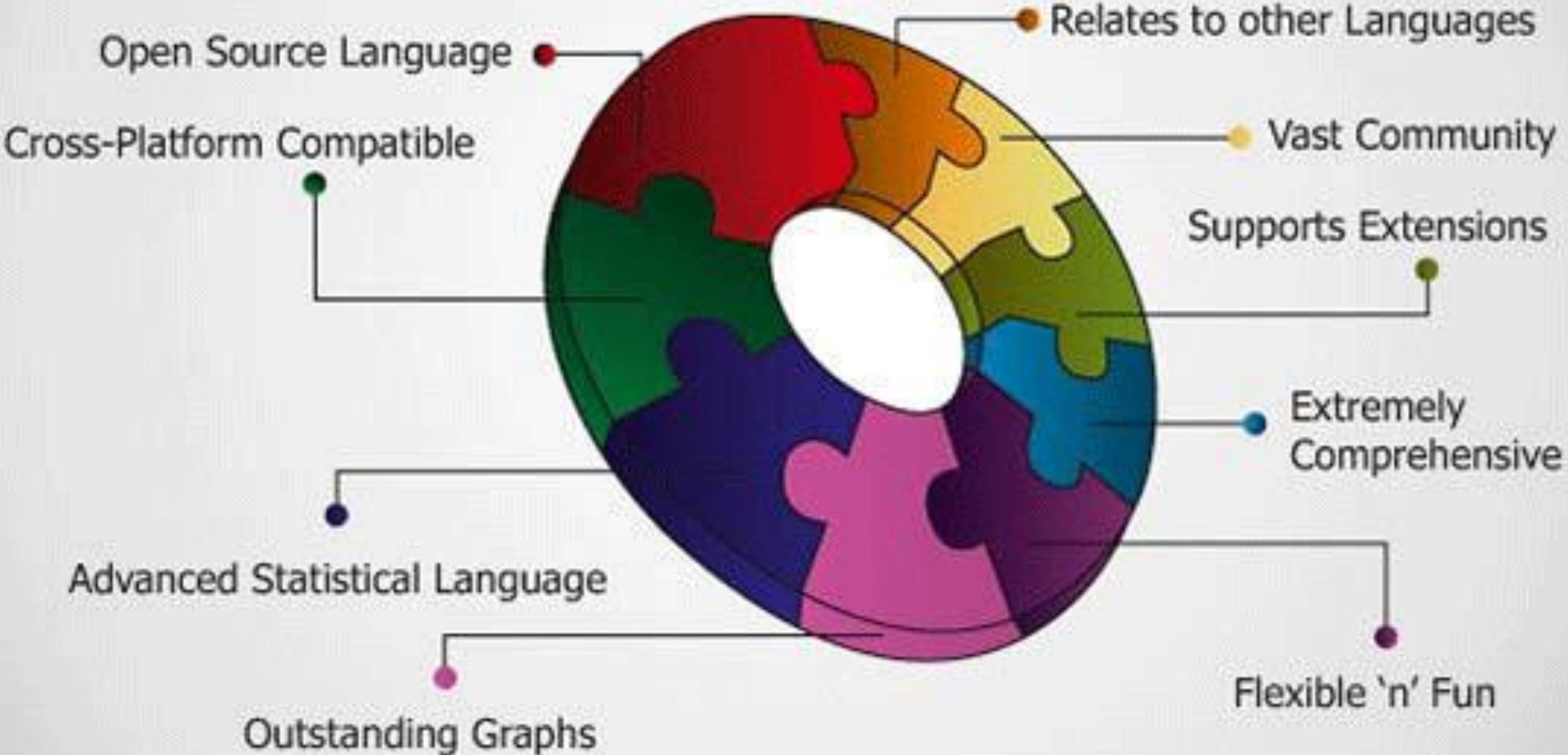
Some useful web content and information related to the programming language R

Magnus Fontes
May 2017

Why Learn R?

edureka!



Open Source Language

Cross-Platform Compatible

Advanced Statistical Language

Outstanding Graphs

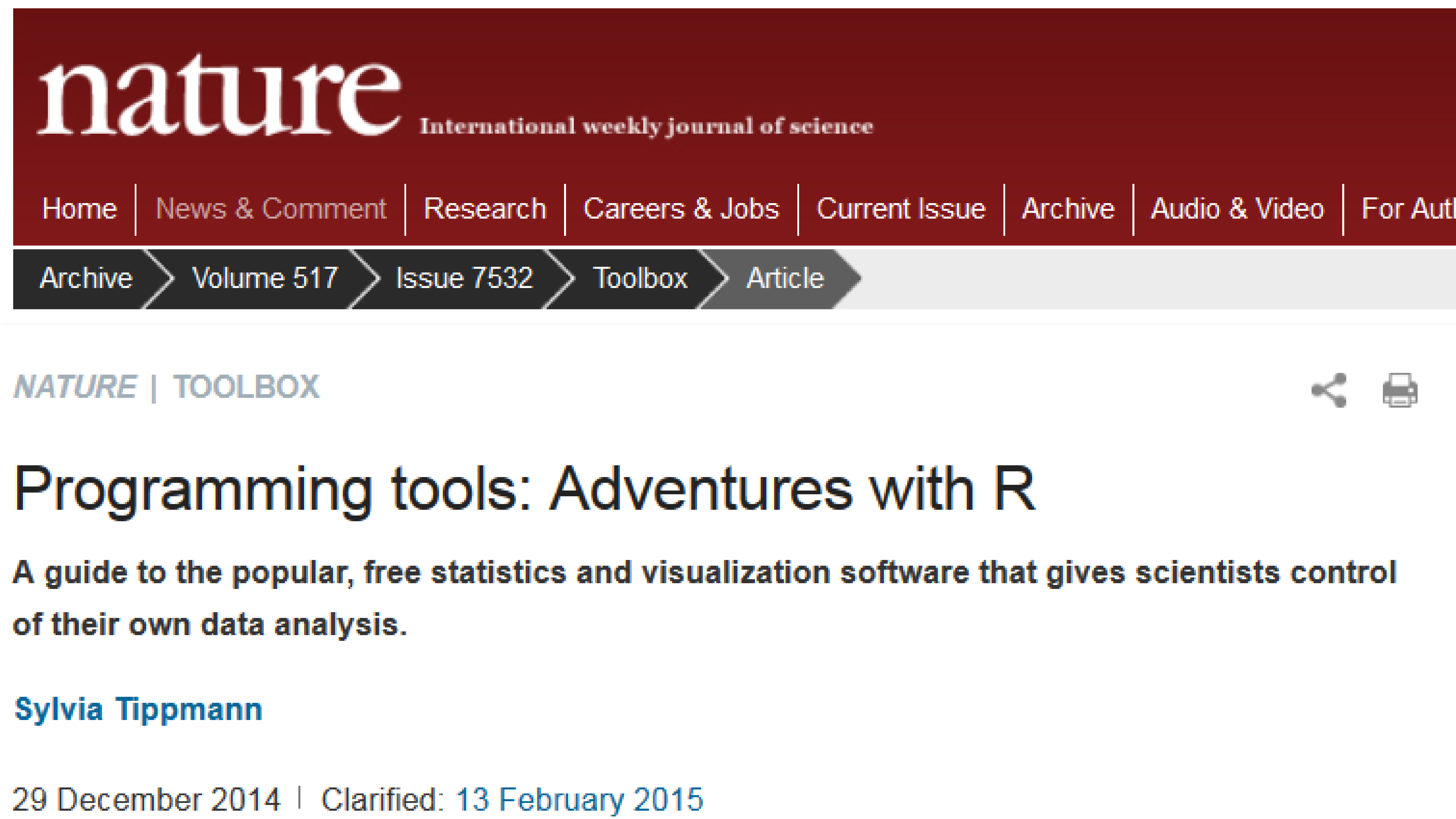
Relates to other Languages

Vast Community

Supports Extensions

Extremely Comprehensive

Flexible 'n' Fun



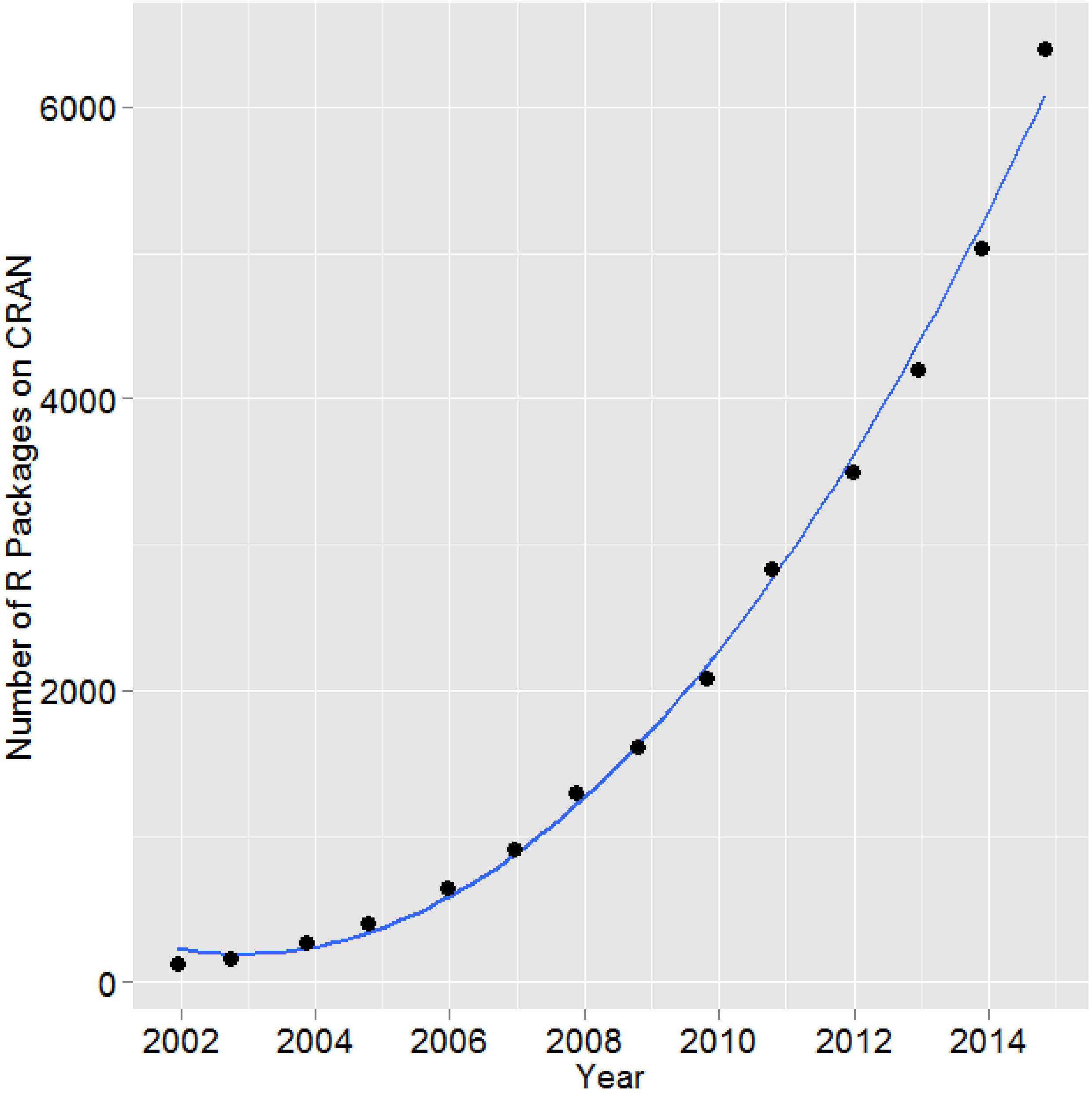
NATURE | TOOLBOX

Programming tools: Adventures with R

A guide to the popular, free statistics and visualization software that gives scientists control of their own data analysis.

