Lightning talks

Améliorer la qualité de son package R avec l'intégration continue

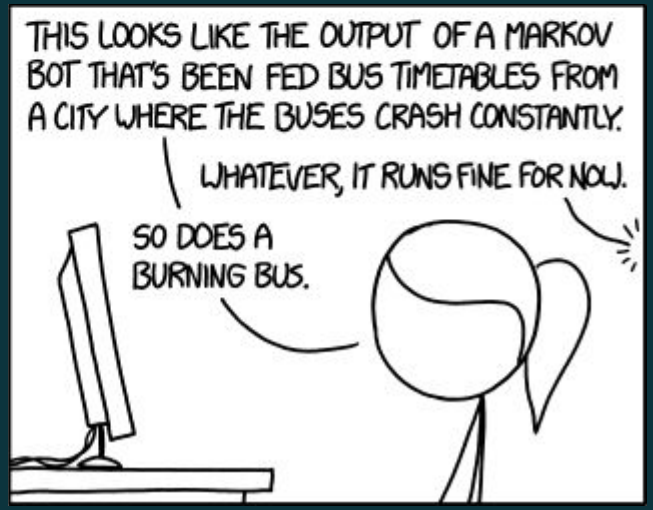
Géraud DUGE de BERNONVILLE, Valtech

Intégration continue avec R

Ou comment améliorer la qualité de son package ?

Géraud Dugé de Bernonville
@geraudster

Qualité ?



<< Chez Moi Ça Marche ! >>

Intégration continue

1

Référentiel de source Commun

CVS

2

Tests

Exécutables

Automatisés

Dans le référentiel de source

devtools
+
testthat

Construction régulière

et automatisée

3

Analyse de code

Tests

Packaging

Rapport

TeamCity

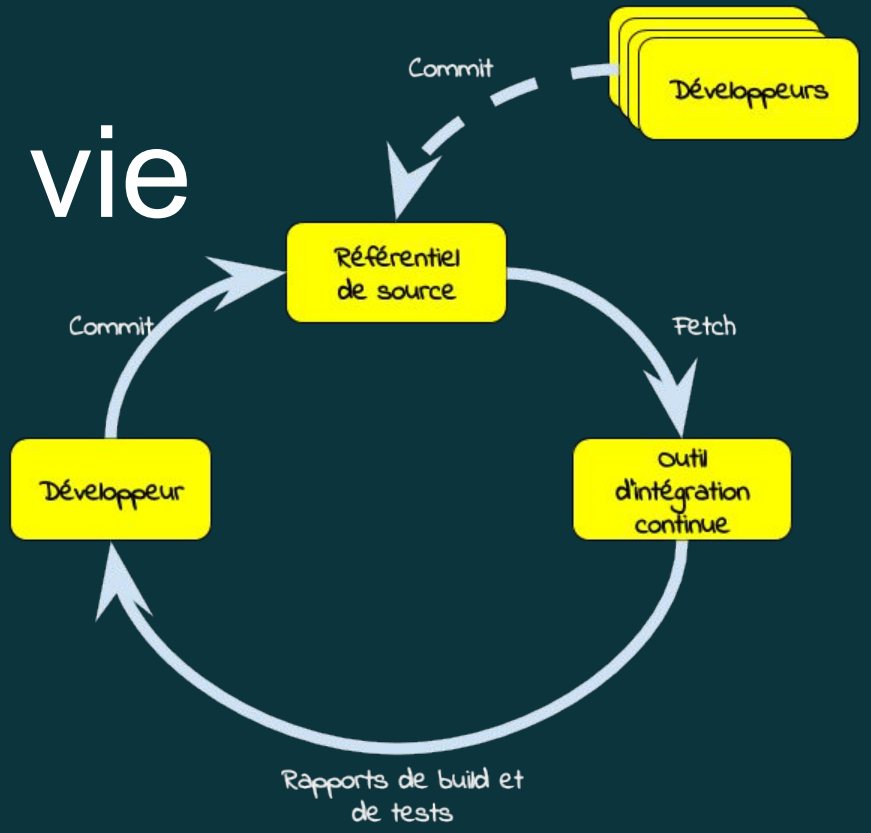
circleci

Travis CI

continuum™

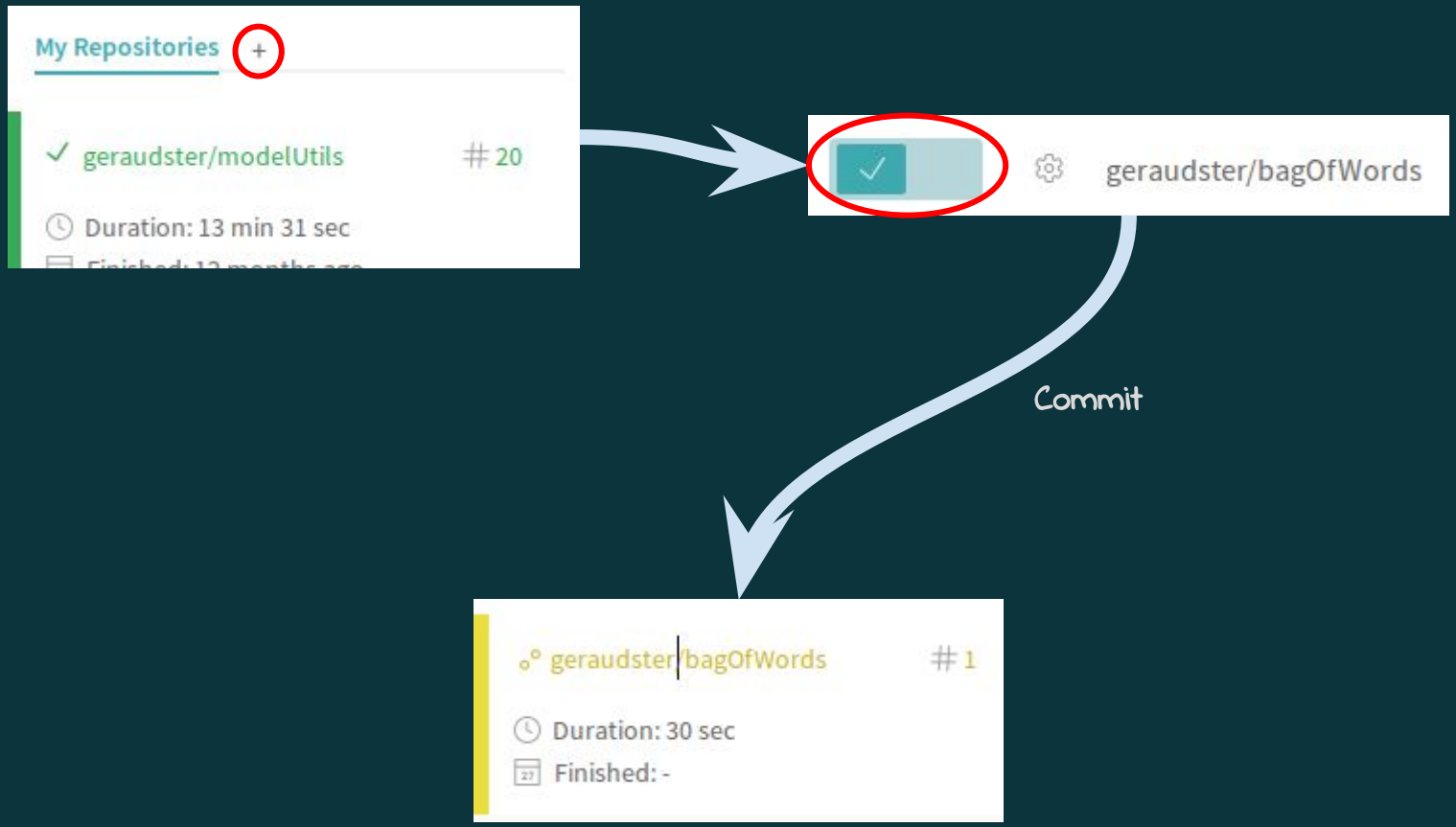
Jenkins

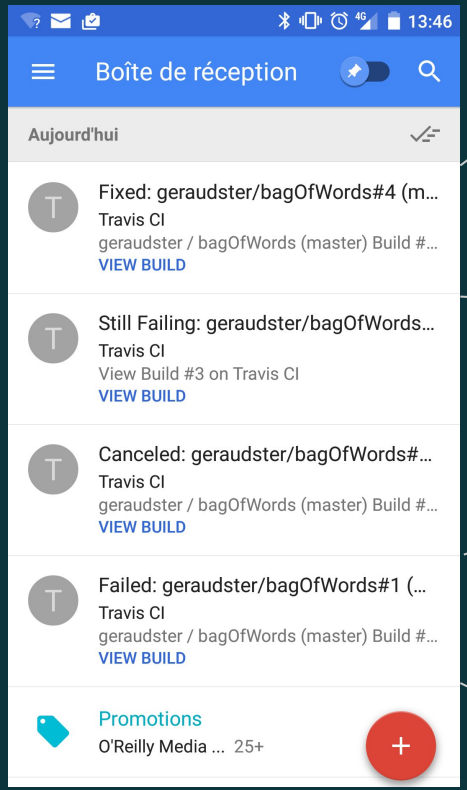
Le cycle de vie du dev

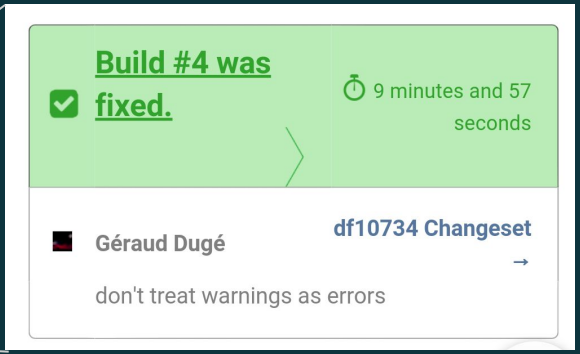


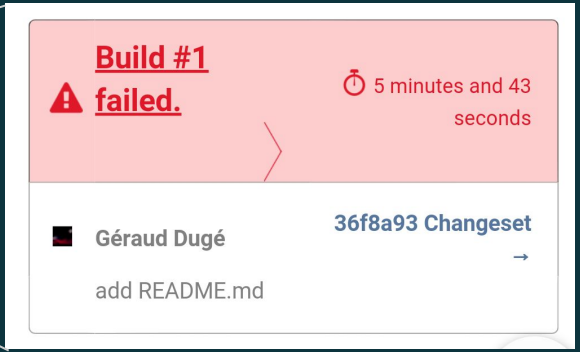
.travis.yml

```
language: r

sudo: false
```


My Repositories
+
geraudster/modelUtils
20
Duration: 13 min 31 sec
geraudster/bagOfWords
Commit
geraudster/bagOfWords
1
Duration: 30 sec
Finished: -


13:46
Boîte de réception
Aujourd'hui
Fixed: geraudster/bagOfWords#4 (m...
Travis CI
geraudster / bagOfWords (master) Build #...
VIEW BUILD
Still Failing: geraudster/bagOfWords...
Travis CI
View Build #3 on Travis CI
VIEW BUILD
Canceled: geraudster/bagOfWords#...
Travis CI
geraudster / bagOfWords (master) Build #...
VIEW BUILD
Failed: geraudster/bagOfWords#1 (...
Travis CI
geraudster / bagOfWords (master) Build #...
VIEW BUILD
Promotions
O'Reilly Media ... 25+


Build #4 was fixed.
9 minutes and 57 seconds
Géraud Dugé
df10734 Changeset
don't treat warnings as errors

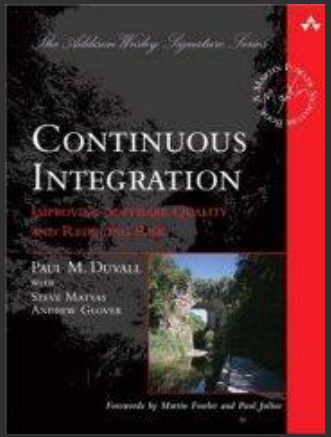

Build #1 failed.
5 minutes and 43 seconds
Géraud Dugé
36f8a93 Changeset
add README.md

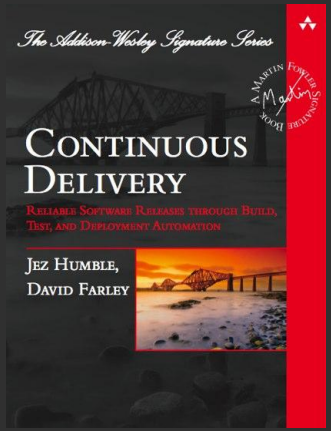
En savoir plus...

http://martinfowler.com/delivery.html
https://travis-ci.org/

@geraudster

MARTINFOWLER.COM





Cinquièmes Rencontres R

Toulouse
du 22 au 24 juin 2016

Lightning talks

Géraud DUGE de BERNONVILLE, Valtech

Nathalie Villa-Vialaneix, INRA, UR875, MIAT

Sébastien Déjean, IMT, Toulouse.

Christophe Bontemps, Toulouse School of Economics -INRA

Cinquièmes Rencontres R

Toulouse

du 22 au 24 juin 2016

Lightning talks

Outils pour chercher de l'information sur R et se former

Nathalie Villa-Vialaneix, INRA, UR875, MIAT

Sébastien Déjean, IMT, Toulouse.

Christophe Bontemps, Toulouse School of Economics -INRA

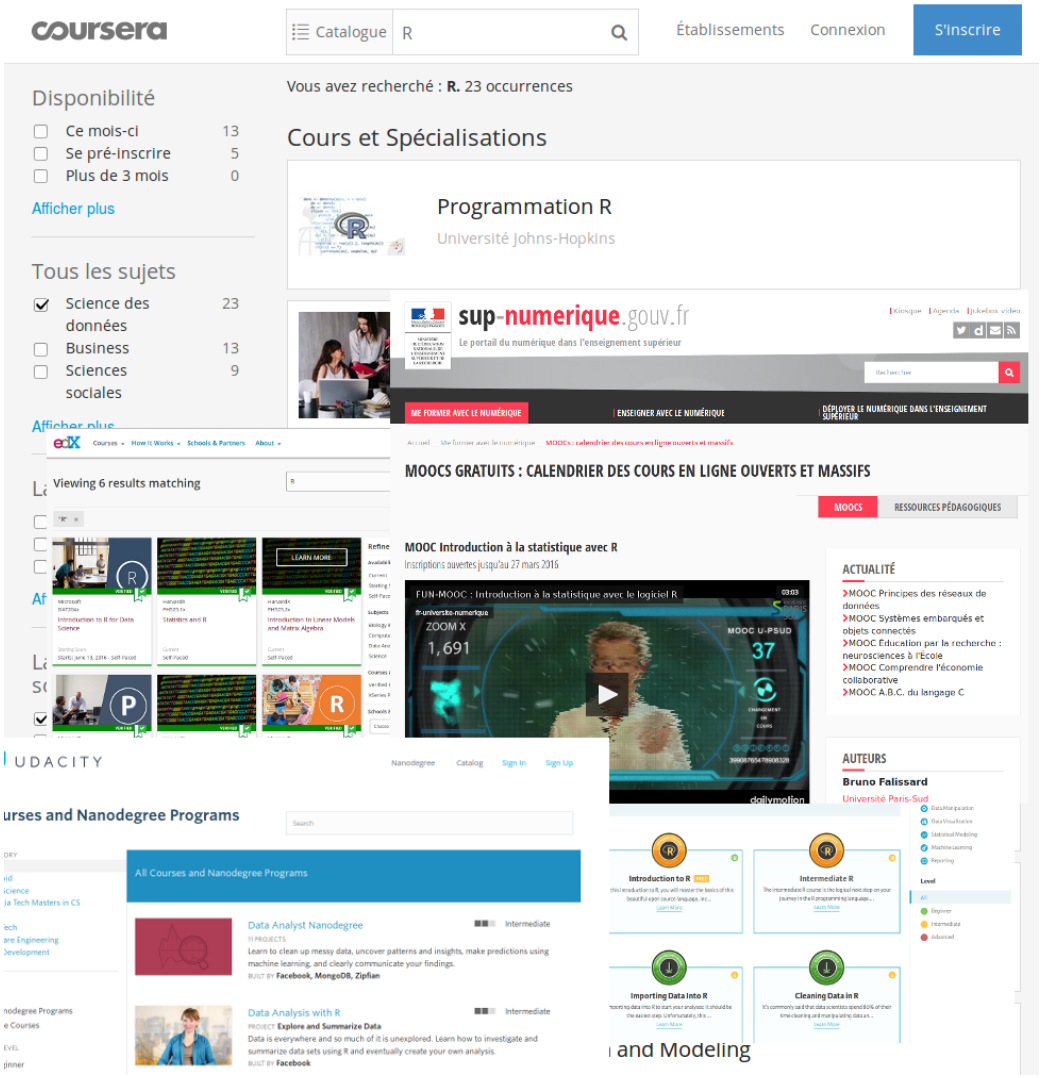
Outils pour chercher de l'information sur R et se former

Nathalie Villa-Vialaneix, Christophe Bontemps & Sébastien Déjean

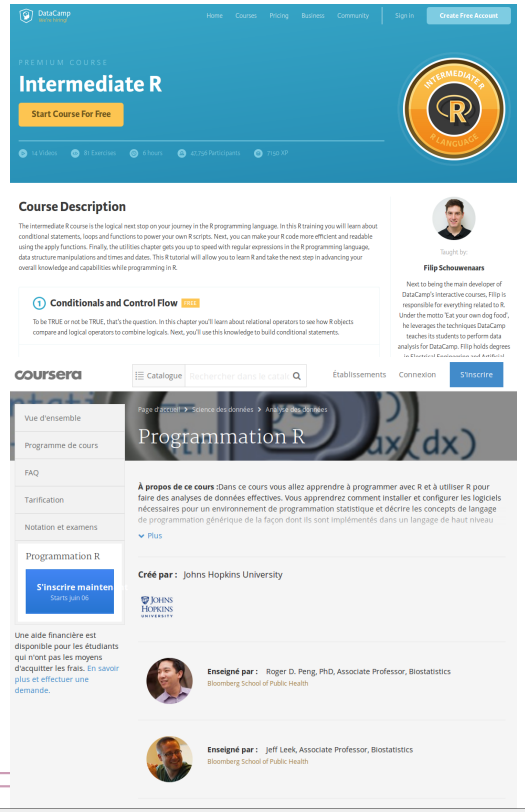
1. où commencer avec R ?
2. à ne pas manquer !
3. à l'aide... !!!

