Introductory course on the R software

Benoit Liquet[1]

[1]University of Pau & Pays de L'Adour
Laboratory of Mathematics and Its Application
France

University Sebalas Maret
Department of Mathematics
Faculty of Mathematics and Natural Science
February 28, 2017

`http://benoit-liquet.github.io`

Audience : who should attend this course ?

People who expect to :

- learn the basics of **R** (obviously !) ;
- want to perform sophisticated data manipulation or non-standard statistical analyses ;
- plan to use it on a regular basis in the future (if not, you will probably forget everything ...).

Level of difficulty

Matlab users (who can read and produce code) will learn something here ... but I will start with the very basics.

Reference for the course
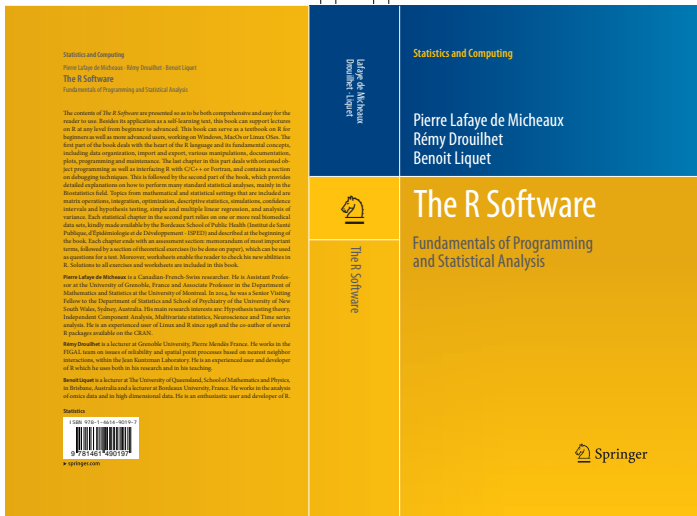
Course format

- Some "theory" and a lot of practice !
- Each people attending the course should bring his/her own laptop.
- "Assignments" will be given for each future session.

Questions ?

Feel free to interrupt me at any time !

Schedule

Day 1 Morning (today), we start with the very basics (Chapter 1) and Chapter 3 - Basic Concepts and Data Organisation).

Day 1 Afternoon : Graphics with 'ggplot2' .

Day 2 Morning : Linear Modelling.

Day 2 Afternoon : Programming in R.

Why use **R** ?

R is :

- free ;
- open-source ;
- cross-platform (Windows, MacOS, Linux, etc.) ;
- rapidly evolving (many packages added each day !).

Warning

R is harder to comprehend than other software on the market. You need to spend time learning the syntax and commands to become an efficient user.

Why use **R** ?

R is especially powerful for data manipulation, calculations and plots.
Its features include :

- an integrated and very well conceived documentation system
  (in English) ;
- efficient procedures for data treatment and storage ;
- a suite of operators for calculations on tables, especially
  matrices (but also arrays) ;
- a vast and coherent collection of statistical procedures for
  data analysis ;
- advanced graphical capabilities ;
- a "simple" and efficient programming language, including
  conditioning, loops, recursion and input-output possibilities.

Installing **R**

Let's start by installing the software !

1. Download the file R-*x*-win.exe (where *x* is the number of the latest version) at the address :
   http://cran.r-project.org/bin/windows/base/

2. Save this executable file on the Windows Desktop and
   double-click the file R-*x*-win.exe (its icon is   ).

3. The software then installs. Follow the instructions displayed on your screen and keep the default options.

4. When the icon   is added to the Desktop, installation is complete.

Installing **R** for MacOS or Linux

Download the file `R-x-snowleopard.pkg` (where *x* is the number of the latest version, i.e. 3.1.0) at the address : `http://cran.r-project.org/bin/macosx/`

I guess that if you use Linux, you will know how to do it !

R and Statistics

Many **classical** and **modern** statistical techniques are implemented in **R**. The most common methods for statistical analysis, such as :

- descriptive statistics ;
- hypothesis testing ;
- analysis of variance ;
- linear regression methods (simple and multiple) ;
- and so on

are directly included at the **core** of the system.

Third part of my book covers the following notions : basic mathematics, descriptive statistics, generation of random values, confidence intervals and hypothesis testing, simple and multiple linear regression, elementary analysis of variance.

Extending **R**

Most **advanced or recent statistical methods** are available through external **packages**, easy to install from R (see Section A2. from my Book).