Sylvia Tippmann

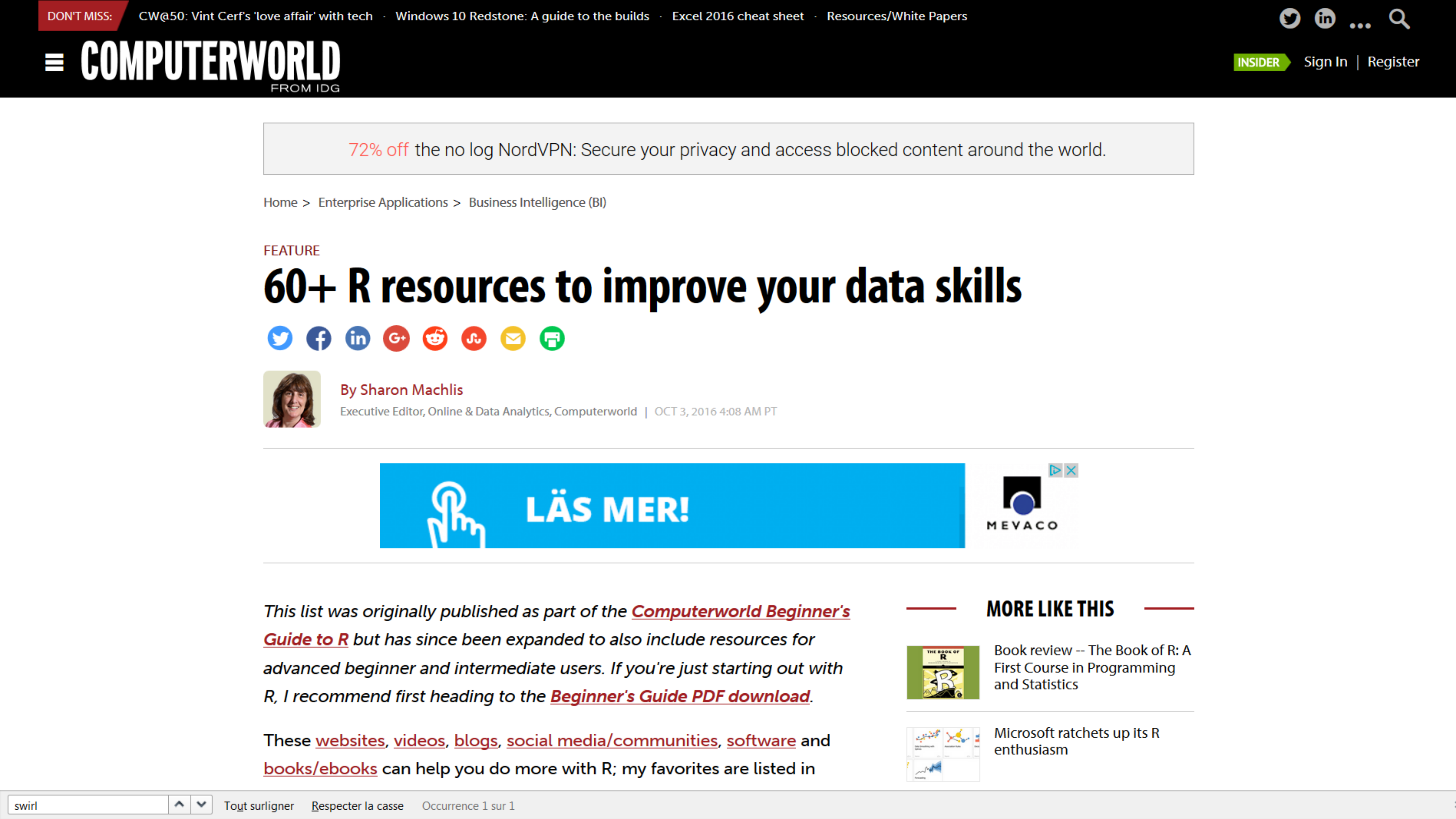
29 December 2014 | Clarified: 13 February 2015



- Versatile
- Packages-Community
- Reproducibility
- Open and free
- Development

“During 2015 alone, R added more functions/procs than SAS Institute has written in its entire history.”

http://r4stats.com/articles/popularity/



FEATURE

60+ R resources to improve your data skills

By Sharon Machlis

Executive Editor, Online & Data Analytics, Computerworld | OCT 3, 2016 4:08 AM PT

This list was originally published as part of the Computerworld Beginner's Guide to R but has since been expanded to also include resources for advanced beginner and intermediate users. If you're just starting out with R, I recommend first heading to the Beginner's Guide PDF download.

These websites, videos, blogs, social media/communities, software and books/ebooks can help you do more with R; my favorites are listed in

MORE LIKE THIS

Book review -- The Book of R: A First Course in Programming and Statistics

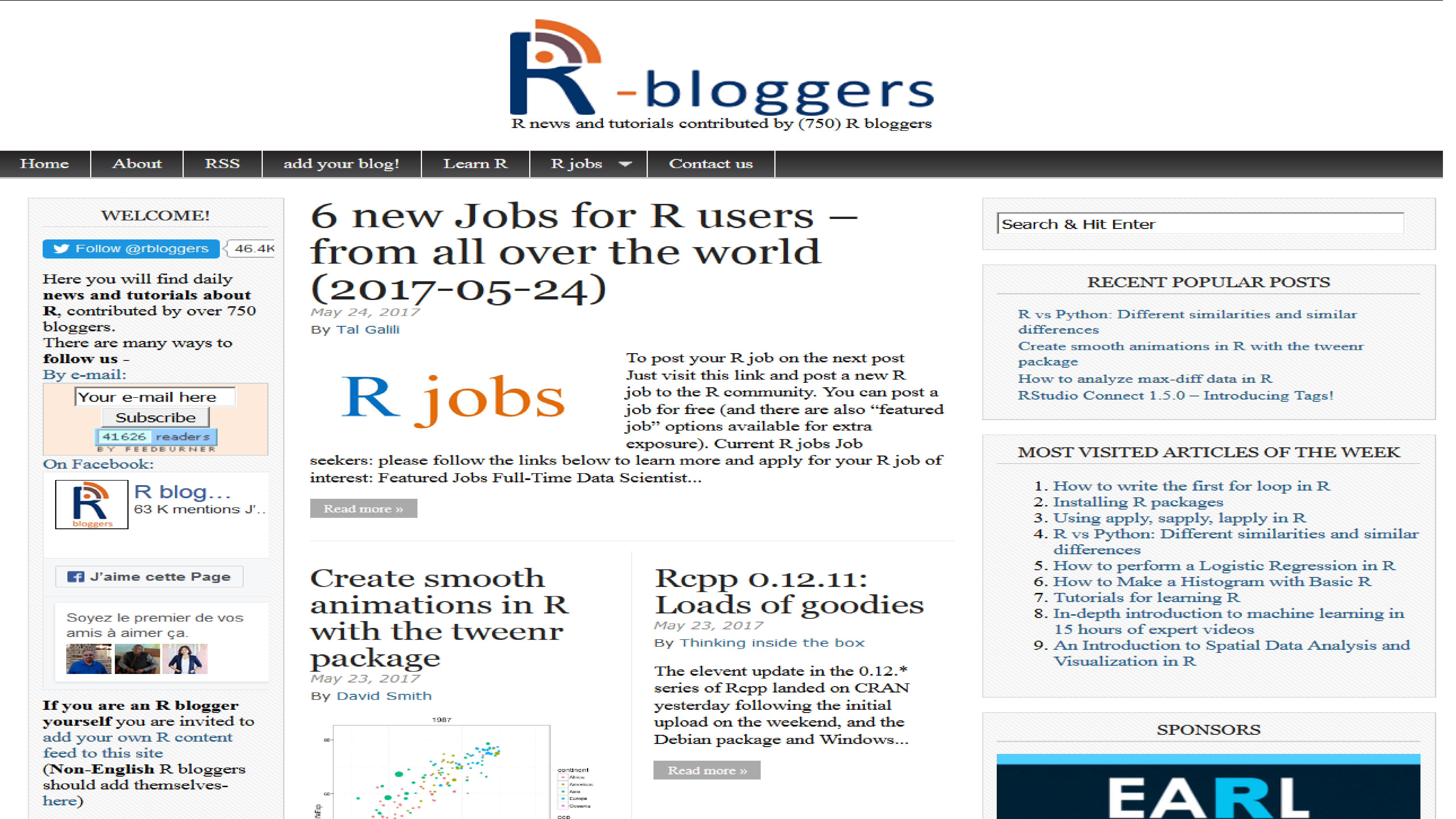
Microsoft ratchets up its R enthusiasm

{swirl}

Learn R, in R.

swirl teaches you R programming and data science interactively, at your own pace, and right in the R console!

Got questions? Join our discussion group!



R news and tutorials contributed by (750) R bloggers

WELCOME!

Here you will find daily **news and tutorials about R**, contributed by over 750 bloggers.
There are many ways to **follow us** -
By e-mail:

On Facebook:

Soyez le premier de vos amis à aimer ça.