Toulouse 2016

Où commencer ? Un MOOC... ?

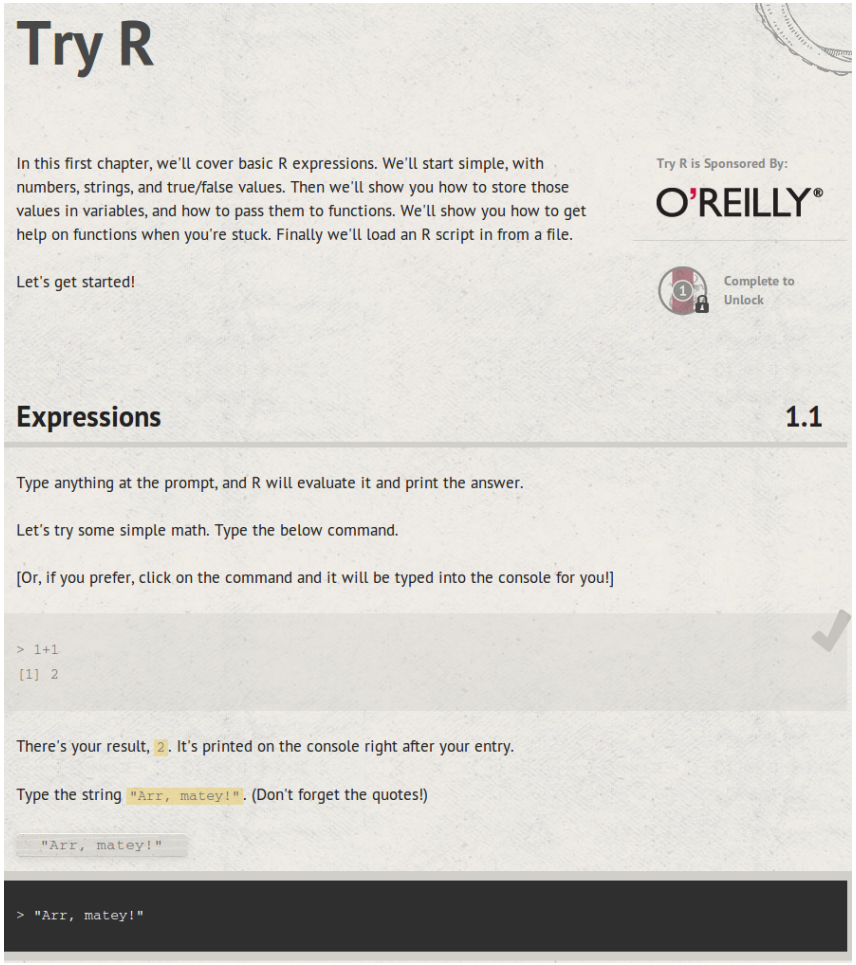


- de nombreux MOOCs sur R ou utilisant R sur coursera, udacity, edX, datacamp, FUN...
- nous recommandons :



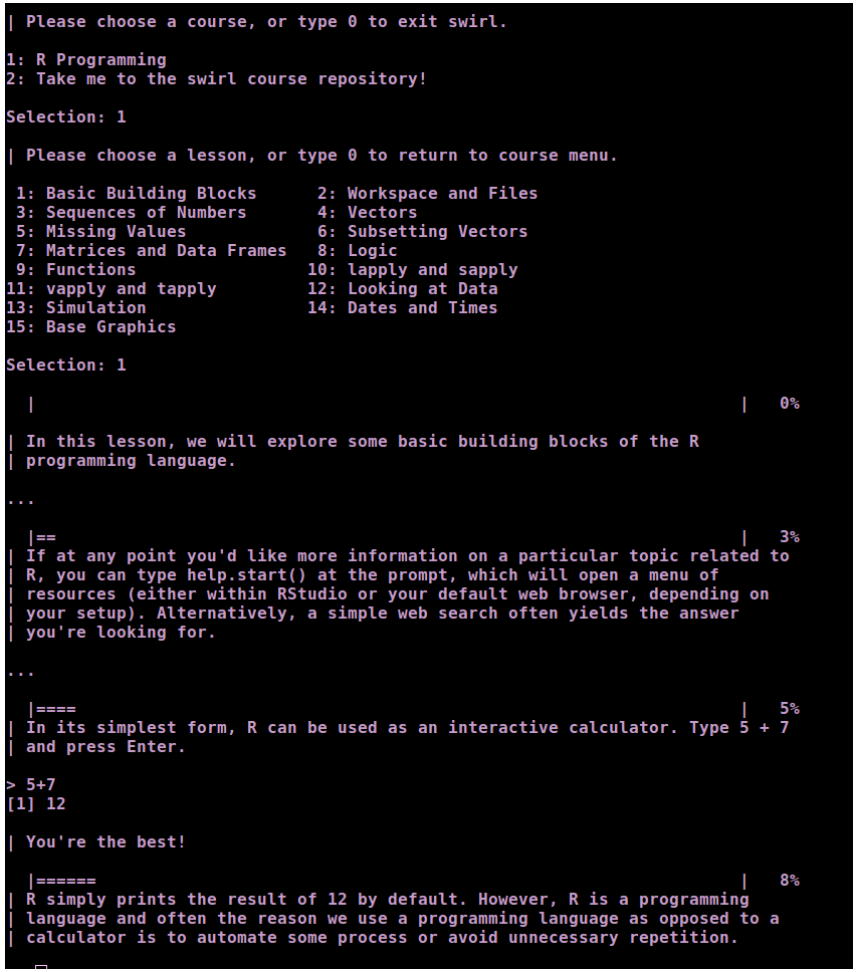
Où commencer ? En interactif...

On line



http://tryr.codeschool.com

Off line




```
install.packages("swirl")
```

À ne pas manquer ! Les manuels du CRAN...

The R Manuals

edited by the R Development Core Team.

The following manuals for R were created on Debian Linux and may differ from the manuals for Mac or Windows on platform-specific pages, platform are part of the respective R installations. The manuals change with R, hence we provide versions for the most recent released R vers a version for the forthcoming R version that is still in development (R-devel).

Here they can be downloaded as PDF files, EPUB files, or directly browsed as HTML:

Manual	R-release	R-patched	R-devel
An Introduction to R is based on the former "Notes on R", gives an introduction to the language and how to use R for doing statistical analysis and graphics.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
R Data Import/Export describes the import and export facilities available either in R itself or via packages which are available from CRAN.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
R Installation and Administration	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
Writing R Extensions covers how to create your own packages, write R help files, and the foreign language (C, C++, Fortran, ...) interfaces.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
A draft of **The R language definition** documents the language *per se*. That is, the objects that it works on, and the details of the expression evaluation process, which are useful to know when programming R functions.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
R Internals: a guide to the internal structures of R and coding standards for the core team working on R itself.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
The R Reference Index: contains all help files of the R standard and recommended packages in printable form. (9MB, approx. 3500 pages)	PDF	PDF	PDF

À ne pas manquer ! Les vues thématiques du CRAN...

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

CRAN Task View: High-Performance and Parallel Computing with R

Maintainer: Dirk Eddelbuettel
Contact: Dirk.Eddelbuettel at R-project.org
Version: 2016-01-28

This CRAN task view contains a list of packages, grouped by topic, that are useful for high-per
related to pushing R a little further: using compiled code, parallel computing (in both explicit a

Unless otherwise mentioned, all packages presented with hyperlinks are available from CRAN,

Several of the areas discussed in this Task View are undergoing rapid change. Please send sug

Suggestions and corrections by Achim Zeileis, Markus Schmidberger, Martin Morgan, Max Kul
Bill Cleveland, Ross Boylan, Ramon Diaz-Uriarte, and Mark Zeligman (as well as others I may

Contributions are always welcome, and encouraged. Since the start of this CRAN task view in
GitHub repository (see below) so that pull requests are also possible.

Direct support in R started with release 2.14.0 which includes a new package **parallel** in
'parallel'. However, and as explained in the package vignette, the parts of parallel which provic
The **parallel** package also contains support for multiple RNG streams following L'Ecuyer et al
The version released for R 2.14.0 contains base functionality: higher-level convenience functior

Parallel computing: Explicit parallelism

- Several packages provide the communications layer required for parallel computing. The
  longer actively maintained, but available from its CRAN archive directory.
- In recent years, the alternative MPI (Message Passing Interface) standard has become the
  access to numerous functions from the MPI API, as well as a number of R-specific extensi
  is now in maintenance mode, and new development is focussed on Open MPI.
- The pbdMPI package provides S4 classes to directly interface MPI in order to support the
  builds on this and uses scalable linear algebra packages (namely BLACS, PBLAS, and Sca
  for distributed data types upon which the pbdDMAT builds to provide distributed dense m
  synchronization) and supports terabyte-sized files. The pbdDEMO package provides exam
  libraries, such as fpmpi, mpiP, or TAU.
- An alternative is provided by the nws (NetWorkSpaces) packages from REvolution Compu
  toolkit for Python.
- The snow (Simple Network of Workstations) package by Tierney et al. can use PVM, MPI,
  provides fault-tolerance extensions to snow.
- The snowfall package by Knaus provides a more recent alternative to snow. Functions car
- The foreach package allows general iteration over elements in a collection without the us
  (using parallel/multicore on single workstations), doSNOW (using snow, see above), doMP
- The future package allows for synchroneous (sequential) and asynchronous (parallel) eva
  Iteration over elements in a collection is supported.
- The Rborist package employs OpenMP pragmas to exploit predictor-level parallelism in th
  of which are performance bottlenecks in the algorithm.
- The h2o package connects to the h2o open source machine learning environment which h
- The randomForestSRC package can use both OpenMP as well as MPI for random forest e

Parallel computing: Implicit parallelism

- The pnmath package by Tierney (link) uses the Open MP parallel processing directives o
  make use of multiple cores --- without any explicit requests from the user. The alternate pr

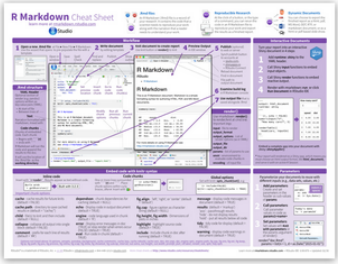
À ne pas manquer ! Les feuilles de triche de RStudio

2 feuilles sur des généralités

R Markdown Cheat Sheet

R Markdown is an authoring format that makes it easy to write reusable reports with R. You combine your R code with narration written in markdown (an easy-to-write plain text format) and then export the results as an html, pdf, or Word file. You can even use R Markdown to build interactive documents and slideshows. (Old Version).

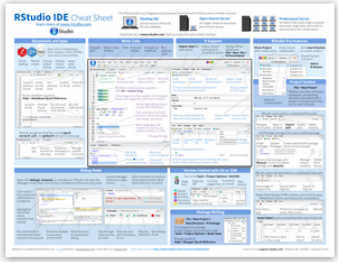
DOWNLOAD



RStudio IDE Cheat Sheet

The RStudio IDE is the most popular integrated development environment for R. Do you want to write, run, and debug your own R code? Work collaboratively on R projects with version control? Build packages or create documents and apps? No matter what you do with R, the RStudio IDE can help you do it faster. This cheat sheet will guide you through the most useful features of the IDE, as well as the long list of keyboard shortcuts built into the RStudio IDE.

DOWNLOAD



Shiny Cheat Sheet

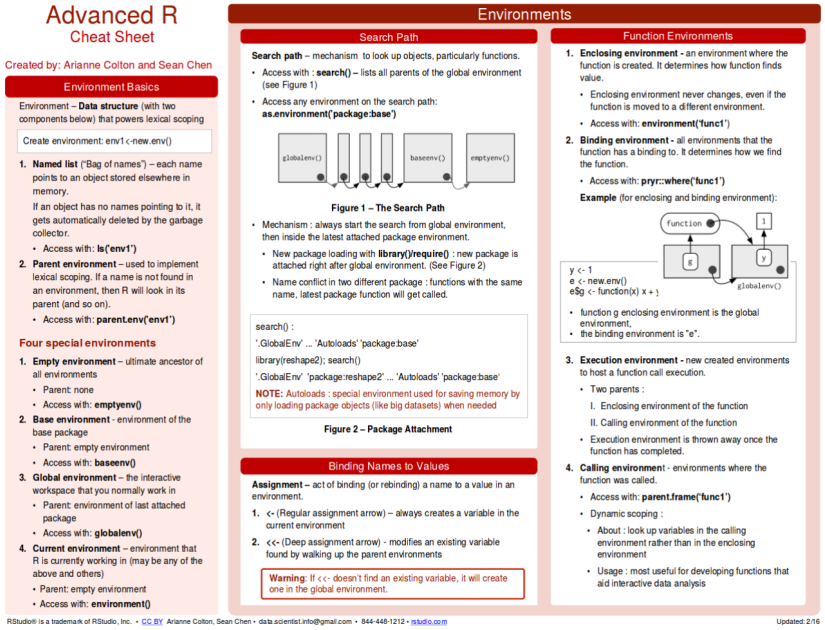
If you're ready to build interactive web apps with R, say hello to Shiny. This cheat sheet provides a tour of the Shiny package and explains how to build and customize an interactive app. Be sure to follow the links on the sheet for even more information.

DOWNLOAD

Data Visualization Cheat Sheet

The ggplot2 package lets you make beautiful and customizable plots of your data. It implements the grammar of graphics, an easy to use system for building plots. See docs.ggplot2.org for detailed examples.

