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

Nathalie Villa-Vialaneix, Christophe Bontemps & Sébastien Déjean
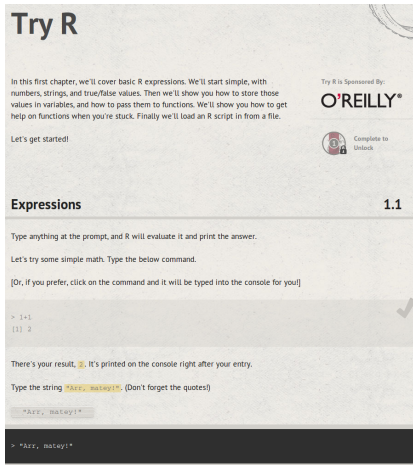
1. où commencer avec R ?

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...

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...

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 t
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 provi
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 extensio
  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 via Knaus provides a more recent alternative to snow.
- The snowfall package by Knaus provides a more recent alternative to snow. Functions can
- 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 synchronous (sequential) and asynchronous (parallel) eval
  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

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

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)
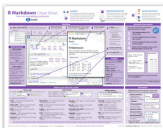
RStudio IDE Cheat Sheet

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.

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.

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.

7 feuilles sur des outils de RStudio

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

edgeR

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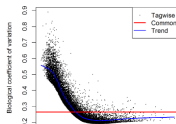
Citation (from within R, enter `citation("edgeR")`):

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] 0.0706
> plotBCV(y)
```

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

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 quantitation data.

Browse packages by topic

All Packages

Bioconductor version 3.3 (Release)

Autocomplete biocViews search:

Package	Maintainer	Title
ABRmNormalization	Jean-Philippe Fortin	Adaptive Robust Regression normalization for Illumina methylation data
BEAT	Kemal Akman	BEAT - BS-Seq Epigenetic 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

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 news and tutorials contributed by (573) R bloggers

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

WELCOME!

R Passes SAS in Scholarly Use (finally)

June 7, 2016

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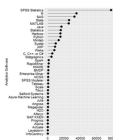
Way back in 2012 I published a forecast that showed that the use of R for scholarly publications would likely pass the use of SAS in 2015. But I didn't believe the forecast since I expected the sharp decline in SAS ... Continue reading →

Read more »

Economist Liquidity Index

June 7, 2016

By Sam Weiss
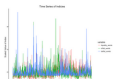
Analyze Face Emotions with R

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

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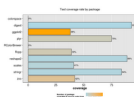
Test coverage of the 10 most downloaded R packages 2014-04-30 Source Introduction How do you know that your code is well tested ? The test coverage is the proportion of source code lines that are executed (covered) when running the tests. It is useful to find the parts of your code that are no exercised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

Read more »

An example of monkey patching a package

August 1, 2013

By R2D2

À ne pas manquer ! Un grand Wiki...

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

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

contre leur concurrent spécialisé

http://rseek.org

et son concurrent le plus célèbre

Conclusion