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)

6 new Jobs for R users – from all over the world (2017-05-24)

May 24, 2017

By Tal Galili

To post your R job on the next post Just visit this link and post a new R job to the R community. You can post a job for free (and there are also "featured job" options available for extra exposure). Current R jobs Job seekers: please follow the links below to learn more and apply for your R job of interest: Featured Jobs Full-Time Data Scientist…

Create smooth animations in R with the tweenr package

May 23, 2017

By David Smith

Rcpp 0.12.11: Loads of goodies

May 23, 2017

By Thinking inside the box

The elevent update in the 0.12.* series of Rcpp landed on CRAN yesterday following the initial upload on the weekend, and the Debian package and Windows…

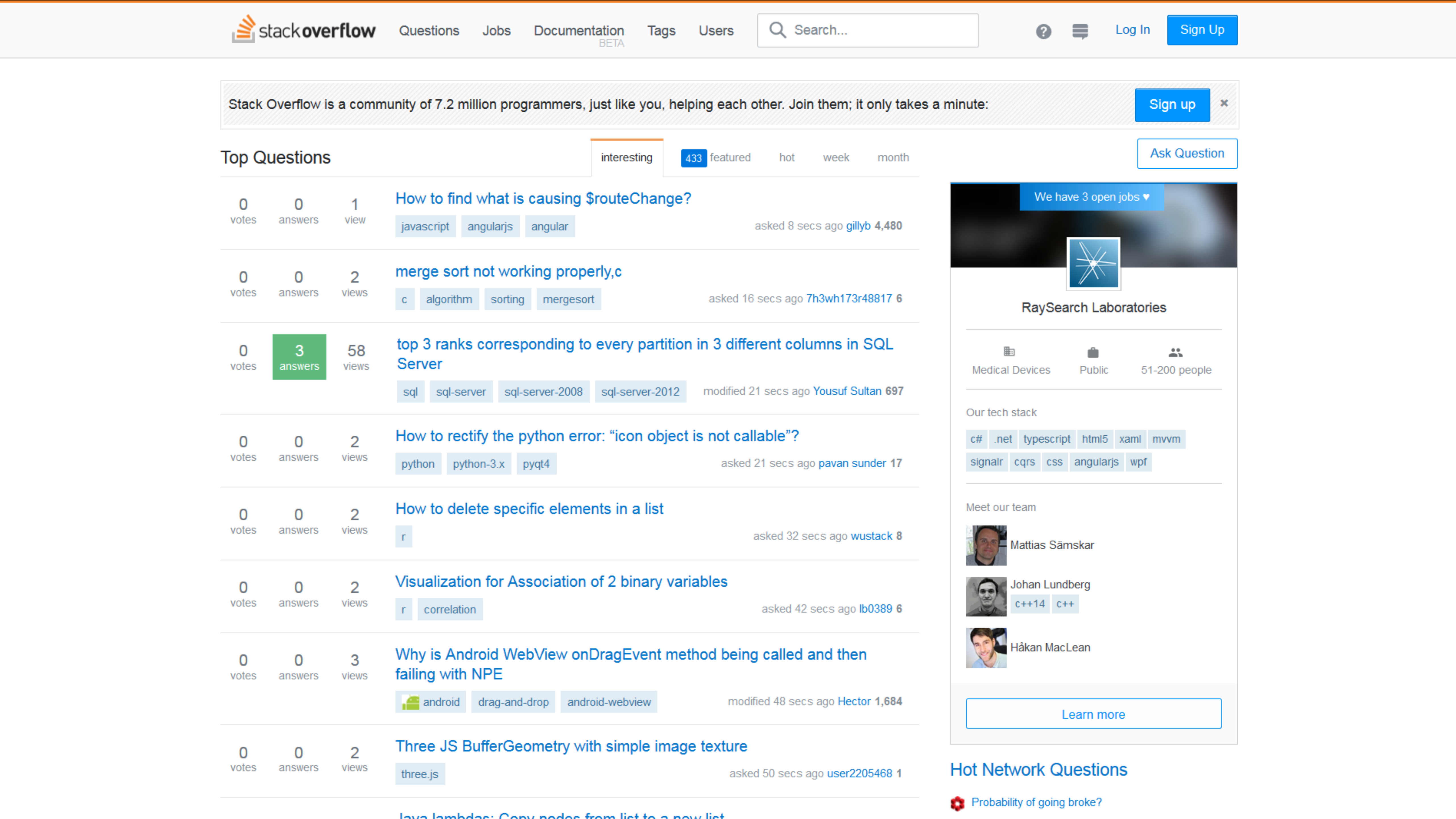
RECENT POPULAR POSTS

- R vs Python: Different similarities and similar differences
- Create smooth animations in R with the tweenr package
- How to analyze max-diff data in R
- RStudio Connect 1.5.0 – Introducing Tags!

MOST VISITED ARTICLES OF THE WEEK

1. How to write the first for loop in R
2. Installing R packages
3. Using apply, sapply, lapply in R
4. R vs Python: Different similarities and similar differences
5. How to perform a Logistic Regression in R
6. How to Make a Histogram with Basic R
7. Tutorials for learning R
8. In-depth introduction to machine learning in 15 hours of expert videos
9. An Introduction to Spatial Data Analysis and Visualization in R

SPONSORS



Stack Overflow is a community of 7.2 million programmers, just like you, helping each other. Join them; it only takes a minute:

Top Questions

0 votes 0 answers 1 view
How to find what is causing $routeChange?
javascript angularjs angular
asked 8 secs ago gillyb 4,480

0 votes 0 answers 2 views
merge sort not working properly,c
c algorithm sorting mergesort
asked 16 secs ago 7h3wh173r48817 6

0 votes 3 answers 58 views
top 3 ranks corresponding to every partition in 3 different columns in SQL Server
sql sql-server sql-server-2008 sql-server-2012
modified 21 secs ago Yousuf Sultan 697

0 votes 0 answers 2 views
How to rectify the python error: “icon object is not callable”?
python python-3.x pyqt4
asked 21 secs ago pavan sunder 17

0 votes 0 answers 2 views
How to delete specific elements in a list
r
asked 32 secs ago wustack 8

0 votes 0 answers 2 views
Visualization for Association of 2 binary variables
r correlation
asked 42 secs ago lb0389 6

0 votes 0 answers 3 views
Why is Android WebView onDragEvent method being called and then failing with NPE
android drag-and-drop android-webview
modified 48 secs ago Hector 1,684

0 votes 0 answers 2 views
Three JS BufferGeometry with simple image texture
three.js
asked 50 secs ago user2205468 1

Java lambdas: Copy nodes from list to a new list

RaySearch Laboratories

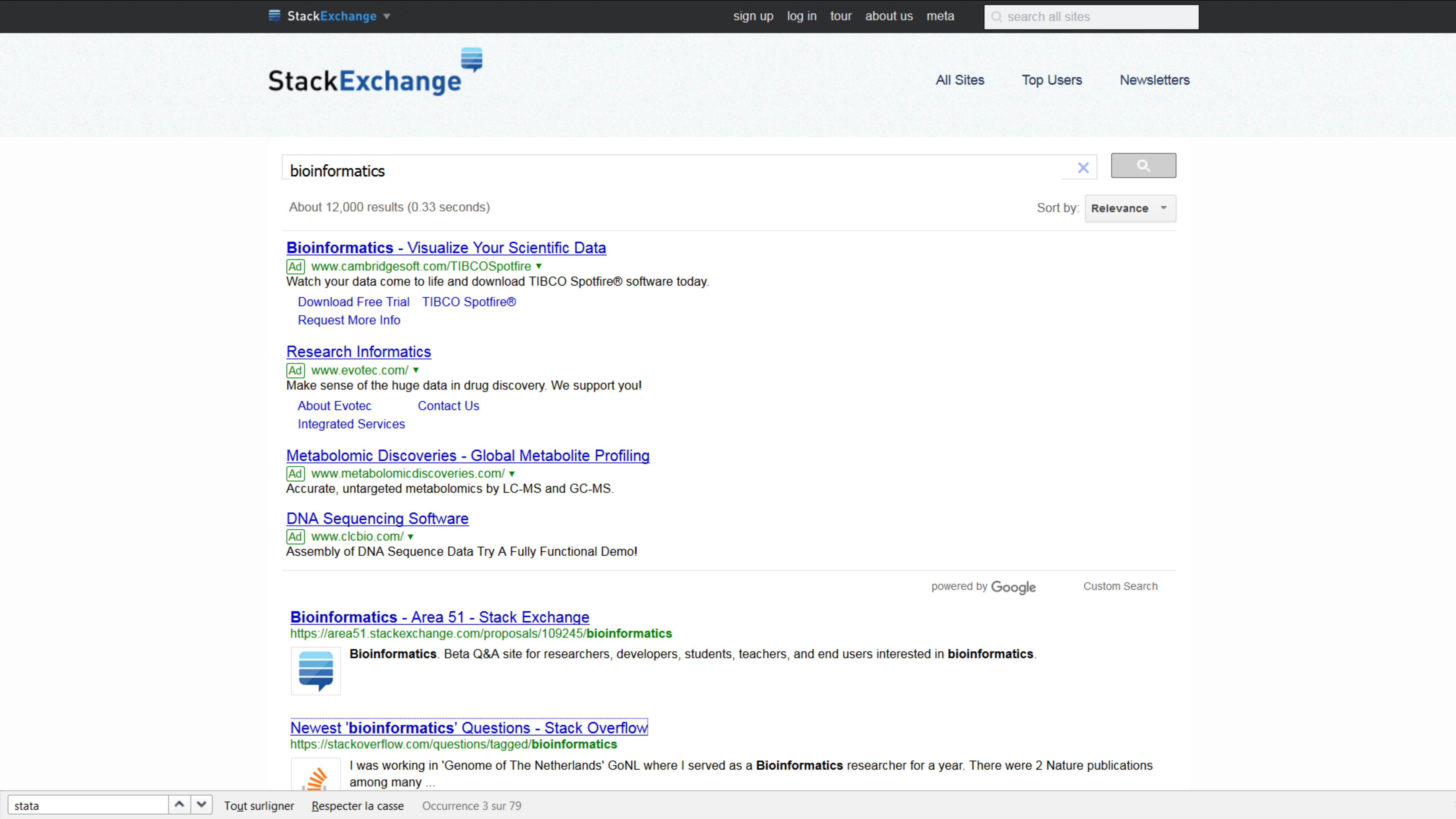
Medical Devices
Public
51-200 people

Our tech stack
c# .net typescript html5 xaml mvvm signalr cqrs css angularjs wpf

Meet our team
Mattias Sämskar
Johan Lundberg c++14 c++
Håkan MacLean

Hot Network Questions

Probability of going broke?