DOWNLOAD



Advanced R
Cheat Sheet

Created by: Arianne Colton and Sean Chen

Environment Basics

Environment – **Data structure** (with two components below) that powers lexical scoping

Create environment: env1 <-new.env()

1. **Named list** ("Bag of names") – each name points to an object stored elsewhere in memory.
   If an object has no names pointing to it, it gets automatically deleted by the garbage collector.
   - Access with: **ls('env1')**
2. **Parent environment** – used to implement lexical scoping. If a name is not found in an environment, then R will look in its parent (and so on).
   - Access with: **parent.env('env1')**

Four special environments

1. **Empty environment** – ultimate ancestor of all environments
   - Parent: none
   - Access with: **emptyenv()**
2. **Base environment** - environment of the base package
   - Parent: empty environment
   - Access with: **baseenv()**
3. **Global environment** – the interactive workspace that you normally work in
   - Parent: environment of last attached package
   - Access with: **globalenv()**
4. **Current environment** – environment that R is currently working in (may be any of the above and others)
   - Parent: empty environment
   - Access with: **environment()**

Environments

Search Path

Search path – mechanism to look up objects, particularly functions.
- Access with : **search()** – lists all parents of the global environment (see Figure 1)
- Access any environment on the search path: **as.environment('package:base')**

Figure 1 – The Search Path

- Mechanism : always start the search from global environment, then inside the latest attached package environment.
  - New package loading with **library()/require()** : new package is attached right after global environment. (See Figure 2)
  - Name conflict in two different package : functions with the same name, latest package function will get called.

search() :
'.GlobalEnv' ... 'Autoloads' 'package:base'
library(reshape2); search()
'.GlobalEnv' 'package:reshape2' ... 'Autoloads' 'package:base'

NOTE: Autoloads : special environment used for saving memory by only loading package objects (like big datasets) when needed

Figure 2 – Package Attachment

Binding Names to Values

Assignment – act of binding (or rebinding) a name to a value in an environment.
1. **<-** (Regular assignment arrow) – always creates a variable in the current environment
2. **<<-** (Deep assignment arrow) - modifies an existing variable found by walking up the parent environments

Warning: If <<- doesn't find an existing variable, it will create one in the global environment.

Function Environments

1. **Enclosing environment** - an environment where the function is created. It determines how function finds value.
   - Enclosing environment never changes, even if the function is moved to a different environment.
   - Access with: **environment('func1')**
2. **Binding environment** - all environments that the function has a binding to. It determines how we find the function.
   - Access with: **pryr::where('func1')**

   Example (for enclosing and binding environment):

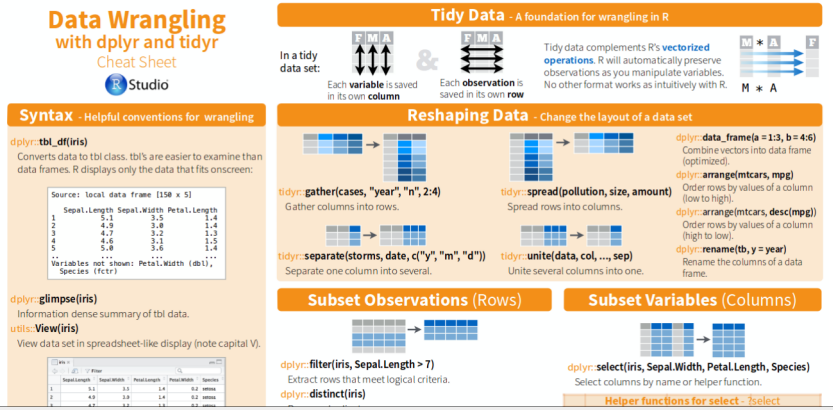
y <- 1
e <- new.env()
e$g <- function(x) x + 1

- function g enclosing environment is the global environment,
- the binding environment is "e".

3. **Execution environment** - new created environments to host a function call execution.
   - Two parents :
     I. Enclosing environment of the function
     II. Calling environment of the function
   - Execution environment is thrown away once the function has completed.
4. **Calling environment** - environments where the function was called.
   - Access with: **parent.frame('func1')**
   - Dynamic scoping :
     - About : look up variables in the calling environment rather than in the enclosing environment
     - Usage : most useful for developing functions that aid interactive data analysis

RStudio® is a trademark of RStudio, Inc. • CC BY Arianne Colton, Sean Chen • data.scientist.info@gmail.com • 844-448-1212 • rstudio.com Updated: 2/16

7 feuilles sur des outils de RStudio



Data Wrangling
with dplyr and tidyr
Cheat Sheet

Syntax - Helpful conventions for wrangling

dplyr::tbl_df(iris)
Converts data to tbl class. tbl's are easier to examine than data frames. R displays only the data that fits onscreen:

```
Source: local data frame [150 x 5]

   Sepal.Length Sepal.Width Petal.Length
1           5.1         3.5          1.4
2           4.9         3.0          1.4
3           4.7         3.2          1.3
4           4.6         3.1          1.5
5           5.0         3.6          1.4
..          ...         ...          ...
Variables not shown: Petal.Width (dbl),
  Species (fctr)
```

dplyr::glimpse(iris)
Information dense summary of tbl data.
utils::View(iris)
View data set in spreadsheet-like display (note capital V).

Tidy Data - A foundation for wrangling in R

In a tidy data set: Each **variable** is saved in its own **column** & Each **observation** is saved in its own **row**

Tidy data complements R's **vectorized operations**. R will automatically preserve observations as you manipulate variables. No other format works as intuitively with R.

Reshaping Data - Change the layout of a data set

tidyr::gather(cases, "year", "n", 2:4)
Gather columns into rows.

tidyr::spread(pollution, size, amount)
Spread rows into columns.

tidyr::separate(storms, date, c("y", "m", "d"))
Separate one column into several.

tidyr::unite(data, col, ..., sep)
Unite several columns into one.

dplyr::data_frame(a = 1:3, b = 4:6)
Combine vectors into data frame (optimized).
dplyr::arrange(mtcars, mpg)
Order rows by values of a column (low to high).
dplyr::arrange(mtcars, desc(mpg))
Order rows by values of a column (high to low).
dplyr::rename(tb, y = year)
Rename the columns of a data frame.

Subset Observations (Rows)

dplyr::filter(iris, Sepal.Length > 7)
Extract rows that meet logical criteria.
dplyr::distinct(iris)

Subset Variables (Columns)

dplyr::select(iris, Sepal.Width, Petal.Length, Species)
Select columns by name or helper function.

Helper functions for select - ?select

À ne pas manquer ! Bioconductor et ses vignettes...

edgeR

platforms some | downloads top 5% | posts 114 / 2 / 3 / 24 | in Bioc 7.5 years
build ok | commits 2.67 | test coverage 41%

This is the **development** version of edgeR; for the stable release version, see edgeR.

Empirical Analysis of Digital Gene Expression Data in R

Bioconductor version: Development (3.4)

Differential expression analysis of RNA-seq expression profiles with biological replication. Implements a range of statistical methodology based on the negative binomial distributions, including empirical Bayes estimation, exact tests, generalized linear models and quasi-likelihood tests. As well as RNA-seq, it be applied to differential signal analysis of other types of genomic data that produce counts, including ChIP-seq, SAGE and CAGE.

Author: Yunshun Chen <yuchen at wehi.edu.au>, Aaron Lun <alun at wehi.edu.au>, Davis McCarthy <dmccarthy at wehi.edu.au>, Xiaobei Zhou <xiaobei.zhou at uzh.ch>, Mark Robinson <mark.robinson at imls.uzh.ch>, Gordon Smyth <smyth at wehi.edu.au>

Maintainer: Yunshun Chen <yuchen at wehi.edu.au>, Aaron Lun <alun at wehi.edu.au>, Mark Robinson <mark.robinson at imls.uzh.ch>, Davis McCarthy <dmccarthy at wehi.edu.au>, Gordon Smyth <smyth at wehi.edu.au>

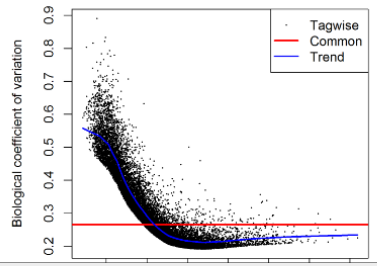
Citation (from within R, enter `citation("edgeR")`):

Extrait de la vignette (104 pages)

4.2.7 Estimating the dispersion

Estimate the genewise dispersion estimates over all genes, allowing for a possible abundance trend. The estimation is also robustified against potential outlier genes.

```
> y <- estimateDisp(y, design, robust=TRUE)
> y$common.dispersion
[1] 0.0705
> plotBCV(y)
```



À ne pas manquer ! Bioconductor et ses outils...

Browse packages by topic

All Packages

Bioconductor version 3.3 (Release)

Autocomplete biocViews search:

- Software (1213)
  - AssayDomain (435)
    - aCGH (16)
    - CellBasedAssays (50)
    - ChIPchip (8)
    - CopyNumberVariation (60)
    - CpGIsland (8)
    - DNAMethylation (50)
    - ExonArray (8)
    - GeneExpression (209)
    - GeneticVariability (30)
    - SNP (57)
    - Transcription (75)
  - BiologicalQuestion (408)
  - Infrastructure (263)
  - ResearchField (298)
  - StatisticalMethod (365)

Packages found under DNAMethylation:

Show All entries Search table:

Package	Maintainer	Title
ARRmNormalization	Jean-Philippe Fortin	Adaptive Robust Regression normalization for Illumina methylation data
BEAT	Kemal Akman	BEAT - BS-Seq Epimutation Analysis Toolkit
BEclear	Markus Merl	Correct for batch effects in DNA methylation data
BiSeq	Katja Hebestreit	Processing and analyzing bisulfite sequencing data
bsseq	Kasper Daniel Hansen	Analyze, manage and store bisulfite sequencing data
bumphunter	Rafael A. Irizarry	Bump Hunter
ChAMP	Yuan Tian, Tiffany Morris	Chip Analysis Methylation Pipeline for Illumina HumanMethylation450 and EPIC
charm	Peter Murakami	Analysis of DNA methylation data from CHARM microarrays
COHCAP	Charles Warden	CpG Island Analysis Pipeline for Illumina Methylation Array and Targeted BS-Seq Data
coMET	Tiphaine Martin	coMET: visualisation of regional epigenome-wide association scan (EWAS) results and DNA co-methylation patterns
conumee	Volker Hovestadt	Enhanced copy-number variation analysis using Illumina 450k methylation arrays
CopyNumber450k	Simon Papillon-Cavanagh	R package for calling CNV from Illumina 450k methylation microarrays
DMRcaller	Nicolae Radu Zabet	Differentially Methylated Regions caller
DMRcate	Tim Peters	Methylation array and sequencing spatial analysis methods
DMRforPairs	Martin Rijlaarsdam	DMRforPairs: identifying Differentially Methylated Regions between unique samples using array based methylation profiles

Explore workflows

Basic Workflows

- Sequence Analysis Import fasta, fastq, BAM, gff, bed, wig, and other sequence formats. Trim, transform, align, and manipulate sequences. Perform quality assessment, ChIP-seq, differential expression, RNA-seq, and other workflows. Access the Sequence Read Archive.
- Oligonucleotide Arrays Import Affymetrix, Illumina, Nimblegen, Agilent, and other platforms. Perform quality assessment, normalization, differential expression, clustering, classification, gene set enrichment, genetical genomics and other workflows for expression, exon, copy number, SNP, methylation and other assays. Access GEO, ArrayExpress, Biomart, UCSC, and other community resources.
- Annotation Resources Introduction to using gene, pathway, gene ontology, homology annotations and the AnnotationHub. Access GO, KEGG, NCBI, Biomart, UCSC, vendor, and other sources.
- Annotating Genomic Ranges Represent common sequence data types (e.g., from BAM, gff, bed, and wig files) as genomic ranges for simple and advanced range-based queries.
- Annotating Genomic Variants Read and write VCF files. Identify structural location of variants and compute amino acid coding changes for non-synonymous variants. Use SIFT and PolyPhen database packages to predict consequence of amino acid coding changes.
- Changing genomic coordinate systems with rtracklayer::liftOver The liftOver facilities developed in conjunction with the UCSC browser track infrastructure are available for transforming data in GRanges formats. This is illustrated here with an image of the NHGRI GWAS catalog that is, as of Oct. 31 2014, distributed with coordinates defined by NCBI build hg38.

Advanced Workflows

- High Throughput Assays Import, transform, edit, analyze and visualize flow cytometric, mass spec, HTqPCR, cell-based, and other assays.
- RNA-Seq workflow: gene-level exploratory analysis and differential expression This lab will walk you through an end-to-end RNA-Seq differential expression workflow, using DESeq2 along with other Bioconductor packages. We will start from the FASTQ files, show how these were aligned to the reference genome, prepare gene expression values as a count matrix by counting the sequenced fragments, perform exploratory data analysis (EDA), perform differential gene expression analysis with DESeq2, and visually explore the results.
- Mass spectrometry and proteomics This lab demonstrates how to access data from proteomics data repositories, how to parse various mass spectrometry data formats, how to identify MS2 spectra and

RNA-seq workflow: gene-level exploratory analysis and differential expression

Michael Love [1], Simon Anders [2,3], Vladislav Kim [3], Wolfgang Huber [3]

[1] Department of Biostatistics, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, US;

[2] Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland;

[3] European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.

Short links

- Counting reads
- Building a DESeqDataSet
- Exploratory analysis and visualization
- Differential expression
- Plotting results
- Annotating and exporting results
- Accounting for unknown batches
- Time course experiments
- Session information and references

Abstract

À ne pas manquer ! Des blogs...

573 bloggers vous donnent les dernières news

R-bloggers

R news and tutorials contributed by (573) R bloggers

Home | About | RSS | add your blog! | Learn R | R jobs | Contact us

WELCOME!

Follow @rbloggers 32.3K

Here you will find daily **news and tutorials about R**, contributed by over 573 bloggers. There are many ways to **follow us -**
By e-mail:

Your e-mail here
Subscribe
30139 readers
BY FEEDBURNER

On Facebook:

R blogg...
34K likes

Liked

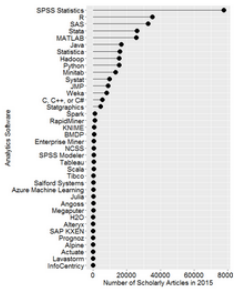
You and 6 other friends like this

If you are an R blogger yourself you are invited to add your own R content feed to this site
(**Non-English** R bloggers should add themselves- here)

JOBS FOR R-USERS

R Passes SAS in Scholarly Use (finally)

June 7, 2016
By muenchen.bob@gmail.com

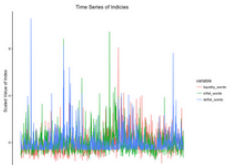


Way back in 2012 I published a forecast that showed that the use of R for scholarly publications would likely pass the use of SAS in 2015. But I didn't believe the forecast since I expected the sharp decline in SAS ... Continue reading →

Read more »

Economist Liquidity Index

June 7, 2016
By Sam Weiss



Analyze Face Emotions with R

June 7, 2016
By Daniel Boller

Search & Hit Enter

RECENT POPULAR POSTS

- What are the Best Machine Learning Packages in R?
- Introduction to R for Data Science :: Session 6 [Linear Regression Model in R + EDA, and Normality Tests]
- Predictive Bookmaker Consensus Model for the UEFA Euro 2016

MOST VISITED ARTICLES OF THE WEEK

1. How to write the first for loop in R
2. Installing R packages
3. R tutorials
4. Predictive Bookmaker Consensus Model for the UEFA Euro 2016
5. In-depth introduction to machine learning in 15 hours of expert videos
6. What are the Best Machine Learning Packages in R?
7. Using apply, sapply, lapply in R
8. How to perform a Logistic Regression in R
9. Bootstrap and cross-validation for evaluating modelling strategies

SPONSORS

EARL

5 search results for "R2D2"

R code coverage support via docker

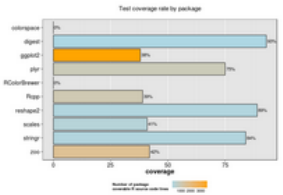
December 17, 2014
By R2D2

R code coverage support via docker 2014-12-17 Source In my previous post , I used a R with built-in code coverage to compute the code coverage of some packages. Today I will show you how to install and use such a R with code coverage support. I just created a public docker container that provides this patched R. If you do not...

Read more »

Test coverage of the 10 most downloaded R packages

May 2, 2014
By R2D2



Test coverage of the 10 most downloaded R packages 2014-04-30 Source Introduction How do you know that your code is well tested ? The test coverage is the proportion of source code lines that are executed (covered) when running the tests. It is useful to find the parts of your code that are no exercised no matter how...