They are all grouped and can be browsed on the website of the *Comprehensive R Archive Network* (CRAN) :
`http://cran.r-project.org/web/packages/available_packages_by_name.html`

See also the Task Views (on the CRAN) that group packages related to some domains of interest :
`http://cran.r-project.org/web/views/`

Official website of **R** : The **C**omprehensive R **A**rchive **N**etwork

`http://cran.r-project.org`

R is Popular with a Large & Steadily Growing User Base

CRAN

Mirrors
What's new?
Task Views
Search

About R

R Homepage
The R Journal

Software

R Sources
R Binaries
Packages
Other

Documentation

Manuals
FAQs
Contributed

CRAN Task Views

|	
Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
NumericalMathematics	Numerical Mathematics
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
SpatioTemporal	Handling and Analyzing Spatio-Temporal Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
WebTechnologies	Web Technologies and Services
gR	gRaphical Models in R

Why Use R ?
R has powerful graphics authoring capabilities
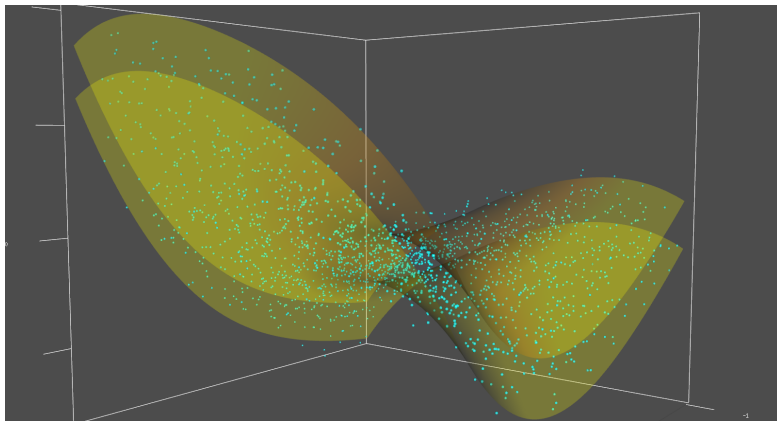
3D visualisation produced with the 'rgl' R package
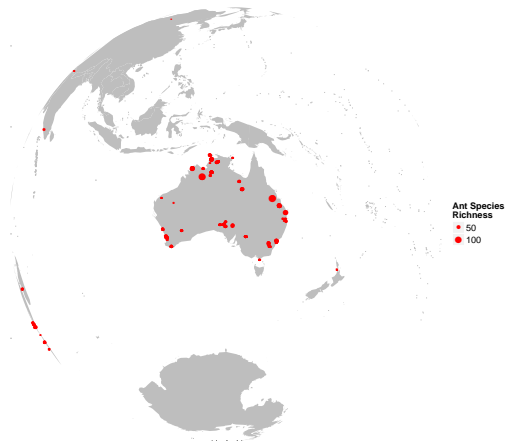
Why Use R ?
R has powerful graphics authoring capabilities
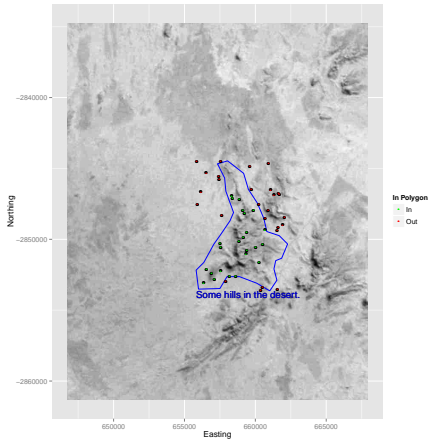
Geospatial Visualisation produced with the R packages 'raster' & 'ggplot2'

Why Use R ?
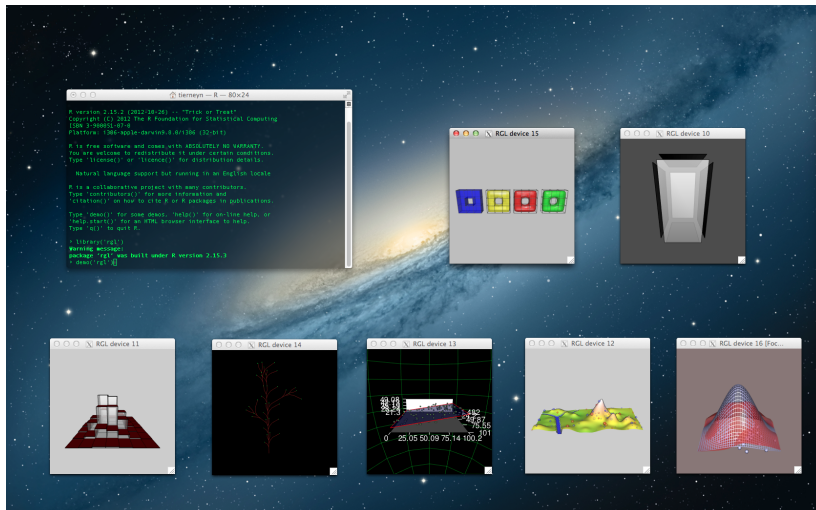R has powerful graphics authoring capabilities