About 12,000 results (0.33 seconds)

Bioinformatics - Visualize Your Scientific Data
Ad www.cambridgesoft.com/TIBCOSpotfire ▾
Watch your data come to life and download TIBCO Spotfire® software today.
Download Free Trial TIBCO Spotfire®
Request More Info

Research Informatics
Ad www.evotec.com/ ▾
Make sense of the huge data in drug discovery. We support you!
About Evotec Contact Us
Integrated Services

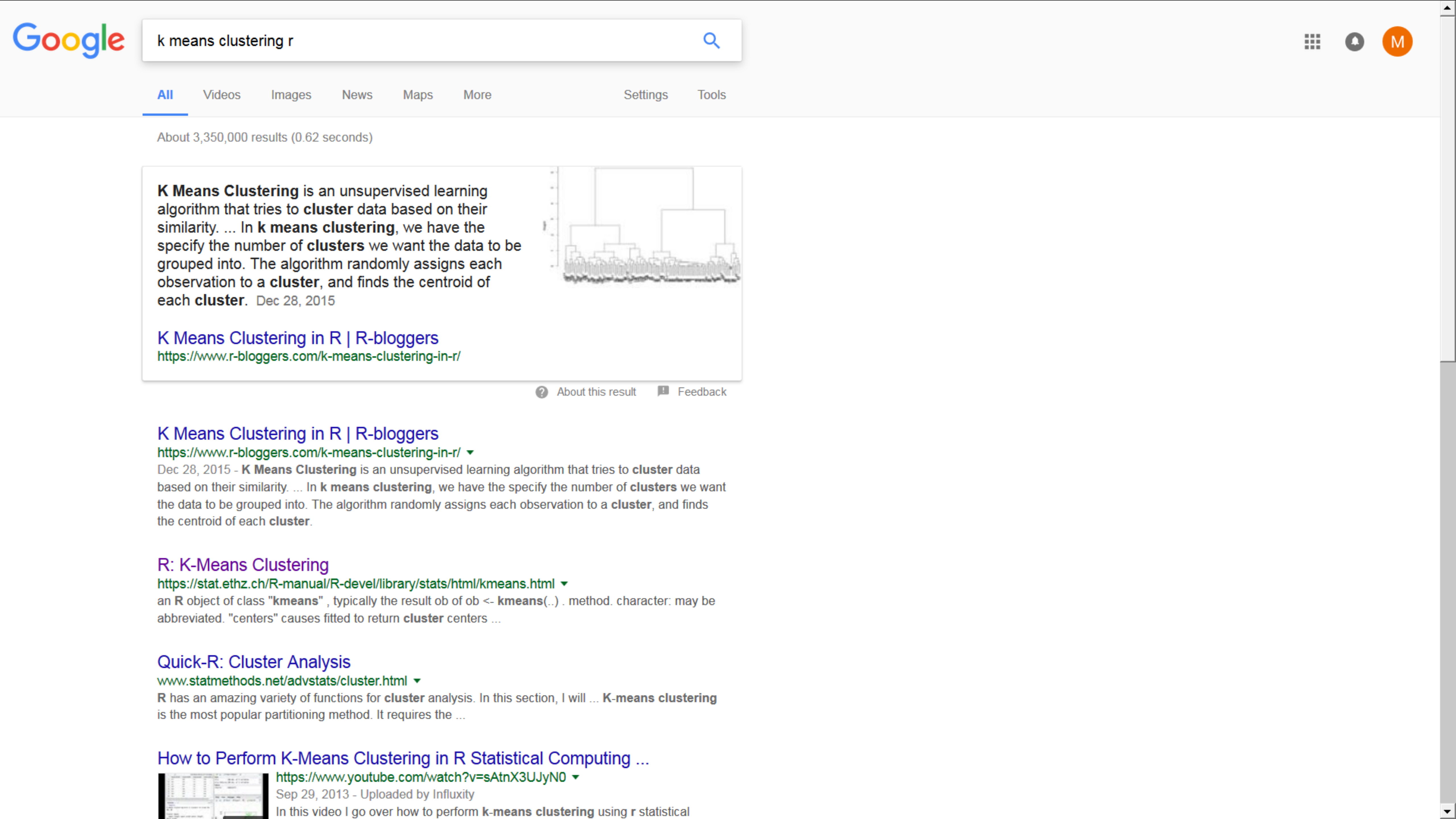
Metabolomic Discoveries - Global Metabolite Profiling
Ad www.metabolomicdiscoveries.com/ ▾
Accurate, untargeted metabolomics by LC-MS and GC-MS.

DNA Sequencing Software
Ad www.clcbio.com/ ▾
Assembly of DNA Sequence Data Try A Fully Functional Demo!

powered by Google Custom Search

Bioinformatics - Area 51 - Stack Exchange
https://area51.stackexchange.com/proposals/109245/**bioinformatics**
Bioinformatics. Beta Q&A site for researchers, developers, students, teachers, and end users interested in **bioinformatics**.

Newest '**bioinformatics**' Questions - Stack Overflow
https://stackoverflow.com/questions/tagged/**bioinformatics**
I was working in 'Genome of The Netherlands' GoNL where I served as a **Bioinformatics** researcher for a year. There were 2 Nature publications among many ...



About 3,350,000 results (0.62 seconds)

K Means Clustering is an unsupervised learning algorithm that tries to **cluster** data based on their similarity. ... In **k means clustering**, we have the specify the number of **clusters** we want the data to be grouped into. The algorithm randomly assigns each observation to a **cluster**, and finds the centroid of each **cluster**. Dec 28, 2015

K Means Clustering in R | R-bloggers
https://www.r-bloggers.com/k-means-clustering-in-r/

K Means Clustering in R | R-bloggers
https://www.r-bloggers.com/k-means-clustering-in-r/ ▾
Dec 28, 2015 - **K Means Clustering** is an unsupervised learning algorithm that tries to **cluster** data based on their similarity. ... In **k means clustering**, we have the specify the number of **clusters** we want the data to be grouped into. The algorithm randomly assigns each observation to a **cluster**, and finds the centroid of each **cluster**.

R: K-Means Clustering
https://stat.ethz.ch/R-manual/R-devel/library/stats/html/kmeans.html ▾
an **R** object of class "**kmeans**" , typically the result ob of ob <- **kmeans**(..) . method. character: may be abbreviated. "centers" causes fitted to return **cluster** centers ...

Quick-R: Cluster Analysis
www.statmethods.net/advstats/cluster.html ▾
R has an amazing variety of functions for **cluster** analysis. In this section, I will ... **K-means clustering** is the most popular partitioning method. It requires the ...

How to Perform K-Means Clustering in R Statistical Computing ...
https://www.youtube.com/watch?v=sAtnX3UJyN0 ▾
Sep 29, 2013 - Uploaded by Influxity
In this video I go over how to perform **k-means clustering** using **r** statistical

R-bloggers

R news and tutorials contributed by (750) R bloggers

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

WELCOME!

Follow @rbloggers 46.4K

Here you will find daily **news and tutorials about R**, contributed by over 750 bloggers.
There are many ways to **follow us** -
By e-mail:

Your e-mail here
Subscribe
41626 readers BY FEEDBURNER

On Facebook:

R blog…
63 K mentions J'…

J'aime cette Page

Soyez le premier de vos amis à aimer ça.

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)

K Means Clustering in R

December 28, 2015

By Teja Kodali

Like 3 Share | in Share 22

(This article was first published on **DataScience+**, and kindly contributed to R-bloggers)

594 SHARES | Share | Tweet

Hello everyone, hope you had a wonderful Christmas! In this post I will show you how to do k means clustering in R. We will use the `iris` dataset from the `datasets` library.

What is K Means Clustering?

K Means Clustering is an unsupervised learning algorithm that tries to cluster data based on their similarity. Unsupervised learning means that there is no outcome to be predicted, and the algorithm just tries to find patterns in the data. In k means clustering, we have the specify the number of clusters we want the data to be grouped into. The algorithm randomly assigns each observation to a cluster, and finds the centroid of each cluster. Then, the algorithm iterates through two steps:

- Reassign data points to the cluster whose centroid is closest.
- Calculate new centroid of each cluster.

These two steps are repeated till the within cluster variation cannot be reduced any further. The within cluster variation is calculated as the sum of the euclidean distance between the data points and their respective cluster centroids.

Exploring the data

The `iris` dataset contains data about sepal length, sepal width, petal length, and petal width of flowers of different species. Let us see what it looks like:

Search & Hit Enter

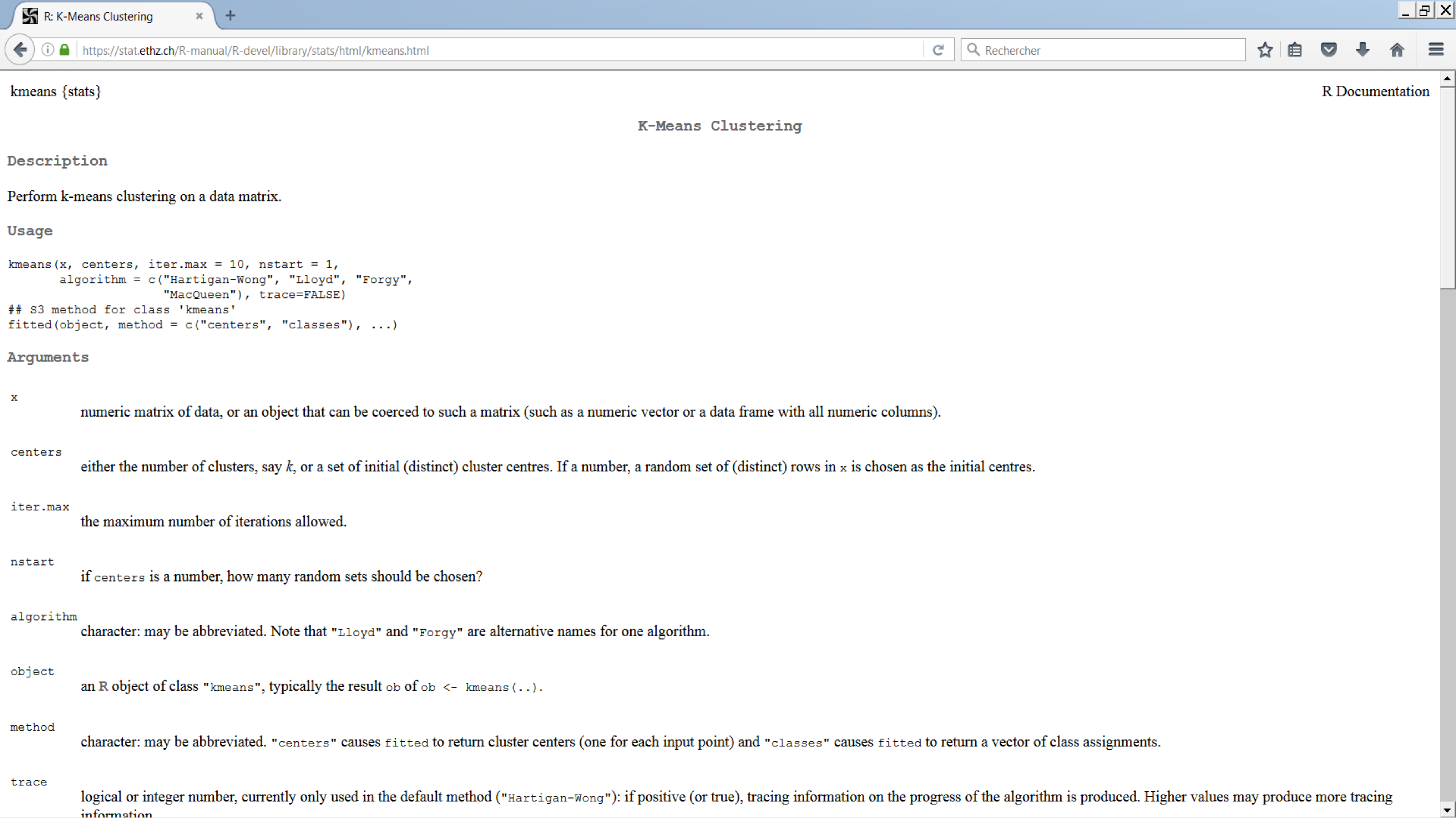
RECENT POPULAR POSTS

- New series: R and big data (concentrating on Spark and sparklyr)
- Text Mining with R: A Tidy Approach
- A Primer in Functional Programming in R Exercises (Part – 1)
- Sankey charts for swinging voters
- Review of Efficient R Programming