Read more »

An example of monkey patching a package

August 1, 2013
By R2D2

À ne pas manquer ! Un grand Wiki...

R Programming

The latest reviewed version was checked on *15 October 2015*. There are template/file changes awaiting review.

Editor's note
This book is still under development. Please help us

Welcome to the *R programming* Wikibook

This book is designed to be a practical guide to the **R programming language**[1].

R is free software designed for statistical computing. There is already great documentation for the standard R packages on the Comprehensive R Archive Network (CRAN)[2] and many resources in specialized books, forums such as Stackoverflow[3] and personal blogs[4], but all of these resources are scattered and therefore difficult to find and to compare. The aim of this Wikibook is to be the place where anyone can share his or her knowledge and tricks on **R**. It is supposed to be organized by task but not by discipline[5]. We try to make a cross-disciplinary book, i.e. a book that can be used by all people applying statistics to some specific fields.

How can you share your R experience ? [edit]

- Explain the syntax of a command
- Compare the different ways of performing each task using **R**.
- Try to make unique examples based on fake data (ie simulated data sets).
- As with any Wikibook please feel free to make corrections, expand explanations, and make additions where necessary.

Some rules :

- Put the name of functions in the code environment : `<code>lm()</code>`
- Put the name of packages in bold : `'''stats'''`
- Except for the default packages, put the name of the package in parentheses after each function : <code>MCMCprobit()</code> ('''MCMCpack''')
- Put examples in the source environment with argument `lang="rsplus"` : `<source lang="rsplus"></source>`
- Use references to package documentations, academic literature and wikipedia.
- Use the citations templates to make citations : Template:Cite book, Template:Cite web, Template:Cite journal
- If you want to add a graph, you should load it on Commons and add the tag `{{Created with R}}` and, if it is applicable, `[[Category:Images with R source code]]`.

Prerequisites [edit]

We assume that readers have a background in statistics. This book is not a book about statistics but a book about how to implement statistical methods using **R**. We try to use terms which are already defined on Wikipedia such that people can refer to the corresponding wikipedia page each time they have some doubts on a notion.

We also assume that readers are familiar with computers and that they know how to use software with a command-line interface. There are some graphical user interfaces for **R** but we are not going to explain how to use them in this textbook. Beginners should have a look at the Sample session for a first session with **R**. They can also have a look at the Statistical Analysis: an Introduction using R book.

See also [edit]

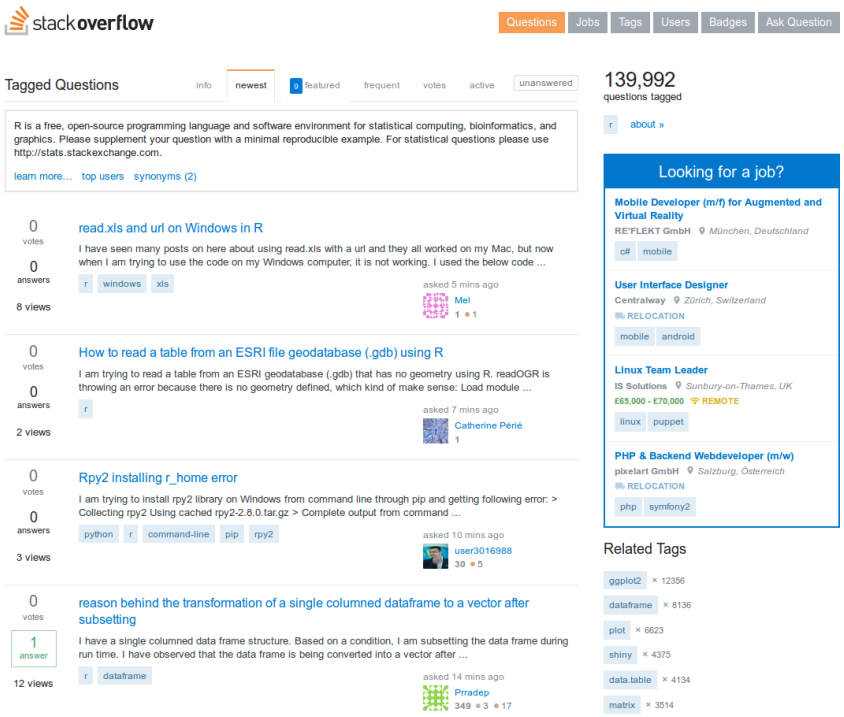
https://en.wikibooks.org/wiki/R_Programming

R Programming

- **R Basics**
  - Introduction
  - Sample Session
  - Manage your workspace
  - Settings
  - Documentation
  - Control Structures
  - Working with functions
  - Debugging
  - Using C or Fortran
  - Utilities
  - Estimation utilities
  - Packages
- **Data Management**
  - Data types
  - Working with data frames
  - Importing and exporting data
  - Text Processing
  - Times and Dates
- Graphics
  - Grammar of graphics
- Publication quality output
- Descriptive Statistics
- Mathematics
  - Optimization
  - Probability Distributions
  - Random Number Generation
- **Statistical Core Methods**
  - Maximum Likelihood
  - Method of Moments
  - Bayesian Methods
  - Bootstrap
  - Multiple Imputation
  - Nonparametric Methods
- **Regression Models**
  - Linear Models
  - Quantile Regression
  - Binomial Models
  - Multinomial Models
  - Tobit And Selection Models
  - Count Data Models

À l'aide !!!! Avec de vrais gens...

En anglais



http://stackoverflow.com/questions/tagged/r

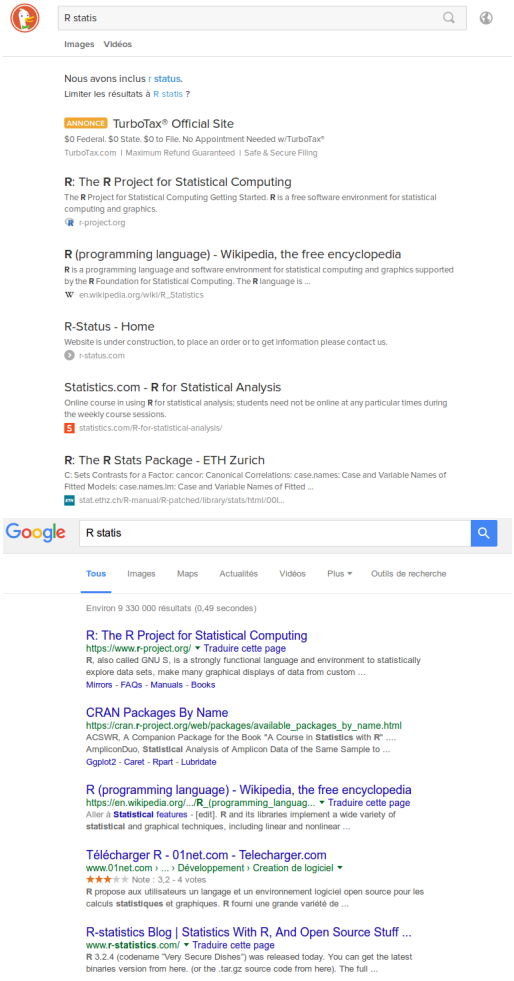
... ou en français

http://forums.cirad.fr/logiciel-R

À l'aide !!!! GI(not)YF

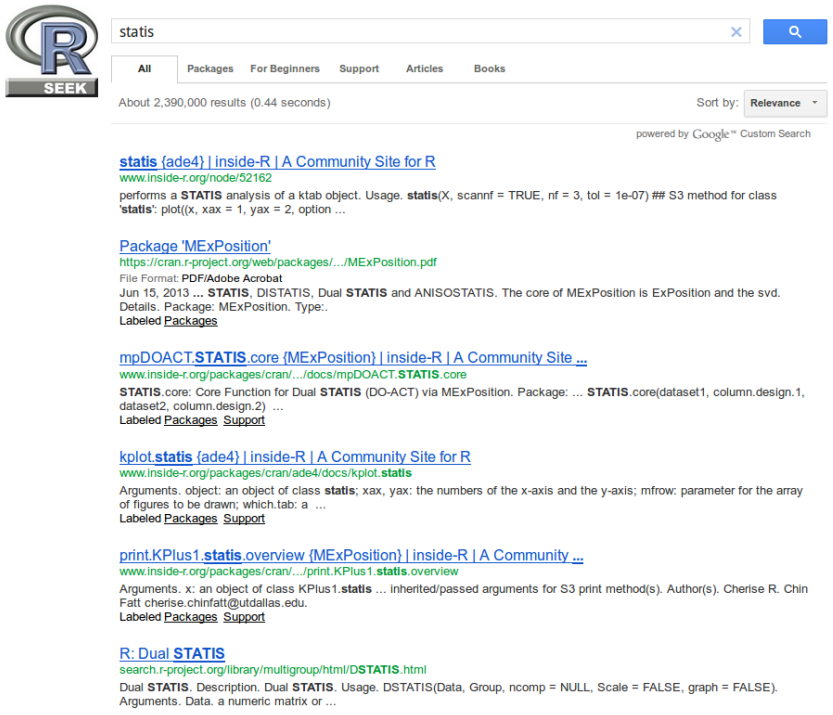
Une bête étude comparée...

avec mon canard préféré



et son concurrent le plus célèbre

contre leur concurrent spécialisé



http://rseek.org

Conclusion

Cinquièmes Rencontres R

Toulouse
du 22 au 24 juin 2016

Lightning talks

Nathalie Villa-Vialaneix, INRA, UR875, MIAT
Sébastien Déjean, IMT, Toulouse.
Christophe Bontemps, Toulouse School of Economics -INRA

Yan Holtz, Centre international d'études supérieures en sciences agronomiques

Cinquièmes Rencontres R
Toulouse
du 22 au 24 juin 2016

Lightning talks

The R Graph Gallery: une core-collection de graphiques R

Yan Holtz, Centre international d'études supérieures en sciences agronomiques

The R Graph Gallery

Rencontres R, Toulouse, June 2016

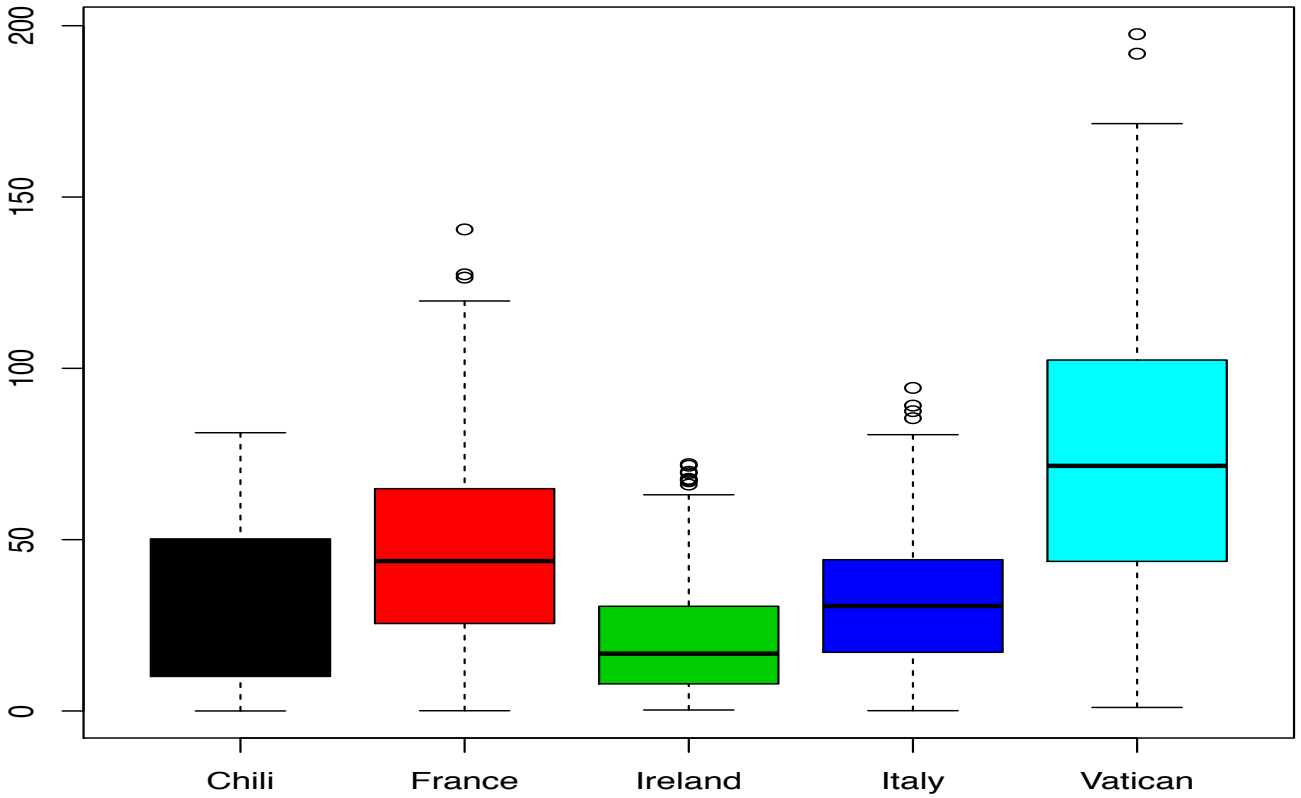
Yan Holtz, holtz@supagro.fr

A quick basic graph

names	country	drink
Chelsea	France	45.629739
Marissa	Ireland	5.906875
Mara	France	36.089158
A'Tiana	Chili	13.052119
...	...	...
Fardowsa	Ireland	5.573123
Makayla	Vatican	83.58251

+

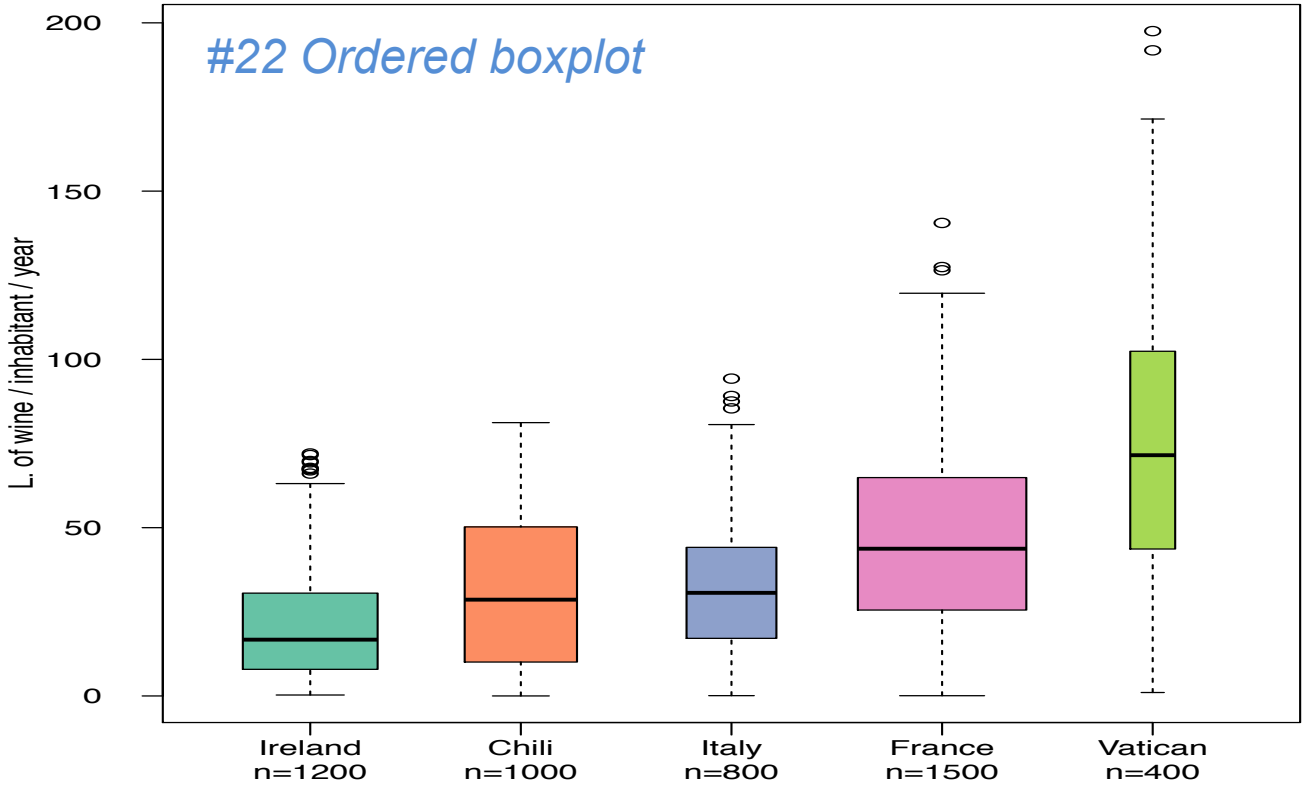
```
boxplot(data$drink ~ data$country, col=c(1:5) )
```



- R = high efficiency to produce simple graphics.
- But more work is needed for an attractive communication

… the custom version!