Ant Species Richness
- 50
- 100

Ways to Use R

- via a command line interface e.g. PowerShell or Terminal

- via the default GUI clients for MS Windows & Mac OS

- via one of many Integrated Development Environments that either have been exclusively written for R or have R language modes e.g.
    - RStudio
    - Tinn-R
    - Sublime Text
    - Atom
    - Emacs Speaks Statistics
    - ...

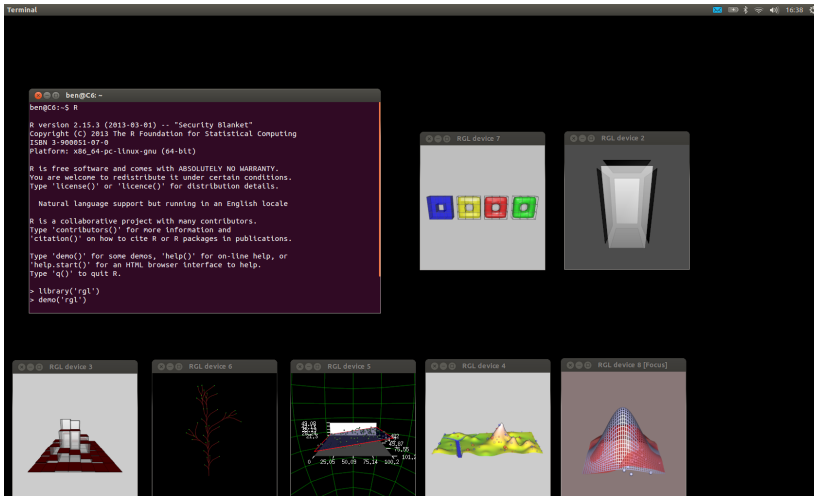- remotely i.e. submitting R scripts to a sever (e.g. HPC facility) to execute
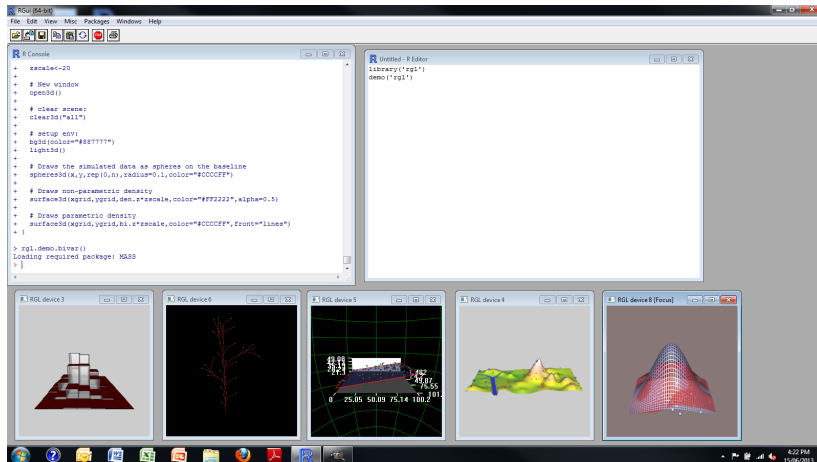
Ways to Use R :
In a termial e.g. on Mac OS

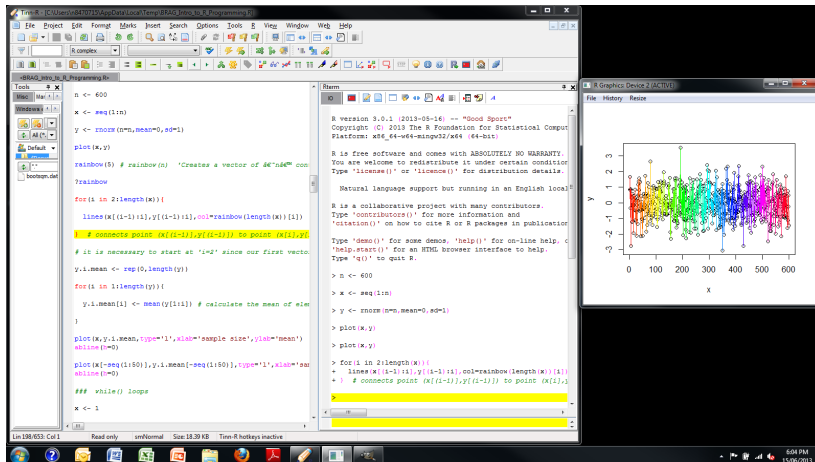# Ways to Use R :
In a termial e.g. on GNU+Linux