MOST VISITED ARTICLES OF THE WEEK

1. How to write the first for loop in R
2. Installing R packages
3. An Introduction to Spatial Data Analysis and Visualization in R
4. Using apply, sapply, lapply in R
5. How to perform a Logistic Regression in R
6. Tutorials for learning R
7. How to Make a Histogram with Basic R
8. Network analysis of Game of Thrones family ties
9. In-depth introduction to machine learning in 15 hours of expert videos

SPONSORS



kmeans {stats} R Documentation

K-Means Clustering

Description

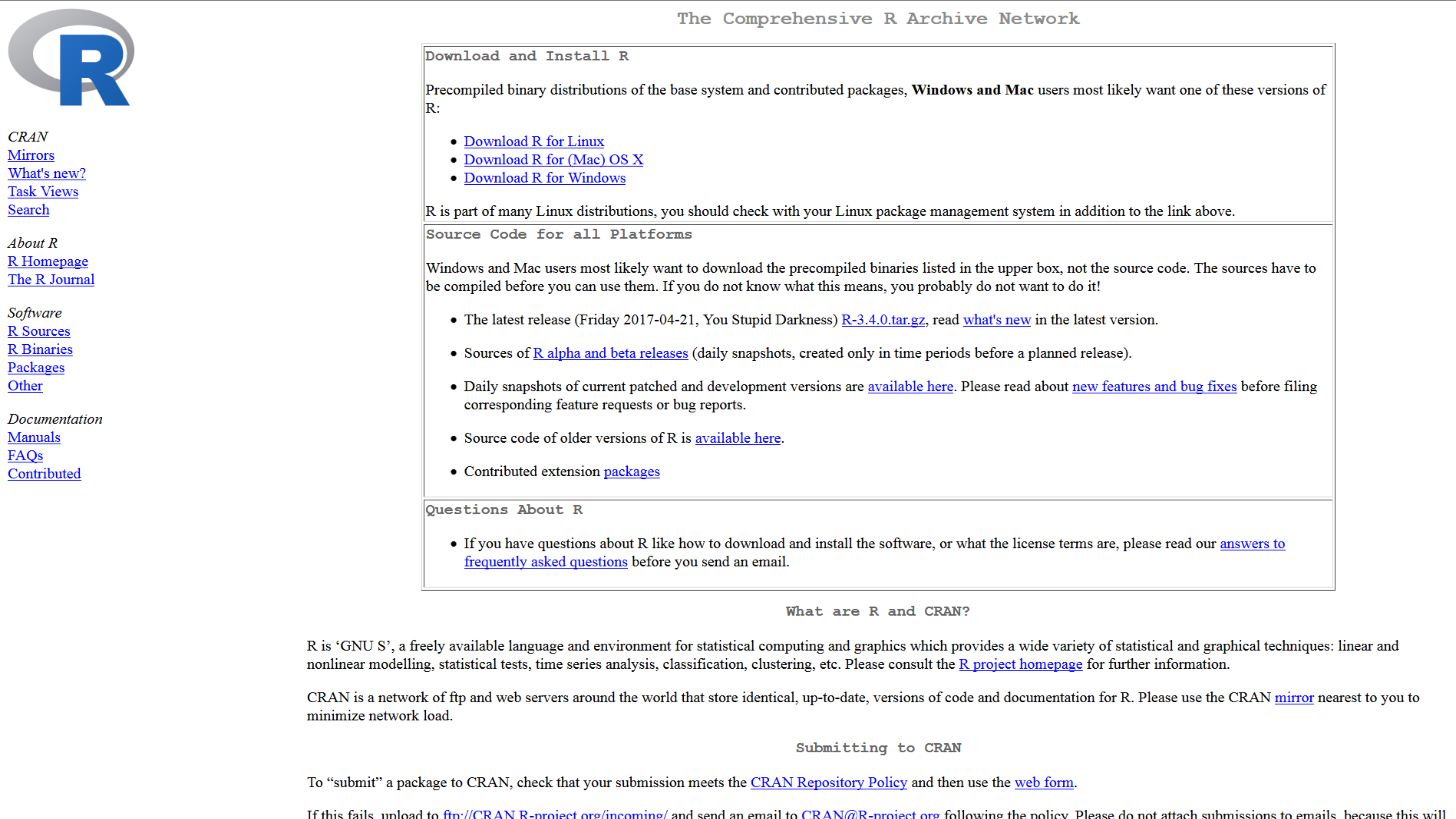
Perform k-means clustering on a data matrix.

Usage

```
kmeans(x, centers, iter.max = 10, nstart = 1,
       algorithm = c("Hartigan-Wong", "Lloyd", "Forgy",
                     "MacQueen"), trace=FALSE)
## S3 method for class 'kmeans'
fitted(object, method = c("centers", "classes"), ...)
```

Arguments

`x`	numeric matrix of data, or an object that can be coerced to such a matrix (such as a numeric vector or a data frame with all numeric columns).
`centers`	either the number of clusters, say *k*, or a set of initial (distinct) cluster centres. If a number, a random set of (distinct) rows in `x` is chosen as the initial centres.
`iter.max`	the maximum number of iterations allowed.
`nstart`	if `centers` is a number, how many random sets should be chosen?
`algorithm`	character: may be abbreviated. Note that `"Lloyd"` and `"Forgy"` are alternative names for one algorithm.
`object`	an R object of class `"kmeans"`, typically the result `ob` of `ob <- kmeans(..)`.
`method`	character: may be abbreviated. `"centers"` causes `fitted` to return cluster centers (one for each input point) and `"classes"` causes `fitted` to return a vector of class assignments.
`trace`	logical or integer number, currently only used in the default method (`"Hartigan-Wong"`): if positive (or true), tracing information on the progress of the algorithm is produced. Higher values may produce more tracing information.



The Comprehensive R Archive Network

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

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Friday 2017-04-21, You Stupid Darkness) R-3.4.0.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

What are R and CRAN?

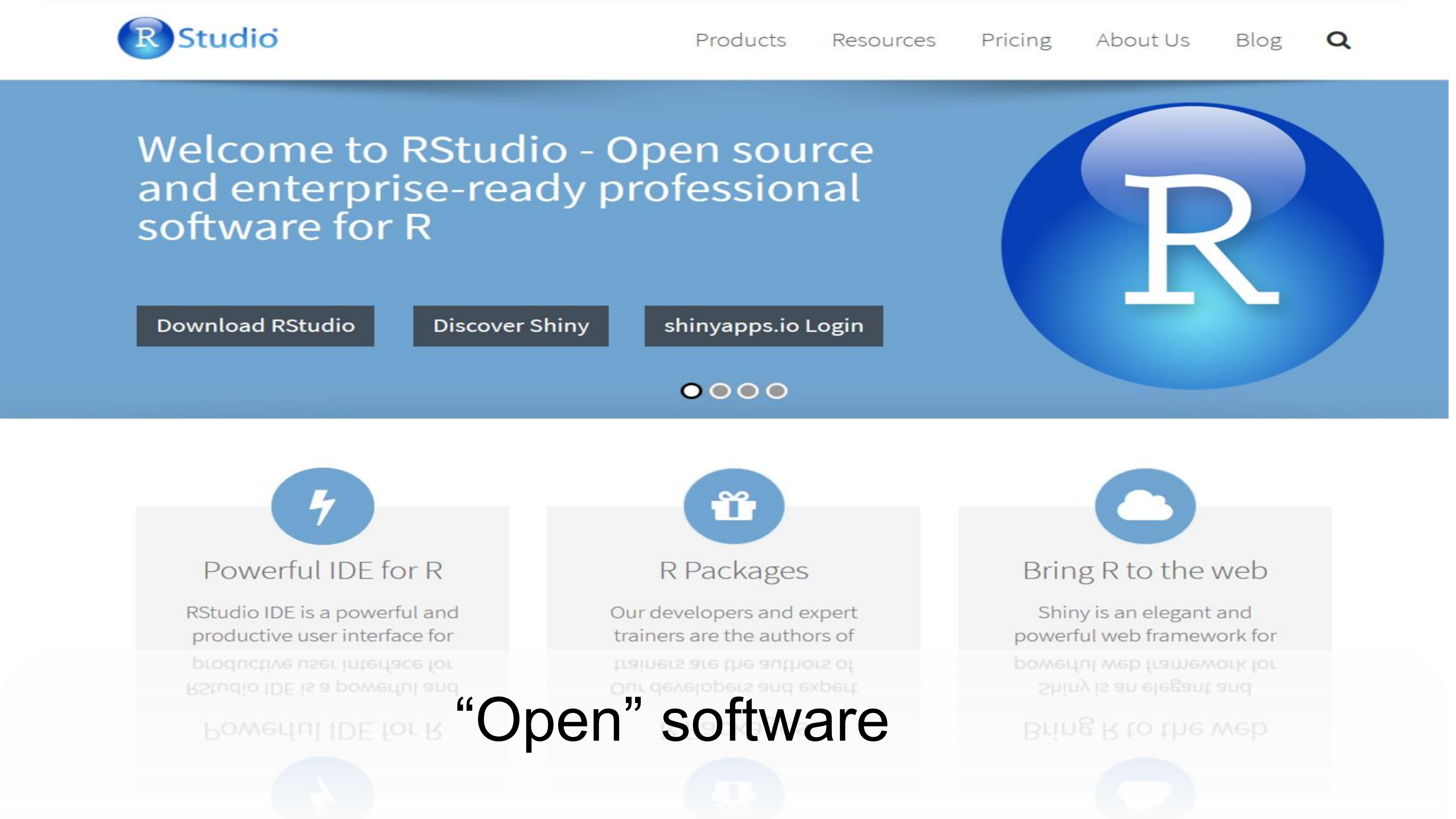
R is ‘GNU S’, a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN mirror nearest to you to minimize network load.