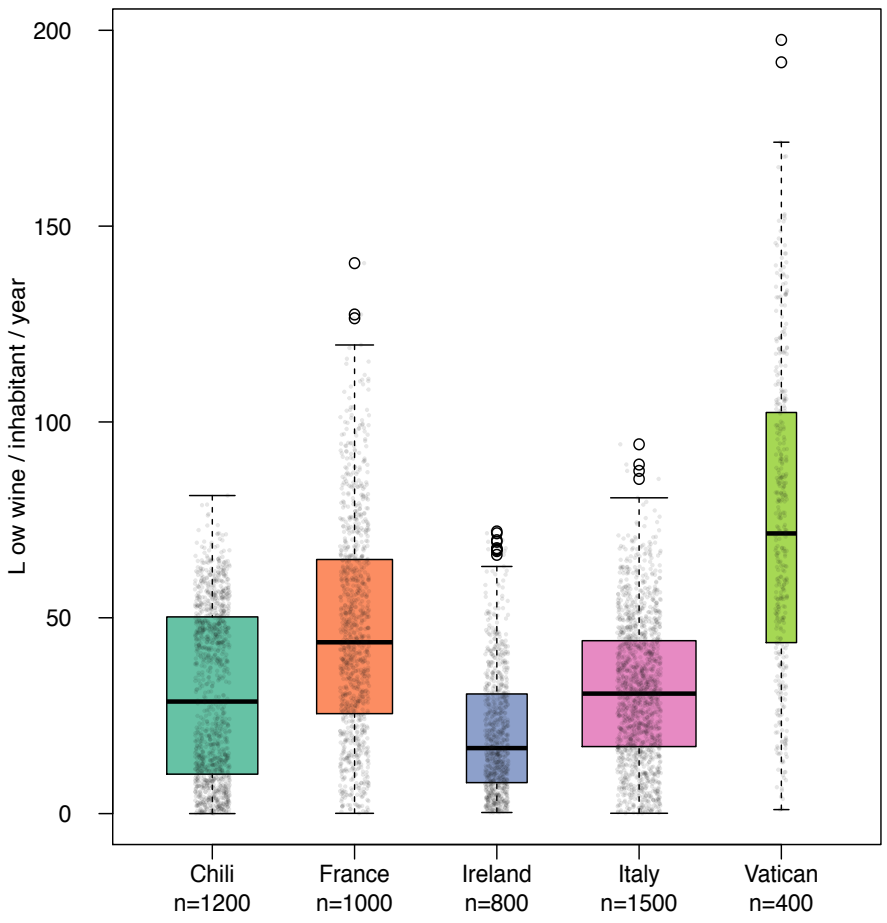
```
library("RColorBrewer")
par(mgp=c(3,1.5,0) , mar=c(4,5,2,2) )
proportion=table(data$country)/nrow(data)
new_order <- with(data, reorder(country , drink, median))
boxplot(data$drink ~ new_order  , ylab="L. of wine / inhabitant / year" , col=brewer.pal(5, "Set2") , las=1 ,
width=proportion, xaxt="n" )
axis(1, at=c(1:5) , labels=paste(levels(data$country),table(data$country),sep="\nn=")  )
```



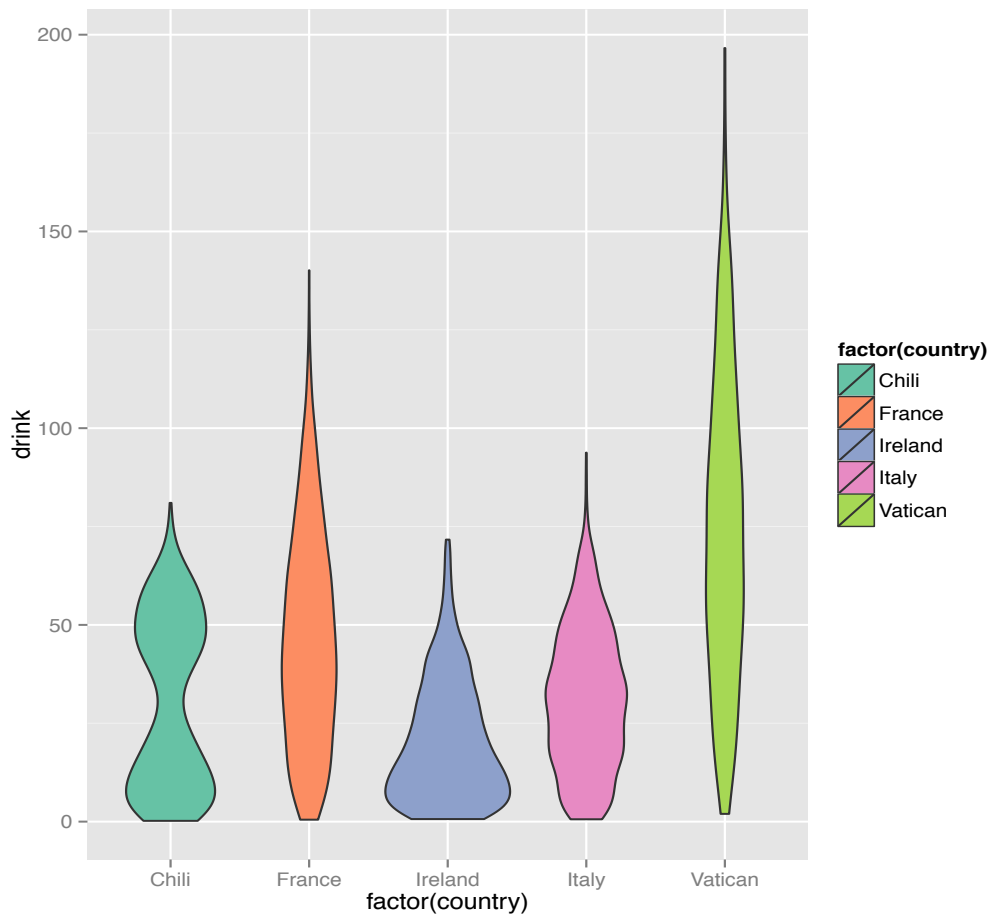
- Better graph=better understanding
- Hard to remember every option!

Dataviz is a science

#96 Boxplot with jitter

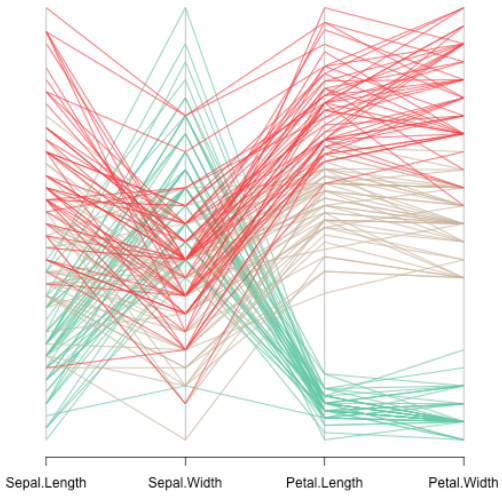


#95 Violin Plot

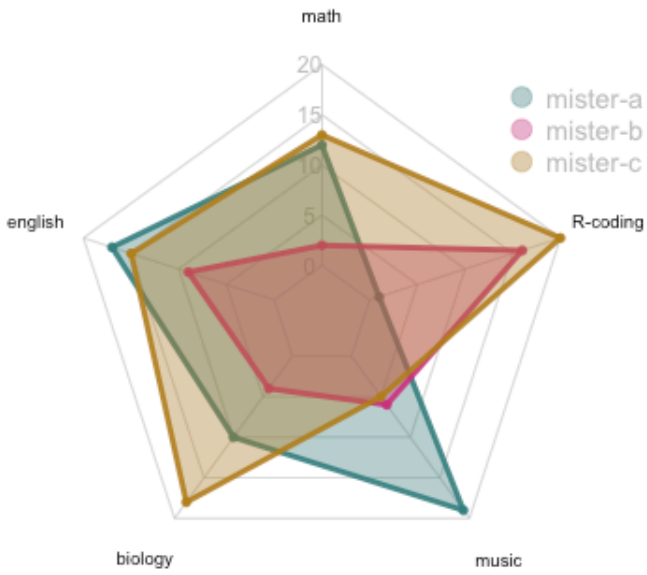


- DataViz is not only aesthetic!
- Do you know every chart possibility?

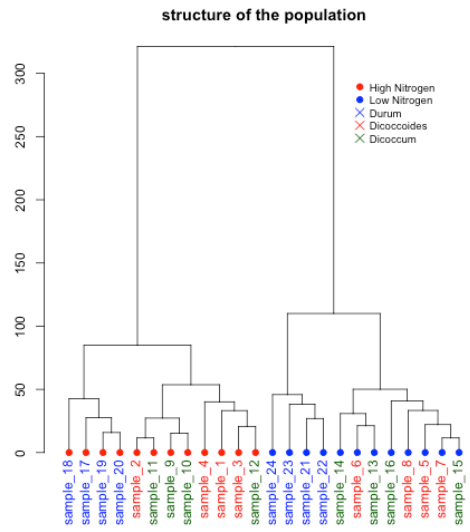
A world of possibilities



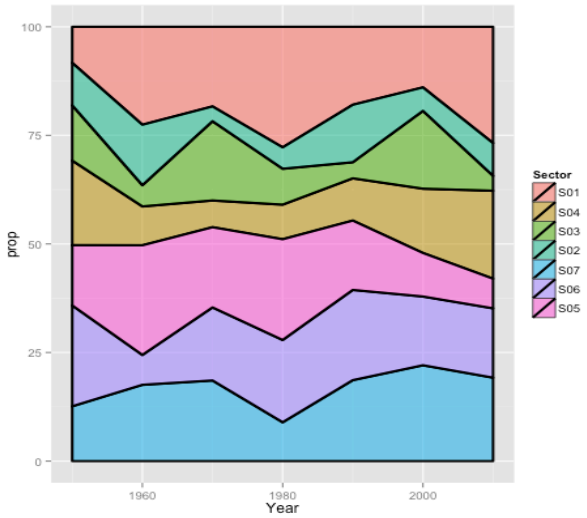
#93 Parallel plot



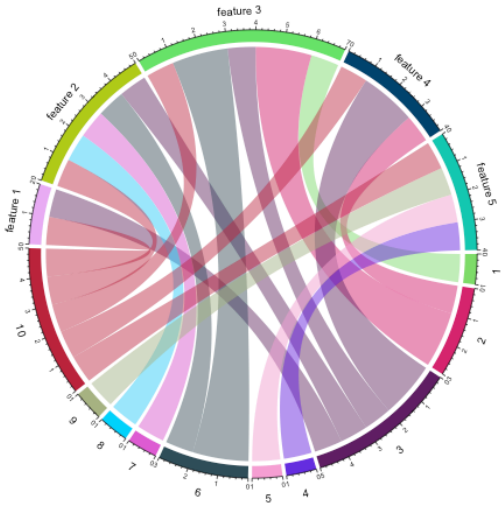
#143 Spider Graph



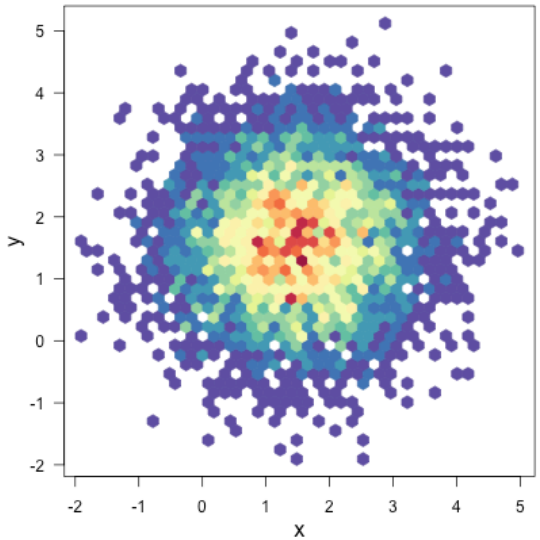
#31 Clustering tree



#136 Stacked area



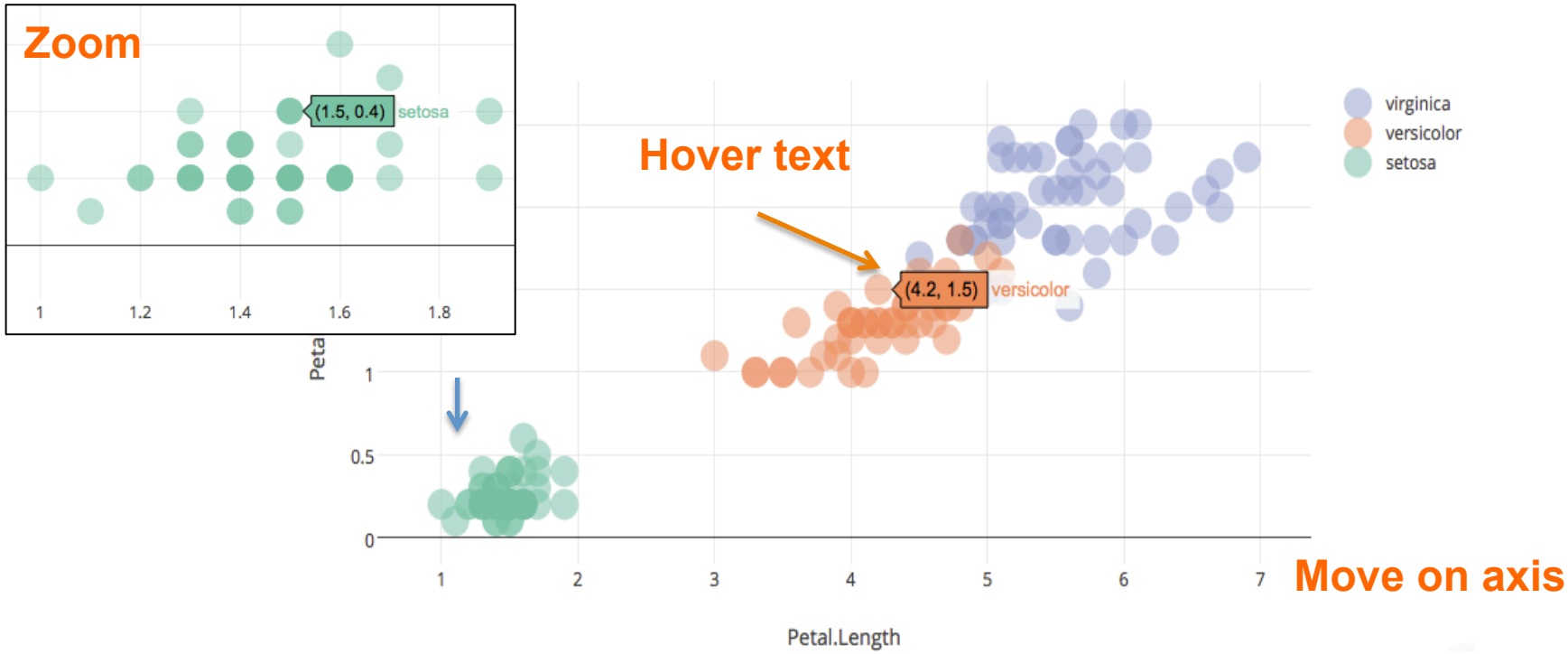
#123 Circular plot



#100 scatter & binning

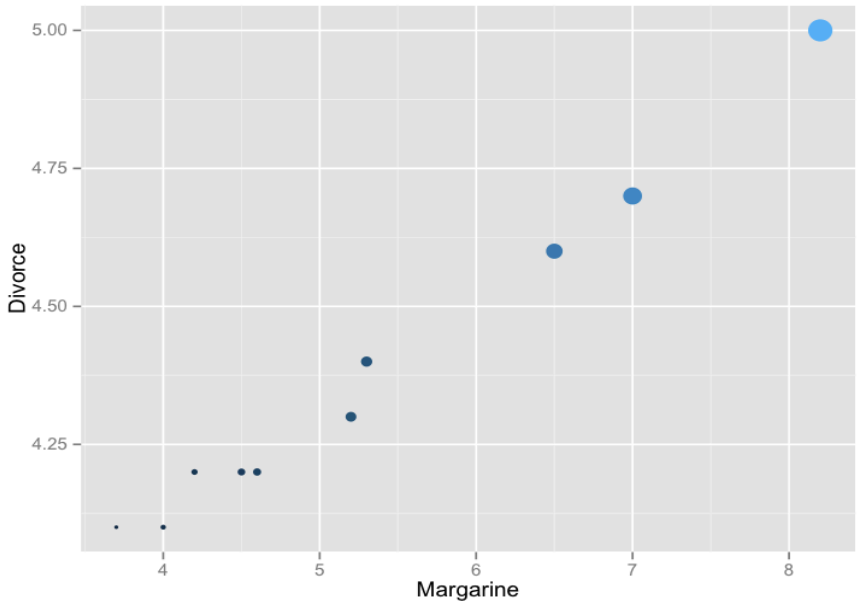
Next step: Interactive plots with the HTML widgets

