/Literate_programming

Tex

$\text{T}_{\text{E}}\text{X}$ (= tau epsilon chi, and pronounced similar to "blecch", not to the state known for 'Tex-Mex' chili) is a computer language designed for use in typesetting; in particular, for typesetting math and other technical (from Greek "techne" = art/craft, the stem of 'technology') material.

tug.org/whatis.html

- It has been around for many years...
-

Tex and Latex

\LaTeX , which is pronounced «Lah-tech» or «Lay-tech» (to rhyme with «blech» or «Bertolt Brecht»), is a document preparation system for high-quality typesetting. It is most often used for medium-to-large technical or scientific documents but it can be used for almost any form of publishing.

<http://www.latex-project.org/about/>

- It also has been around for many years...
-

What is R?

R is a programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

Wikipedia:

[https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))

What is RStudio?

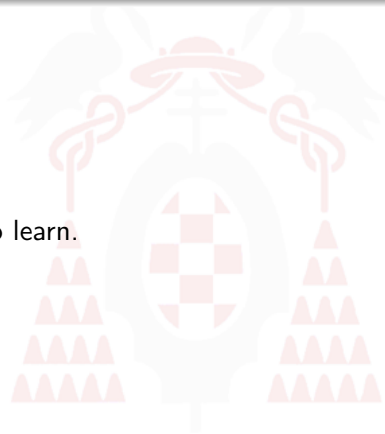
R is a programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

Wikipedia:

[https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))

Markdown

Extremely easy to learn.



knitr

<http://yihui.name/knitr/demo/minimal/>



knitr

<http://yihui.name/knitr/demo/minimal/>



knitr

knitr extracts R code in the input document, evaluates it and writes the results to the output document

<http://yihui.name/knitr/>

Rnw, Markdown, HTML and LaTeX

Git and GitHub



Git

Git

[https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))

GitHub

GitHub

www.github.com

Publish Websites, etc.



Conclusions

- Learn Bash and other tools for it: awk, make, sort,
- Learn Git, R or Python.
- If you use R/Python for your research automate as much as possible. It is much easier to modify.
- Try to stick to Open/libre software GNU/Linux, R, Python, etc.