Submitting to CRAN

To “submit” a package to CRAN, check that your submission meets the CRAN Repository Policy and then use the web form.

If this fails, upload to ftp://CRAN.R-project.org/incoming/ and send an email to CRAN@R-project.org following the policy. Please do not attach submissions to emails, because this will



Welcome to RStudio - Open source and enterprise-ready professional software for R

Powerful IDE for R

RStudio IDE is a powerful and productive user interface for

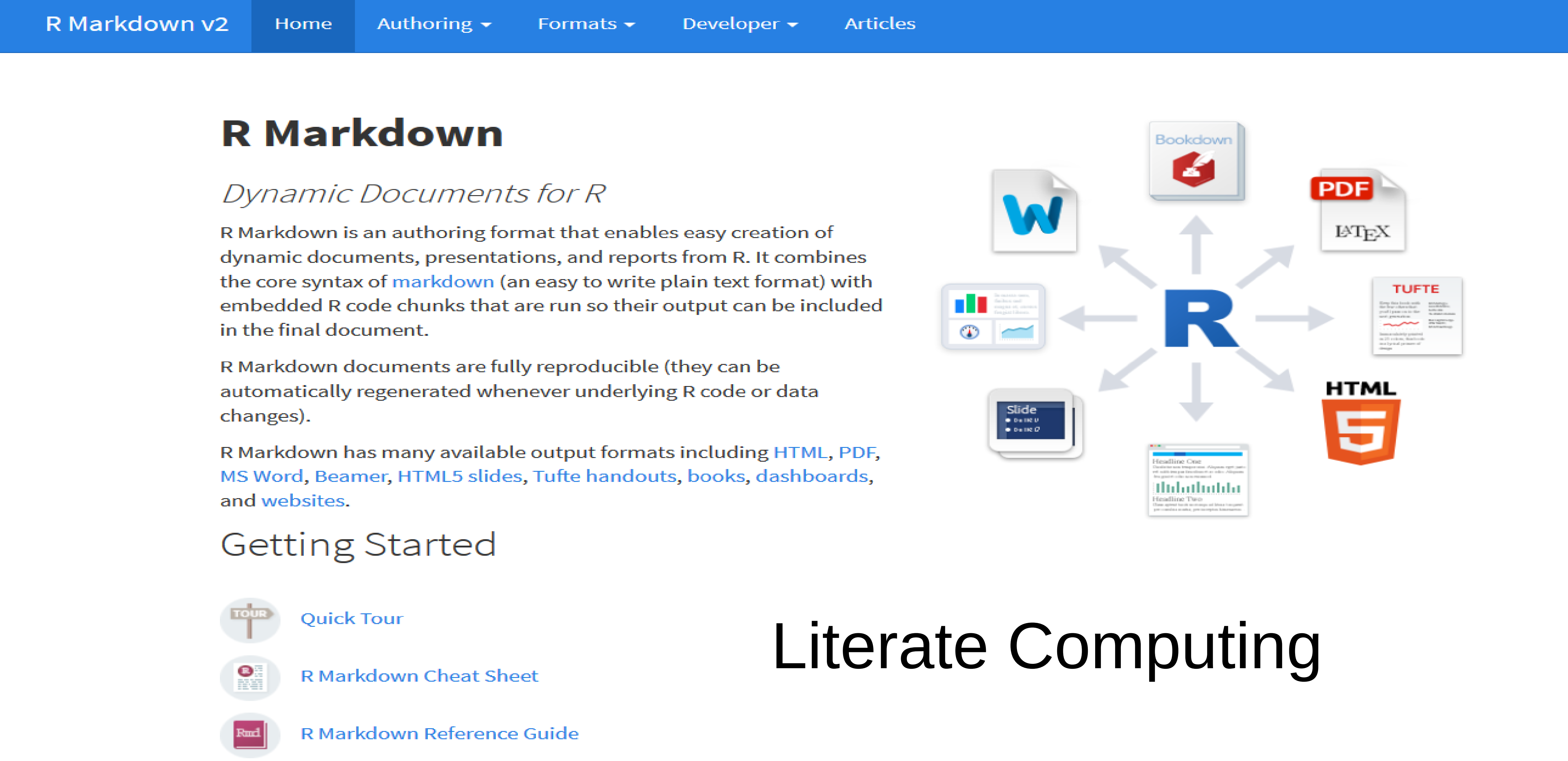
R Packages

Our developers and expert trainers are the authors of

Bring R to the web

Shiny is an elegant and powerful web framework for

“Open” software



R Markdown

Dynamic Documents for R

R Markdown is an authoring format that enables easy creation of dynamic documents, presentations, and reports from R. It combines the core syntax of markdown (an easy to write plain text format) with embedded R code chunks that are run so their output can be included in the final document.

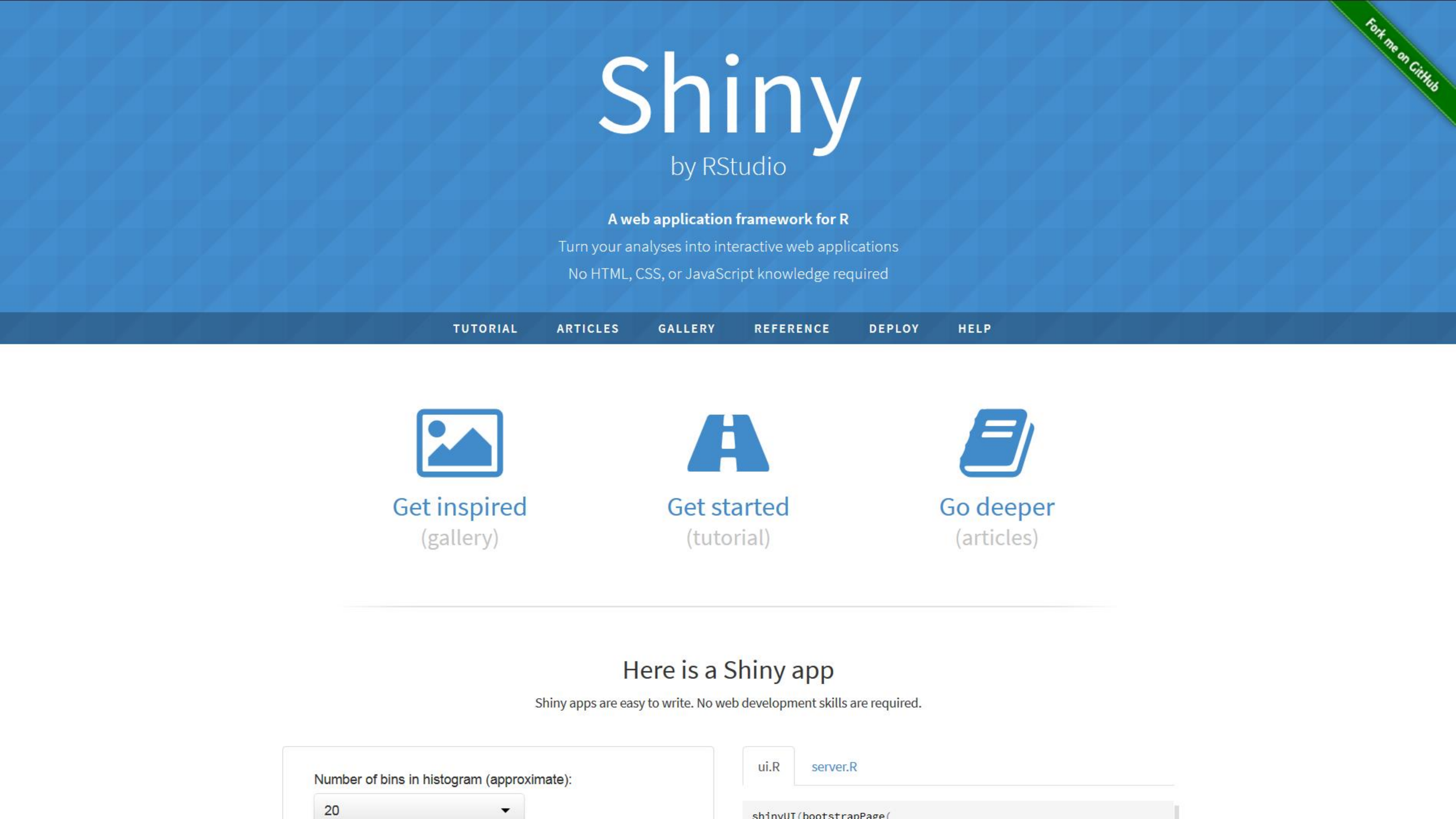
R Markdown documents are fully reproducible (they can be automatically regenerated whenever underlying R code or data changes).

R Markdown has many available output formats including HTML, PDF, MS Word, Beamer, HTML5 slides, Tufte handouts, books, dashboards, and websites.