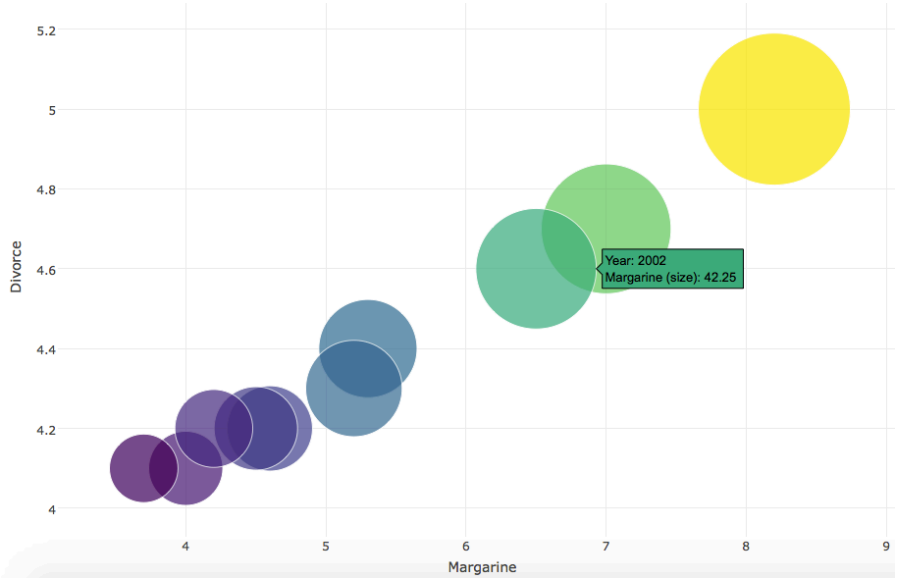
HTML widgets: Dygraph - Leaflet - Plotly - rbokeh - visNetwork - networkD3 - DataTables - Threejs - Rglwidget - DiagrammeR - HighCharter - MetricsGraphics – www.htmlwidgets.org/



- Interactivity=better access to information
- Bring your data to life!
- Interactivity with R is easy

Plotly






```
qplot(
        data=data, x=Margarine, y=Divorce,
        colour = Margarine,
        size=Margarine
)
```

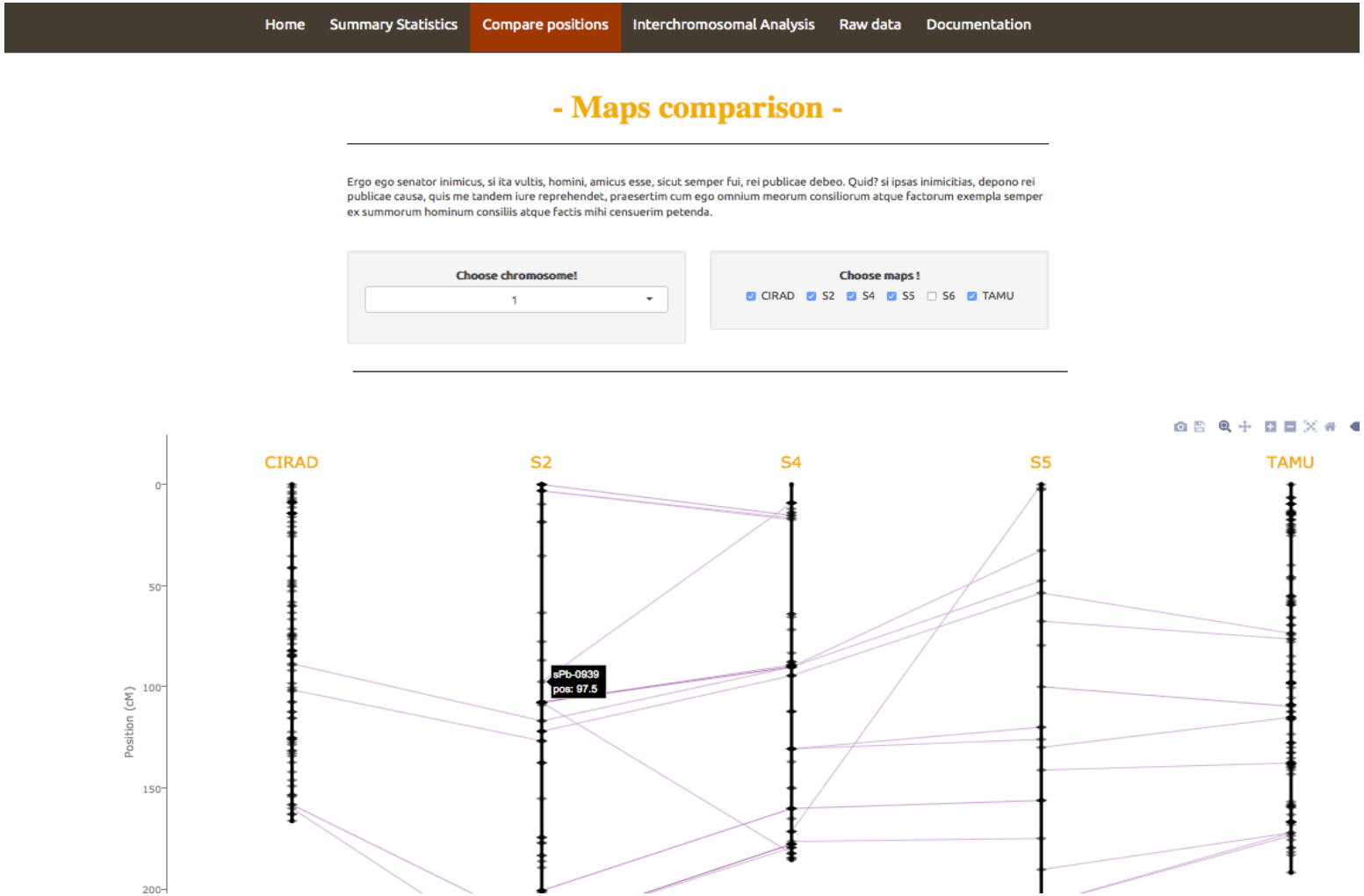
```
plot_ly(
        data, x=Margarine , y=Divorce ,
        size=Margarine,
        color=Margarine,
        mode="markers" ,
        text=paste("Year: ",year,sep="") ,
        hoverinfo="text”
)
```

- Plotly : syntax is close to ggplot2
- Possibility to share your graphics

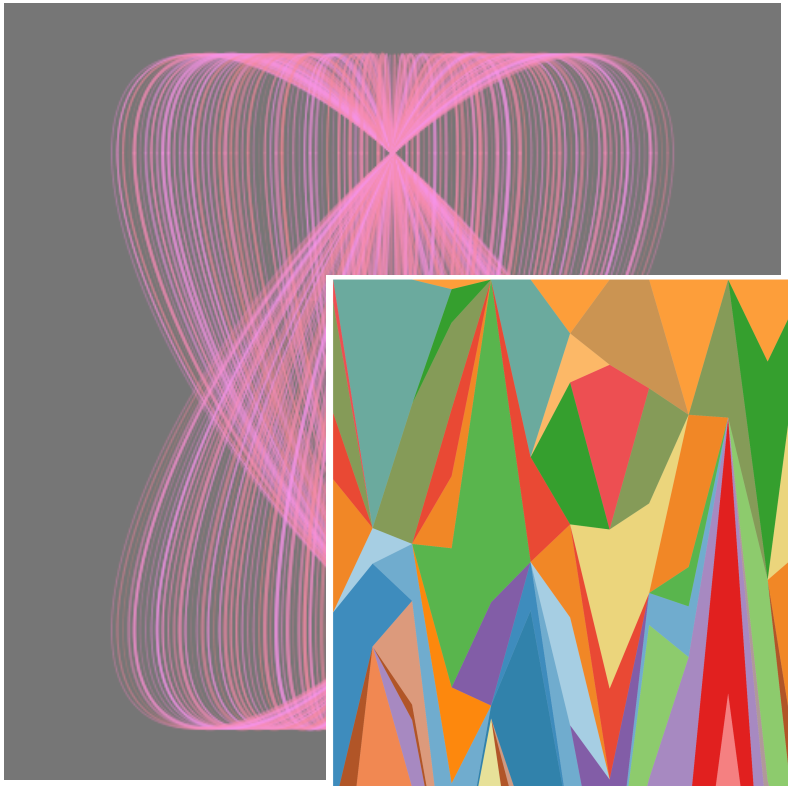
http://www.tylervigen.com/spurious-correlations

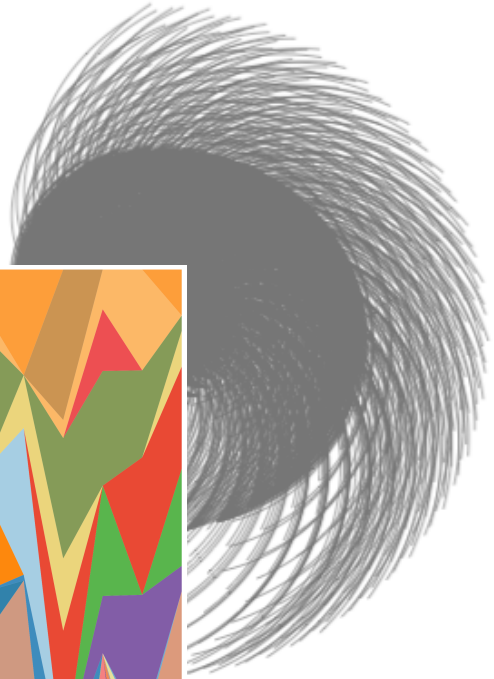
Shiny

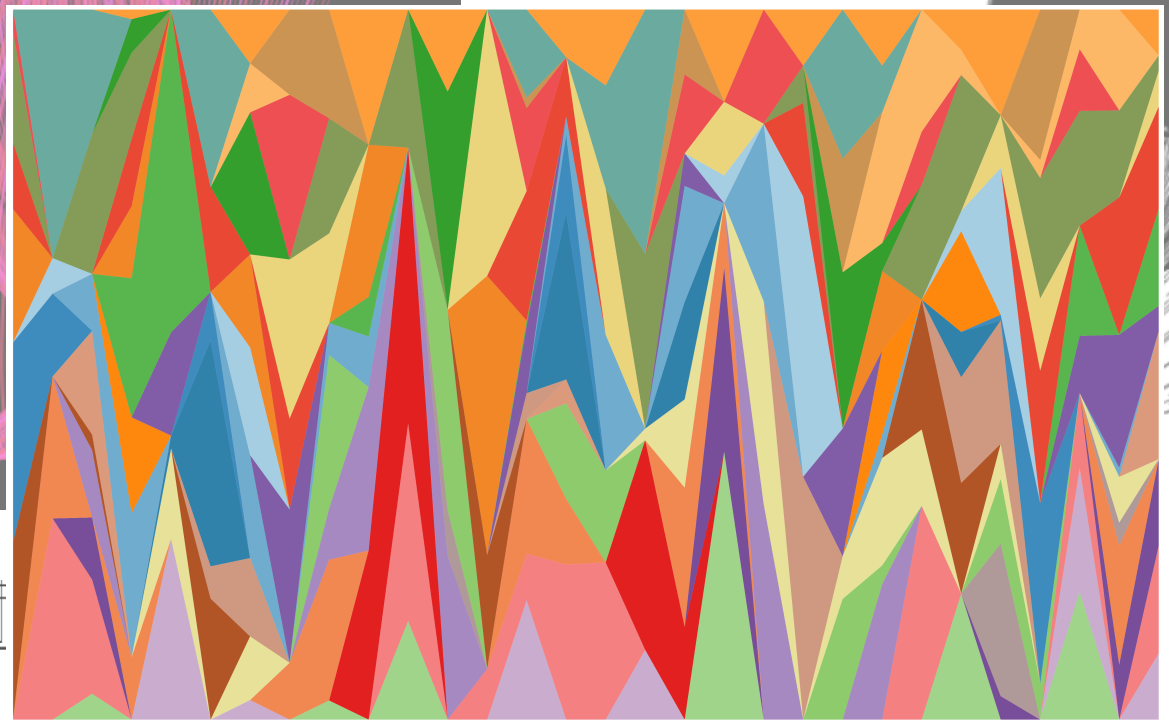
- Shiny: Turn your R analyses into interactive web applications
- http://shiny.rstudio.com/
- Combine the functionality of shiny, the aesthetics of ggplot2 & the interactivity of plotly!