Liquet, Benoit   Introductory course on the R software
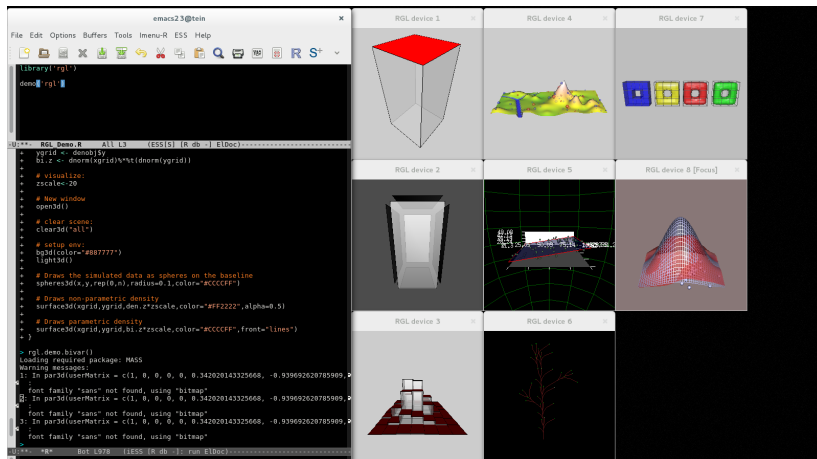
Ways to Use R :
Default Windows Client

Ways to Use R :
Tinn-R Integrated Development Environnement

Ways to Use R :
Emacs Speaks Statistics Integrated Development Environement

Ways to Use R :
RStudio Integrated Development Environement

The R Graphical User Interface

The R Graphical User Interface (*i.e.* its set of menus) is very limited, and completely nonexistent on some operating systems, when compared to other standard software (SPSS say). This minimality can set back some new users. However, this drawback is limited since :

- it has the didactic advantage that it incites users to know well the statistical procedures they wish to use ;
- there are additional tools which extend the GUI

Please, download and install RStudio :
https://www.rstudio.com/ide/download/desktop

For this course I encourage you to use the RStudio IDE

Because :...

- it's comparatively intuitive and easy to learn

- feature rich

- available for most major operating systems (MS Windows, Mac OS, various flavours of GNU+Linux)

However, if you have already begun your journey learning R using a different IDE and wish to continue to use it please feel free to do so, provided you feel confident to open and execute .R files with this IDE.

R and other software

R can interact with the following software :

- use R from within Excel :
  http://rcom.univie.ac.at/download.html#RExcel
- R from within SAS : http:
  //support.sas.com/rnd/app/studio/Rinterface2.html
- R from within SPSS : http://www.ibm.com/
  developerworks/library/ba-call-r-spss
- R from Matlab : http://www.mathworks.com/
  matlabcentral/fileexchange/5051-matlab-r-link
- Matlab from R : install package R.matlab

R and plots

Let's play a little bit ! Try these commands :

```
demo(image)
example(contour)
demo(graphics)
demo(persp)
demo(plotmath)
demo(Hershey)
install.packages(""lattice)
require("lattice") # Load package, previously installed.
demo(lattice)
example(wireframe)
install.packages("rgl")
require("rgl")
demo(rgl) #  Interact using your mouse.
example(persp3d)
```

Liquet, Benoit   Introductory course on the R software

The Plan
Feel free to use this time to pursue something that interests you

Course organised into 4 instructory modules and, collaborative exercise.

Module :

1. Introduction to R & RStudio & Rcommander
2. Graphics with the R package 'ggplot2'
3. Linear Modelling in R
4. Programming in R

Liquet, Benoit   Introductory course on the **R** software

Module 1
Introduction to R & RStudio & Rcommander

Key Learning Outcomes

Familiarisation with

- R using Rcommander
- Command Line Computing
- RStudio Integrated Development Environment
- Commands and arguments
- Common Object Classes in R
- Assigning values to Objects
- Saving & Loading R Workspaces
- R Base Graphics
- Data Input

Your first steps in **R**

Let's start our journey using Rcommander

1. Open **R** via Rsudio or via a terminal
2. use the following command in the console
   library(Rcmdr)
   Commander()
3. After my demo on Rcommander, Please, do by yourself the contents of Section 1.5 of my Book.

   http:
   //biostatisticien.eu/springeR/Rbook-chap1.pdf

Warning

- **R** Commander is a tool mostly dedicated for beginners.
- **R** is case sensitive.

Thank you for your attention !