Getting Started

- Quick Tour
- R Markdown Cheat Sheet
- R Markdown Reference Guide

Literate Computing



Shiny

by RStudio

A web application framework for R

Turn your analyses into interactive web applications

No HTML, CSS, or JavaScript knowledge required

Get inspired
(gallery)

Get started
(tutorial)

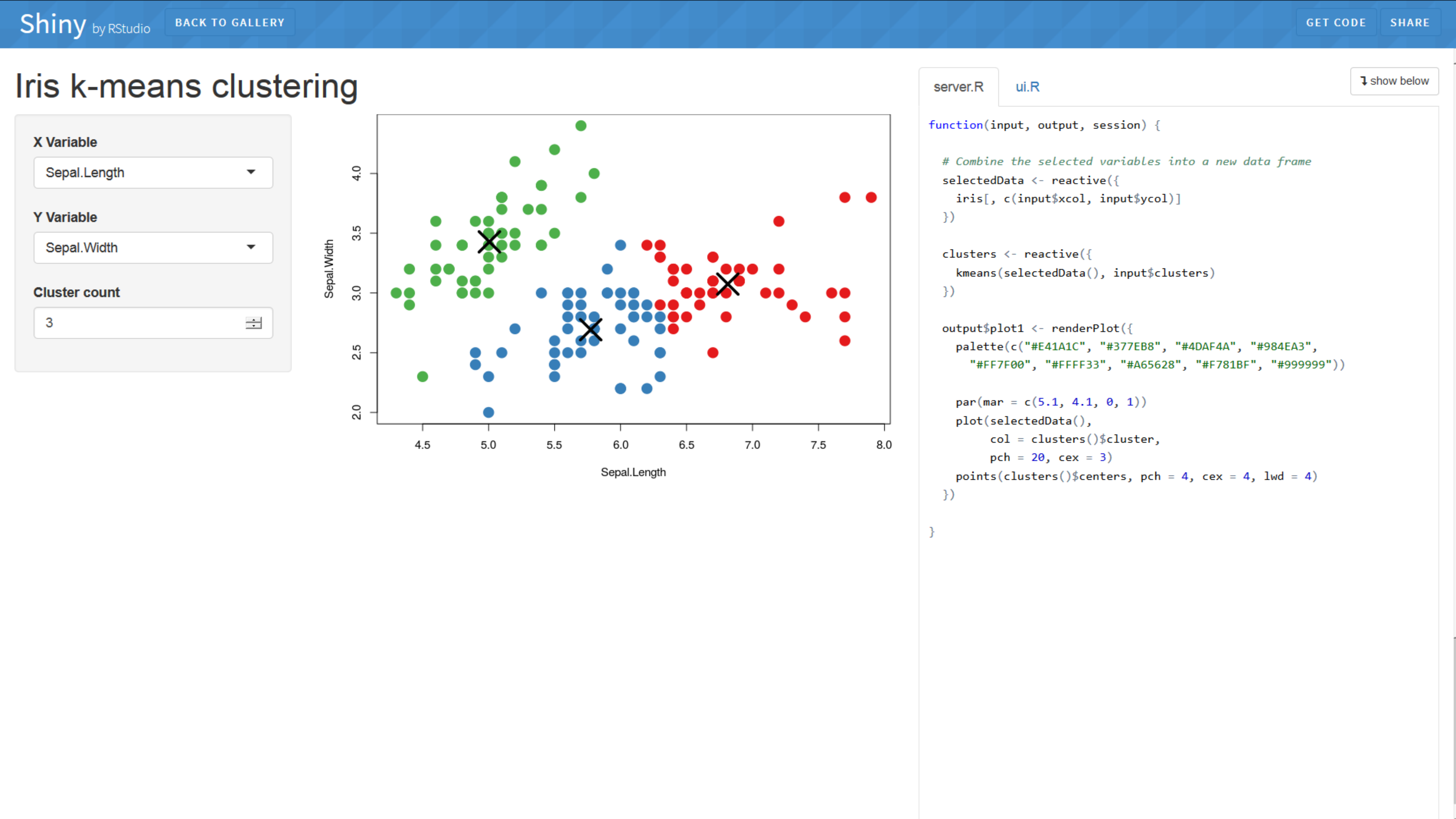
Go deeper
(articles)

Here is a Shiny app

Shiny apps are easy to write. No web development skills are required.

Number of bins in histogram (approximate): 20

```
shinyUI(bootstrapPage(
```



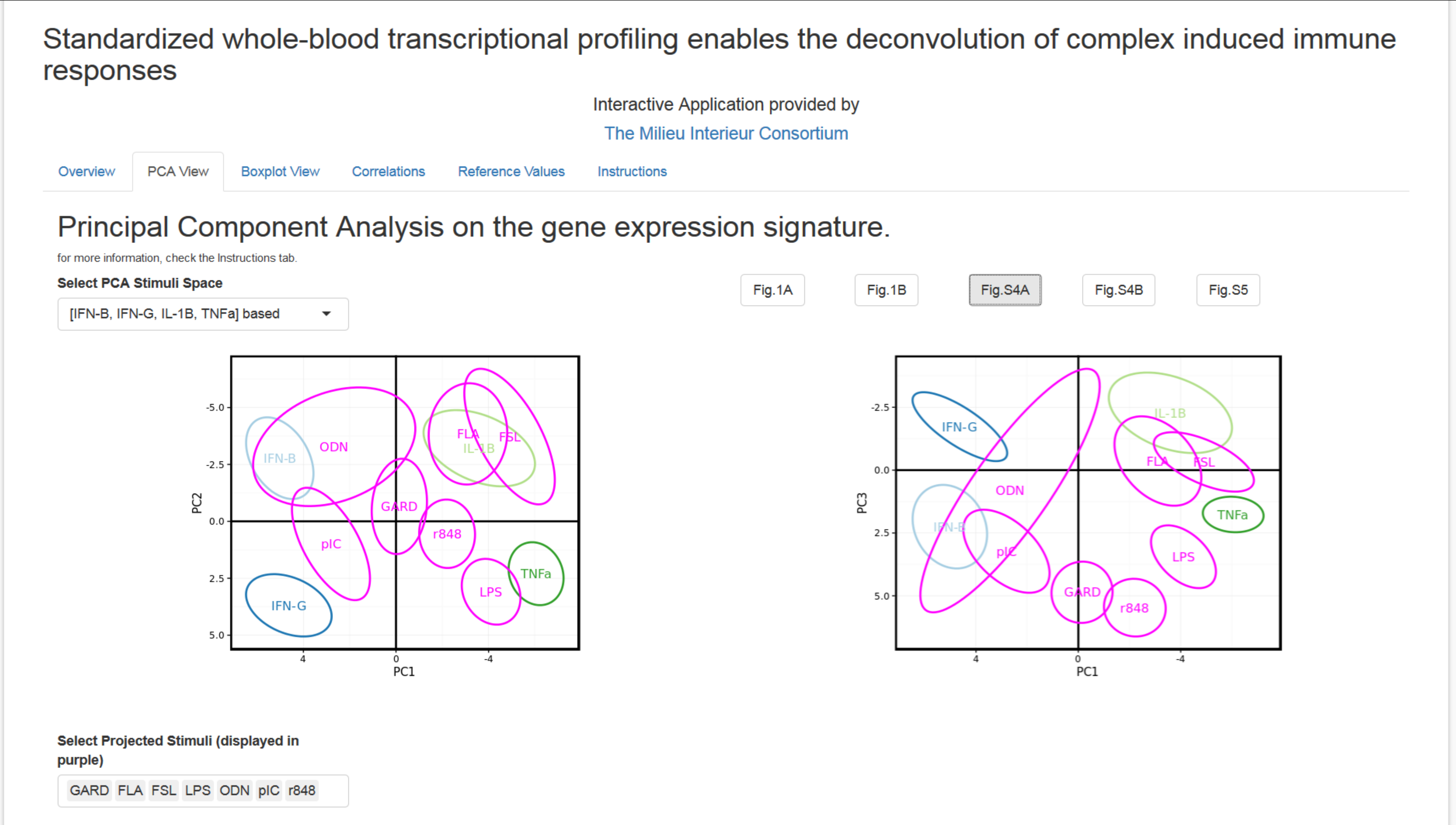
Iris k-means clustering

X Variable
Sepal.Length

Y Variable
Sepal.Width

Cluster count
3

```
function(input, output, session) {

  # Combine the selected variables into a new data frame
  selectedData <- reactive({
    iris[, c(input$xcol, input$ycol)]
  })

  clusters <- reactive({
    kmeans(selectedData(), input$clusters)
  })

  output$plot1 <- renderPlot({
    palette(c("#E41A1C", "#377EB8", "#4DAF4A", "#984EA3",
      "#FF7F00", "#FFFF33", "#A65628", "#F781BF", "#999999"))

    par(mar = c(5.1, 4.1, 0, 1))
    plot(selectedData(),
         col = clusters()$cluster,
         pch = 20, cex = 3)
    points(clusters()$centers, pch = 4, cex = 4, lwd = 4)
  })

}
```



Standardized whole-blood transcriptional profiling enables the deconvolution of complex induced immune responses

Interactive Application provided by
The Milieu Interieur Consortium

Principal Component Analysis on the gene expression signature.

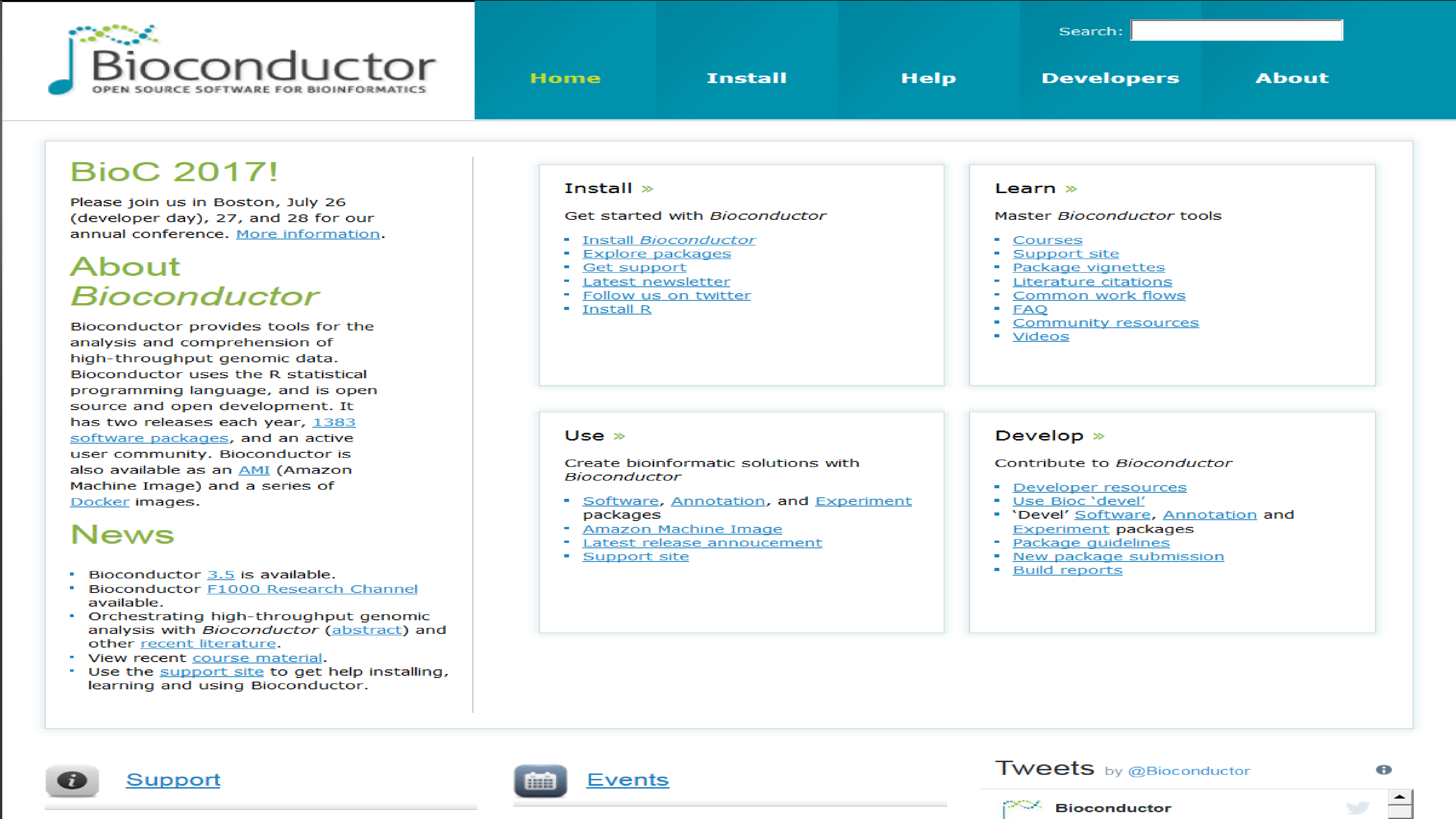
for more information, check the Instructions tab.