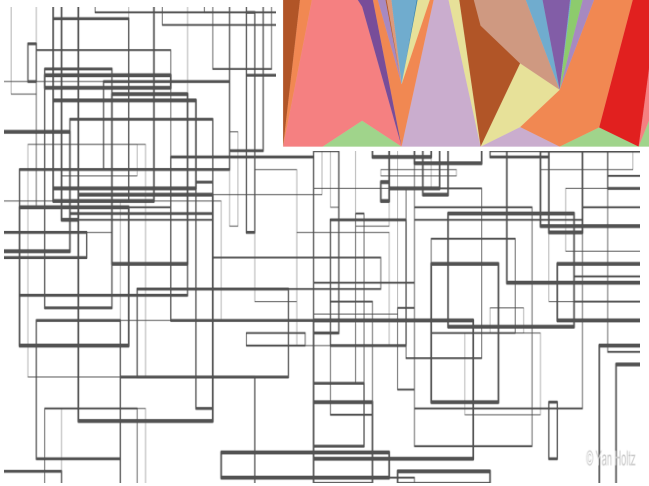


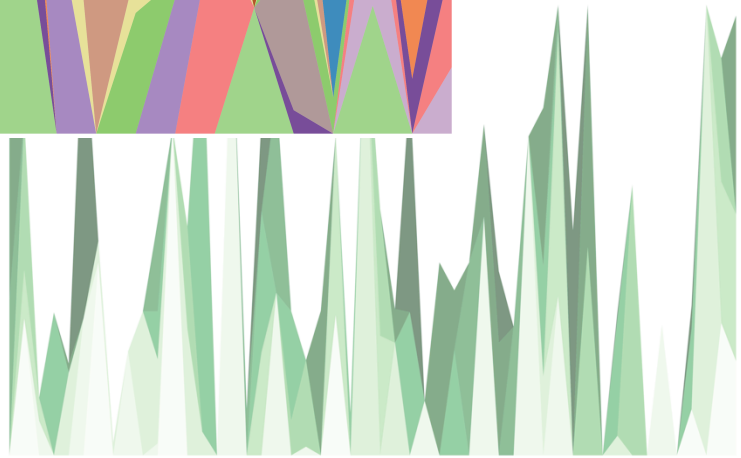
From Dataviz to DataArt



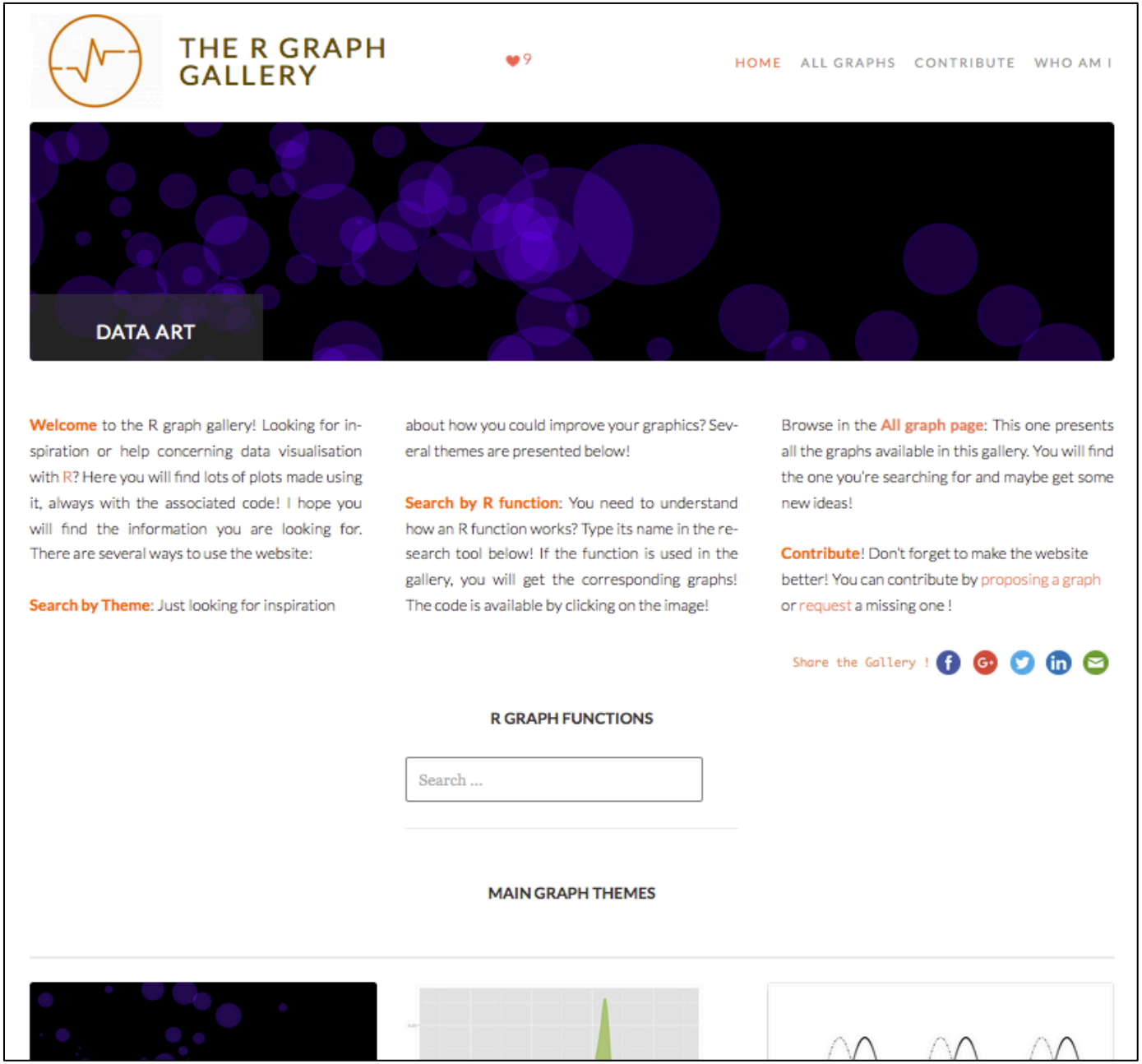








Welcome page of the gallery

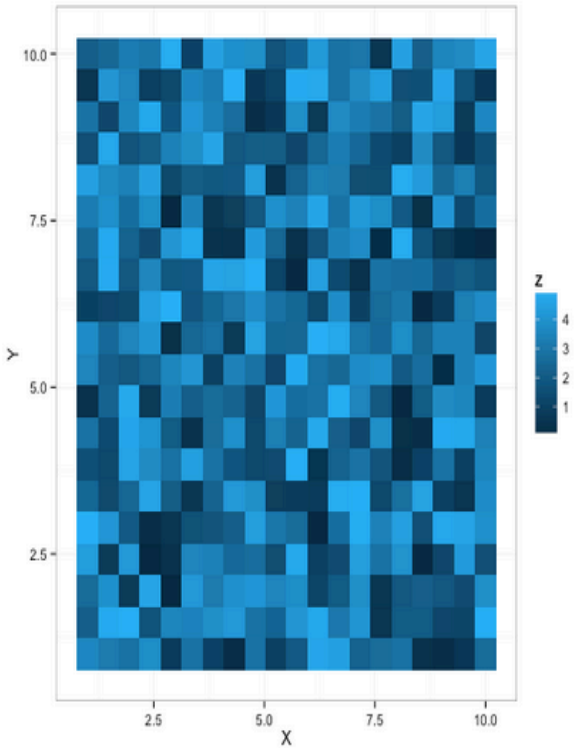


- >170 charts
- Organized by theme
- Search bar for function or graph type

www.r-graph-gallery.com

Example of a graphic page

- Reproducible code
- Link to relevant libraries, functions and related graphics
- Each graph targets a specific topic



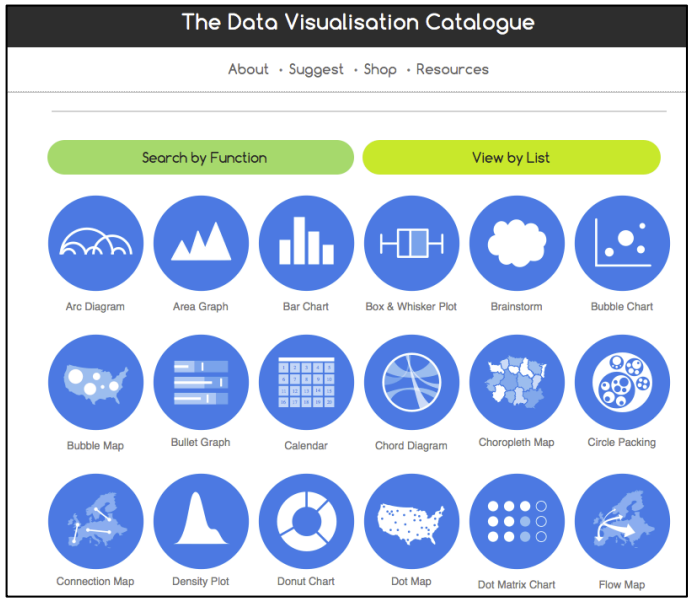
This is a levelplot made using the ggplot2 library. Each square represents a value of the input matrix.

You can also draw levelplot without ggplot2. Have a look to graph #78 if you have a square matrix, or to graph #27 if your data are organized in 3 columns !

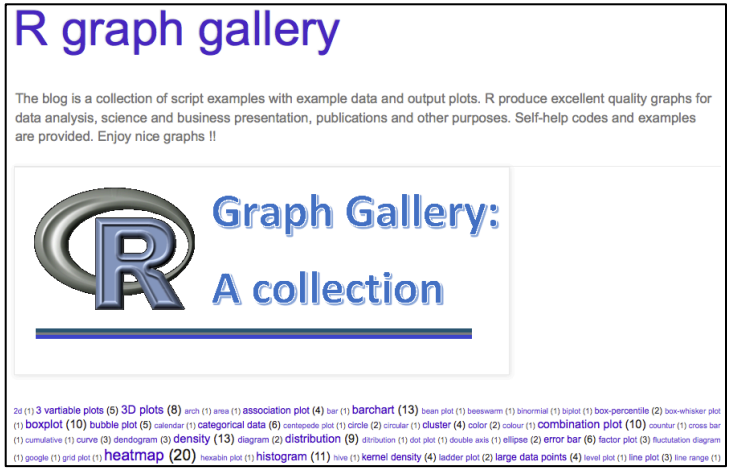
```
## Example data
x <- seq(1,10, length.out=20)
y <- seq(1,10, length.out=20)
data <- expand.grid(X=x, Y=y)
data$Z <- runif(400, 0, 5)

# Levelplot with ggplot2
library(ggplot2)
ggplot(data, aes(X, Y, z= Z)) + geom_tile(aes(fill = Z)) + theme_bw()

# To change the color of the gradation :
ggplot(data, aes(X, Y, z= Z)) + geom_tile(aes(fill = Z)) +
    theme_bw() +
    scale_fill_gradient(low="white", high="blue")
```

http://www.r-graph-gallery.com/79-levelplot-with-ggplot2/

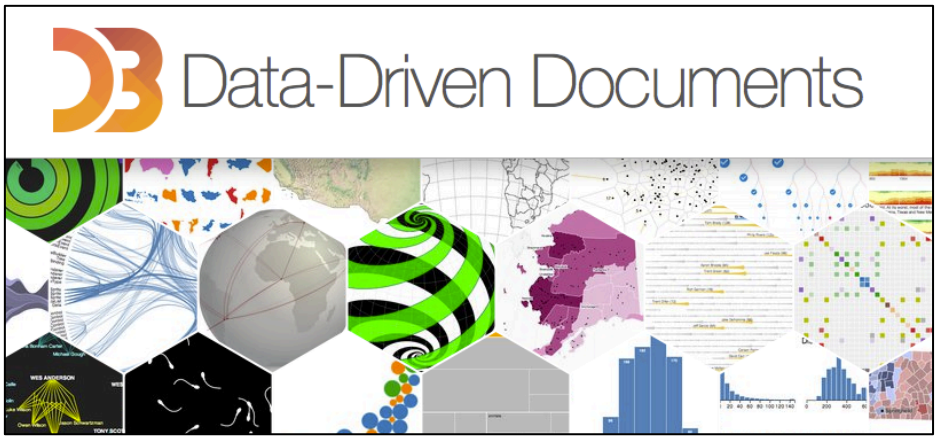
Other resources



http://www.datavizcatalogue.com/



http://rgraphgallery.blogspot.fr/



https://d3js.org/

The R Graph Gallery

Rencontres R, Toulouse, June 2016

Yan Holtz, holtz@supagro.fr

DataViz is a key step in data science

Dataviz is not trivial

R = amazing graphing tool

Get inspired & contribute!

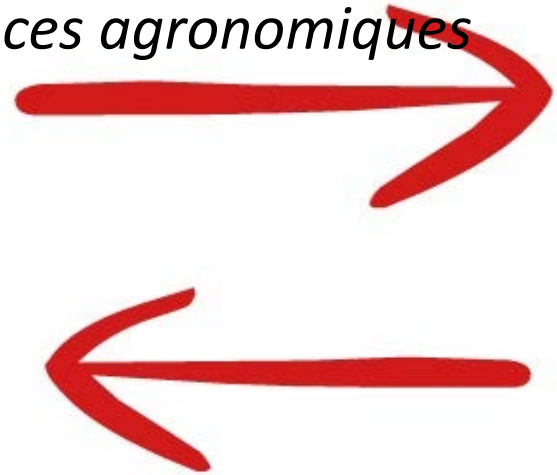
Cinquièmes Rencontres R

Toulouse

du 22 au 24 juin 2016

Lightning talks

Yan Holtz, Centre international d'études supérieures en sciences agronomiques



Géraud DUGE de BERNONVILLE, Valtech

Cinquièmes Rencontres R

Toulouse

du 22 au 24 juin 2016

Lightning talks

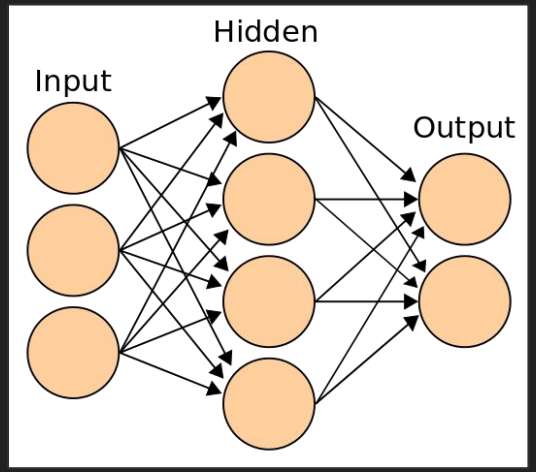
Deep Learning avec R et H2O

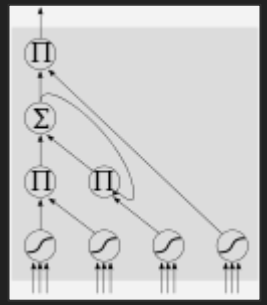
Géraud DUGE de BERNONVILLE, Valtech

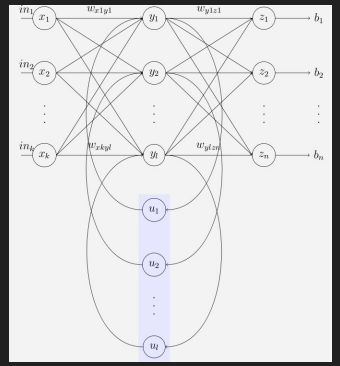
Deep learning avec R et H2O

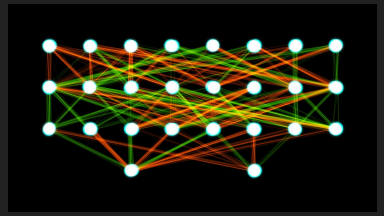
Géraud Dugé de Bernonville
@geraudster

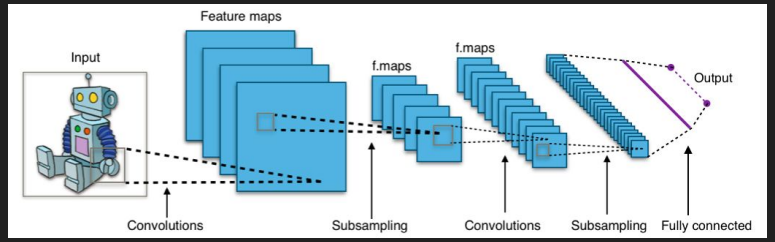
Deep Learning Origins











Utilisation

Régression
Classification
Détection d'anomalie
Détection d'objet
Traitement du langage
Feature extraction
Réduction de dimensions
...

Analyse statistique
+ Machine Learning
+ Distribué
+ Open source
= **H2O**

H_2O.ai

Mise en oeuvre
@Drivendata.org

C'est parti

```
h2o.init(min_mem_size = '5G', nthreads = 4)
```

Data

```
trainset <- h2o.uploadFile('trainset_values.csv')

summary(trainset$status_group)

status_group
 functional              :32259
 non functional          :22824
 functional needs repair: 4317
```

Tentative 1

```
h2o.deeplearning(predictors, 'status_group', trainset)
```

Score:
0.7801
Rank:
259/1665

Paramètres

Hidden

Epochs

L1

L2

Activation

Dropout ratio

....

?

Grid

```
h2o.grid('deeplearning', x = predictors, y = 'status_group',
         training_frame = trainset,
         hyper_params = list(
                            hidden = list(rep(100,2), rep(200,2), rep(300,2), ...),
                            activation = c('Rectifier','RectifierWithDropout'),
                            l1 = c(0, 1e-5)))
```

Score:
0.7931
Rank:
234/1665

Et aussi dans H2O

Autoencoder, Variable Importance
PCA, K-means, Linear models
Gradient Boosting, Random Forests

Et aussi !

Scalabilité (Hadoop, Spark)

H2O Flow

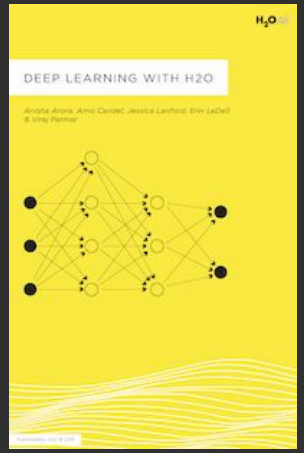
Export de modèles Java

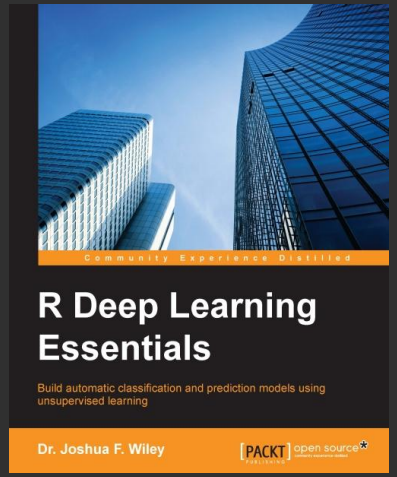
...

En savoir plus...

http://h2o.ai
R Deep Learning Essentials

@geraudster





Cinquièmes Rencontres R

Toulouse
du 22 au 24 juin 2016

Lightning talks

Géraud DUGE de BERNONVILLE, Valtech

Nicolas Turenne, laboratoire interdisciplinaire science innovation société

Cinquièmes Rencontres R

Toulouse

du 22 au 24 juin 2016

Lightning talks

Groupe de Travail « Analyse des données textuelles sous R »

Nicolas Turenne, laboratoire interdisciplinaire science innovation société

LISIS LABORATOIRE INTERDISCIPLINAIRE
Sciences Innovations Sociétés

Groupe d'échange Text Data Analytics

Nicolas Turenne

Rencontres R – Toulouse – 22-24 Juin 2016

LISIS LABORATOIRE INTERDISCIPLINAIRE
Sciences Innovations Sociétés

Groupe d'échange Text Data Analytics

Nicolas Turenne

Rencontres R – Toulouse – 22-24 Juin 2016

Cet ouvrage est un guide complet pour l'analyse de données textuelles. Il s'appuie sur un ensemble de bibliothèques géré par le langage R, logiciel libre de traitement des données et d'analyse statistiques.

Didactique, *Analyse de données textuelles sous R* présente les étapes préalables d'opérations de base comme le chargement des données, le découpage en tokens ou la conversion en matrice terme-document. Il détaille également les tâches plus complexes comme l'association des segments de discours à des locuteurs d'un entretien, l'extraction et la visualisation des thèmes, la classification des mots, l'indexation et la recherche des documents ou l'extraction des relations entre entités nommées. Cet ensemble d'opérations doit pouvoir s'intégrer dans un écosystème de plateformes d'analyse comme un système de recommandation, d'aide à la traduction ou d'analyse des médias sociaux.

L'auteur

Docteur en informatique, Nicolas Turenne est chercheur à l'INRA au sein du Laboratoire Interdisciplinaire Sciences Innovation Sociétés à l'université de Paris-Est. Il est spécialiste en extraction et gestion de connaissances à partir de textes en utilisant des modèles stochastiques et relationnels.

iSTE editions

9 781784 051075

Nicolas Turenne

Analyse de données textuelles sous R

COLLECTION SCIENCES COGNITIVES

Analyse de données textuelles sous R

Nicolas Turenne

iSTE editions

Cet ouvrage est un guide complet pour l'analyse de données textuelles. Il s'appuie sur un ensemble de bibliothèques géré par le langage R, logiciel libre de traitement des données et d'analyse statistiques.

Didactique, *Analyse de données textuelles sous R* présente les étapes préalables d'opérations de base comme le chargement des données, le découpage en tokens ou la conversion en matrice terme-document. Il détaille également les tâches plus complexes comme l'association des segments de discours à des locuteurs d'un entretien, l'extraction et la visualisation des thèmes, la classification des mots, l'indexation et la recherche des documents ou l'extraction des relations entre entités nommées. Cet ensemble d'opérations doit pouvoir s'intégrer dans un écosystème de plateformes d'analyse comme un système de recommandation, d'aide à la traduction ou d'analyse des médias sociaux.

L'auteur

Docteur en informatique, Nicolas Turenne est chercheur à l'INRA au sein du Laboratoire Interdisciplinaire Sciences Innovation Sociétés à l'université de Paris-Est. Il est spécialiste en extraction et gestion de connaissances à partir de textes en utilisant des modèles stochastiques et relationnels.

iSTE editions

9 781784 051075

Nicolas Turenne

Analyse de données textuelles sous R

COLLECTION SCIENCES COGNITIVES

Analyse de données textuelles sous R

Nicolas Turenne

iSTE editions

Nom	Thème	Maintenance	Contact
DeducerText	interface graphique qui utilise tm	Ian Fellows	<ian@fellstat.com>
koRpus	indices de lisibilité d'un corpus	m.eik michalke	<meik.michalke@hhu.de>
LanguageR	modèles de langage	R. H. Baayen	<harald.baayen@uni-tuebingen.de>
maptpx	model selection for latent topics in text	Matt Taddy	taddy@chicagobooth.edu
maxent	classification de textes	Timothy P. Jurka	<tpjurka@ucdavis.edu>
openNLP	extraction d'entités nommées	Kurt Hornik	<Kurt.Hornik@R-project.org>
qdap	analyse quantitative du discours	Tyler Rinker	<tyler.rinker at gmail.com>
RcmdrPlugin.temis	plug-in à R commander; extraction de cooccurrents; classifications de termes	Milan Bouchet-Valat	<nalimilan@club.fr>
RTextTools	apprentissage automatique pour la classification automatique	Timothy P. Jurka	<tpjurka at ucdavis.edu>
smdc	similarité entre documents	Masaaki TAKADA	<tkdmah@gmail.com>
stm	structural topic model	Brandon Stewart	bstewart@fas.harvard.edu
stringdist	similarité entre chaînes	Mark van der Loo	<mark.vanderloo@gmail.com>
textcat	catégorisation de documents par n-grammes	Kurt Hornik	<Kurt.Hornik@R-project.org>
textir	analyse de sentiment par régression logistique	Matt Taddy	taddy@chicagobooth.edu
textometry	scores de spécificité d'association	Matthieu Decorde	<matthieu.decorde at ens-lyon.fr>
tm	matrice termes documents; cooccurrence de mots	Ingo Feinerer	ingo.feinerer@tuwien.ac.at
tm.plugin.factiva	parsing du format factiva	Milan Bouchet-Valat	<nalimilan@club.fr>
topicmodels	Latent Dirichlet Allocation (LDA) models and Correlated Topics Models (CTM)	Bettina Grün	<Bettina.Gruen at jku.at>
lsa	Latent Semantic Analysis	Fridolin Wild	f.wild@open.ac.uk
twitteR	téléchargement de tweets	Jeff Gentry	<geoffjentry@gmail.com>
wordnet	ontologie lexicale	Kurt Hornik	<Kurt.Hornik@R-project.org>
x-ent	extraction d'entités nommées généraliste	Nicolas Turenne & Tien Phan	<nicolas.inra@yahoo.fr> <phantien84@gmail.com>
zipfR	modèle de distribution de grands nombre d'événements rares; interpolation	Stefan Evert & Marco Baroni	<stefan.evert@uos.de> <marco.baroni@unitn.it>

Nom	Thème	Maintenance	Contact
DeducerText	interface graphique qui utilise tm	Ian Fellows	<ian@fellstat.com>
koRpus	indices de lisibilité d'un corpus	m.eik michalke	<meik.michalke@hhu.de>
LanguageR	modèles de langage	R. H. Baayen	<harald.baayen@uni-tuebingen.de>
maptpx	model selection for latent topics in text	Matt Taddy	taddy@chicagobooth.edu
maxent	classification de textes	Timothy P. Jurka	<tpjurka@ucdavis.edu>
openNLP	extraction d'entités nommées	Kurt Hornik	<Kurt.Hornik@R-project.org>
qdap	analyse quantitative du discours	Tyler Rinker	<tyler.rinker at gmail.com>
RcmdrPlugin.temis	plug-in à R commander; extraction de cooccurrents; classifications de termes	Milan Bouchet-Valat	<nalimilan@club.fr>
RTextTools	apprentissage automatique pour la classification automatique	Timothy P. Jurka	<tpjurka at ucdavis.edu>
smdc	similarité entre documents	Masaaki TAKADA	<tkdmah@gmail.com>
stm	structural topic model	Brandon Stewart	bstewart@fas.harvard.edu
stringdist	similarité entre chaînes	Mark van der Loo	<mark.vanderloo@gmail.com>
textcat	catégorisation de documents par n-grammes	Kurt Hornik	<Kurt.Hornik@R-project.org>
textir	analyse de sentiment par régression logistique	Matt Taddy	taddy@chicagobooth.edu
textometry	scores de spécificité d'association	Matthieu Decorde	<matthieu.decorde at ens-lyon.fr>
tm	matrice termes documents; cooccurrence de mots	Ingo Feinerer	ingo.feinerer@tuwien.ac.at
tm.plugin.factiva	parsing du format factiva	Milan Bouchet-Valat	<nalimilan@club.fr>
topicmodels	Latent Dirichlet Allocation (LDA) models and Correlated Topics Models (CTM)	Bettina Grün	<Bettina.Gruen at jku.at>
lsa	Latent Semantic Analysis	Fridolin Wild	f.wild@open.ac.uk
twitteR	téléchargement de tweets	Jeff Gentry	<geoffjentry@gmail.com>
wordnet	ontologie lexicale	Kurt Hornik	<Kurt.Hornik@R-project.org>
x-ent	extraction d'entités nommées généraliste	Nicolas Turenne & Tien Phan	<nicolas.inra@yahoo.fr> <phantien84@gmail.com>
zipfR	modèle de distribution de grands nombre d'événements rares; interpolation	Stefan Evert & Marco Baroni	<stefan.evert@uos.de> <marco.baroni@unitn.it>

https://www.facebook.com/groups/rTextData/members/

Text Data Analytics with R | Nicolas | Accueil | Retrouver des amis

Nicolas Turenne
Modifier le profil
Vos publications

FAVORIS
Fil d'actualité
Messages
Évènements
Groupes de vente

GROUPES
Text Data Analytic...
Créer un groupe

APPLICATIONS
Vidéo en direct
Jeux
Ce jour-là
Suggérer des modi...
Pokes
Photos
Enregistré
Retrouver des amis

PAGES
Actualité des Pages
Pages suggérées
Créer une Page

ÉVÈNEMENTS
Créer un évènement

Analyse de données textuelles sous R

Text Data Analytics with R
Groupe public

Membre | Partager | Notifications | ...

Discussion | Membres | Évènements | Photos | Fichiers | Rechercher dans ce groupe

Membres 93 Administrateurs (1) 1 bloqué(s) | Par défaut | + Ajouter | Trouver un membre

Nicolas Turenne
A créé le groupe le 4 janvier 2016

Nico Bayá
Sciences Po
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Damien Gb
Paris
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Jorge N Rosy
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Norbert Tsopze
Lycée classique de dschang
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Anje Müller Gjesdal
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

https://www.facebook.com/groups/rTextData/members/

Text Data Analytics with R | Nicolas | Accueil | Retrouver des amis

Nicolas Turenne
Modifier le profil
Vos publications

FAVORIS
Fil d'actualité
Messages
Évènements
Groupes de vente

GROUPES
Text Data Analytic...
Créer un groupe

APPLICATIONS
Vidéo en direct
Jeux
Ce jour-là
Suggérer des modi...
Pokes
Photos
Enregistré
Retrouver des amis

PAGES
Actualité des Pages
Pages suggérées
Créer une Page

ÉVÈNEMENTS
Créer un évènement

Analyse de données textuelles sous R

Text Data Analytics with R
Groupe public

Membre | Partager | Notifications | ...

Discussion | Membres | Évènements | Photos | Fichiers | Rechercher dans ce groupe

Membres 93 Administrateurs (1) 1 bloqué(s) | Par défaut | + Ajouter | Trouver un membre

Nicolas Turenne
A créé le groupe le 4 janvier 2016

Nico Bayá
Sciences Po
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Damien Gb
Paris
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Jorge N Rosy
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Norbert Tsopze
Lycée classique de dschang
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Anje Müller Gjesdal
Membre
Ajouté(e) par Nicolas Turenne il y a environ 5 mois

Text Data Analytics with R

Nicolas Accueil Re

Comparison of String Distance Algorithms

To choose an good algorithm for fuzzy string matching and string distances can be tough. In this tutorial I describe and compare various fuzzy string matching algorithms using the R package stringdist.

JOYOFDATA.DE

J'aime · Commenter · Partager

4 · Vu par 40 personnes

1 partage

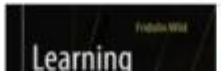
Votre commentaire...

Appuyez sur Entrée pour publier.

Fridolin Wild
14 avril · Oxford, Angleterre, Royaume-Uni

New book is out:
'Learning Analytics in R with SNA, LSA, and MPIA'

This book introduces Meaningful Purposive Interaction Analysis (MPIA) theory, which combines social network analysis (SNA) with latent semantic analysis (LSA) to help create and analyse a meaningful learning landscape from the digital traces left by a learning community in the co-construction of knowledge. ... Afficher la suite



Learning Analytics in R with SNA, LSA, and MPIA | Fridolin Wild |

Jessica Tanon
10 mars · Le Kremlin-Bicêtre, Ile-de-France

Hi everyone ! Anyone knows a good method to classify a word as leetspeak or not ? Or a scoring method that could help me do that ?

J'aime · Commenter · Partager

1 · Vu par 60 personnes

Afficher 7 autres commentaires

Jessica Tanon I'm just looking for a mathematical method that would be able to say helloworld isn't LP, hello1world123 isn't either, but h3770wor7d is, and I'm really not sure how to approach this...
J'aime · Répondre · 10 mars, 16:44

Kevin Bretonnel Cohen Character ngrams.
J'aime · Répondre · 2 · 11 mars, 13:51

Votre commentaire...

Appuyez sur Entrée pour publier.

Prateek Pawan
7 mars

Anyone know how to extract text data from PDF files and then do the text mining on it?

J'aime · Commenter · Partager

1 · Vu par 63 personnes

Afficher 4 autres commentaires

Youya Tralala http://stackoverflow.com/.../use-r-to-convert-pdf-files...

Use R to convert PDF files to text files for

Text Data Analytics with R

Nicolas Accueil Re

Comparison of String Distance Algorithms

To choose an good algorithm for fuzzy string matching and string distances can be tough. In this tutorial I describe and compare various fuzzy string matching algorithms using the R package stringdist.

JOYOFDATA.DE

J'aime · Commenter · Partager

4 · Vu par 40 personnes

1 partage

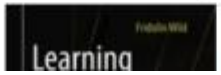
Votre commentaire...

Appuyez sur Entrée pour publier.

Fridolin Wild
14 avril · Oxford, Angleterre, Royaume-Uni

New book is out:
'Learning Analytics in R with SNA, LSA, and MPIA'

This book introduces Meaningful Purposive Interaction Analysis (MPIA) theory, which combines social network analysis (SNA) with latent semantic analysis (LSA) to help create and analyse a meaningful learning landscape from the digital traces left by a learning community in the co-construction of knowledge. ... Afficher la suite



Learning Analytics in R with SNA, LSA, and MPIA | Fridolin Wild |

Jessica Tanon
10 mars · Le Kremlin-Bicêtre, Ile-de-France

Hi everyone ! Anyone knows a good method to classify a word as leetspeak or not ? Or a scoring method that could help me do that ?

J'aime · Commenter · Partager

1 · Vu par 60 personnes

Afficher 7 autres commentaires

Jessica Tanon I'm just looking for a mathematical method that would be able to say helloworld isn't LP, hello1world123 isn't either, but h3770wor7d is, and I'm really not sure how to approach this...
J'aime · Répondre · 10 mars, 16:44

Kevin Bretonnel Cohen Character ngrams.
J'aime · Répondre · 2 · 11 mars, 13:51

Votre commentaire...

Appuyez sur Entrée pour publier.

Prateek Pawan
7 mars

Anyone know how to extract text data from PDF files and then do the text mining on it?

J'aime · Commenter · Partager

1 · Vu par 63 personnes

Afficher 4 autres commentaires

Youya Tralala http://stackoverflow.com/.../use-r-to-convert-pdf-files...

Use R to convert PDF files to text files for

Futures animations

- Webinaires ?
- Hackathon ?
- Atelier de formation ?

Futures animations

- Webinaires ?
- Hackathon ?
- Atelier de formation ?

Lightning talks