Select PCA Stimuli Space

[IFN-B, IFN-G, IL-1B, TNFa] based

Select Projected Stimuli (displayed in purple)

GARD FLA FSL LPS ODN pIC r848



BioC 2017!

Please join us in Boston, July 26 (developer day), 27, and 28 for our annual conference. More information.

About *Bioconductor*

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, 1383 software packages, and an active user community. Bioconductor is also available as an AMI (Amazon Machine Image) and a series of Docker images.

News

- Bioconductor 3.5 is available.
- Bioconductor F1000 Research Channel available.
- Orchestrating high-throughput genomic analysis with *Bioconductor* (abstract) and other recent literature.
- View recent course material.
- Use the support site to get help installing, learning and using Bioconductor.

Install »

Get started with *Bioconductor*

- Install *Bioconductor*
- Explore packages
- Get support
- Latest newsletter
- Follow us on twitter
- Install R

Learn »

Master *Bioconductor* tools

- Courses
- Support site
- Package vignettes
- Literature citations
- Common work flows
- FAQ
- Community resources
- Videos

Use »

Create bioinformatic solutions with *Bioconductor*

- Software, Annotation, and Experiment packages
- Amazon Machine Image
- Latest release annoucement
- Support site

Develop »

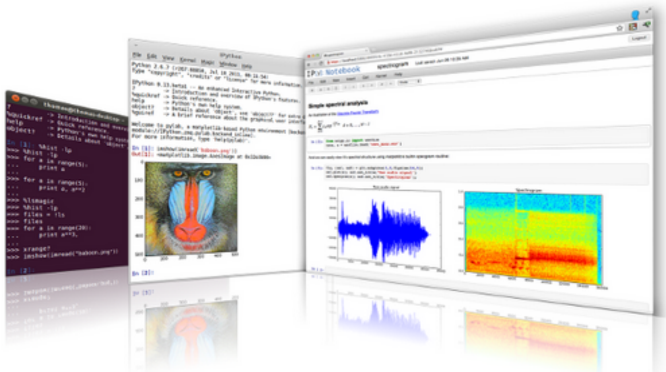
Contribute to *Bioconductor*

- Developer resources
- Use Bioc 'devel'
- 'Devel' Software, Annotation and Experiment packages
- Package guidelines
- New package submission
- Build reports

IP[y]: IPython
Interactive Computing

IPython provides a rich architecture for interactive computing with:

- A powerful interactive shell.
- A kernel for Jupyter.
- Support for interactive data visualization and use of GUI toolkits.
- Flexible, embeddable interpreters to load into your own projects.
- Easy to use, high performance tools for parallel computing.



To get started with IPython in the Jupyter Notebook, see our official example collection. Our notebook gallery is an excellent way to see the many things you can do with IPython while learning about a variety of topics, from basic programming to advanced statistics or quantum mechanics.

To learn more about IPython, you can download our talks and presentations, or read our extensive documentation. IPython is open source (BSD license), and is used by a range of other projects; add your project to that list if it uses IPython as a library, and please don't forget to cite the project.

IPython supports Python 2.7 and 3.3 or newer. Our older 1.x series supports Python 2.6 and 3.2.

Jupyter and the future of IPython

IPython is a growing project, with increasingly language-agnostic components. IPython 3.x was the last monolithic release of IPython, containing the notebook server, qtconsole, etc. As of IPython 4.0, the

Google Custom Search **Search** ×

JUPYTERCON

jupytercon

NOTEBOOK VIEWER

Share your notebooks

COMMUNITY

Stack Overflow
Mailing list
File a bug
Reddit

Tweet

FOR DEVELOPERS

Mailing list
Development Chat Room
Development information
Travis CI
Wiki

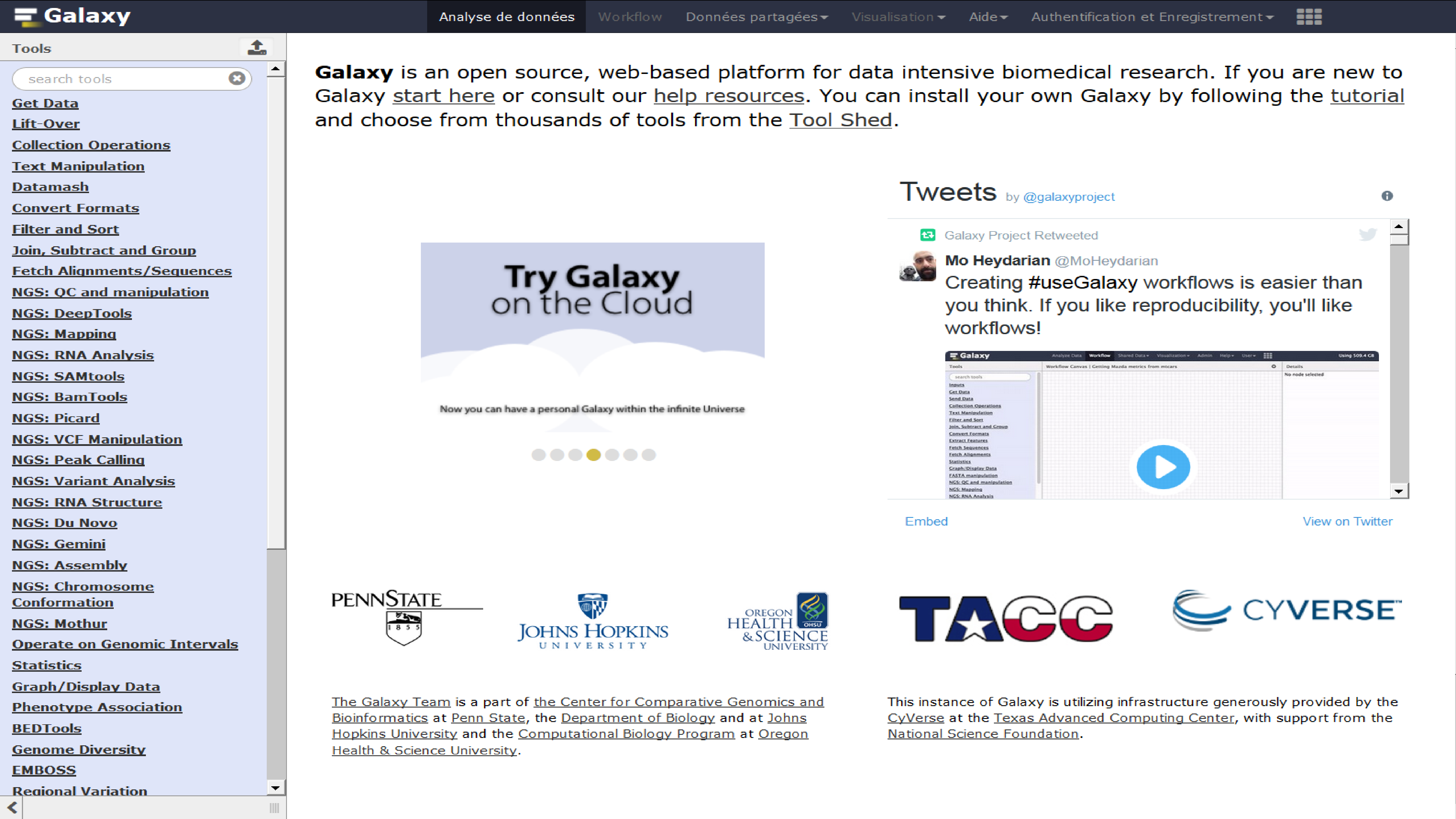
June 20th - June 24th 2017, Berkeley, CA.
CfP closes on March ~~25th~~ 31st!

Julia is part of **Google Summer of Code 2017**!
Check out our project ideas and get involved!

Julia is a high-level, high-performance dynamic programming language for numerical computing. It provides a sophisticated compiler, distributed parallel execution, numerical accuracy, and an extensive mathematical function library. Julia's Base library, largely written in Julia itself, also integrates mature, best-of-breed open source C and Fortran libraries for linear algebra, random number generation, signal processing, and string processing. In addition, the Julia developer community is contributing a number of external packages through Julia's built-in package manager at a rapid pace. IJulia, a collaboration between the Jupyter and Julia communities, provides a powerful browser-based graphical notebook interface to Julia.

Les programmes Julia sont organisés autour de la distribution multiple (multiple dispatch) ; en définissant les fonctions et en les surchargeant pour différentes combinaisons de types d'arguments, qui peuvent également être définis par l'utilisateur. Pour une discussion plus approfondie de la justification et des avantages de Julia sur d'autres systèmes, voir les points suivants ou lire l'introduction dans le manuel en ligne.

JuliaCon 2016, la conférence annuelle sur Julia, a eu lieu du 21 au 25 juin au MIT. Voici une vidéo aléatoire de notre liste de lecture YouTube de ces présentations. Cliquez sur l'icône de sélection pour voir les autres vidéos.



Get Data
Lift-Over
Collection Operations
Text Manipulation
Datamash
Convert Formats
Filter and Sort
Join, Subtract and Group
Fetch Alignments/Sequences
NGS: QC and manipulation
NGS: DeepTools
NGS: Mapping
NGS: RNA Analysis
NGS: SAMtools
NGS: BamTools
NGS: Picard
NGS: VCF Manipulation
NGS: Peak Calling
NGS: Variant Analysis
NGS: RNA Structure
NGS: Du Novo
NGS: Gemini
NGS: Assembly
NGS: Chromosome Conformation
NGS: Mothur
Operate on Genomic Intervals
Statistics
Graph/Display Data
Phenotype Association
BEDTools
Genome Diversity
EMBOSS
Regional Variation

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

Tweets by @galaxyproject

Galaxy Project Retweeted

Mo Heydarian @MoHeydarian

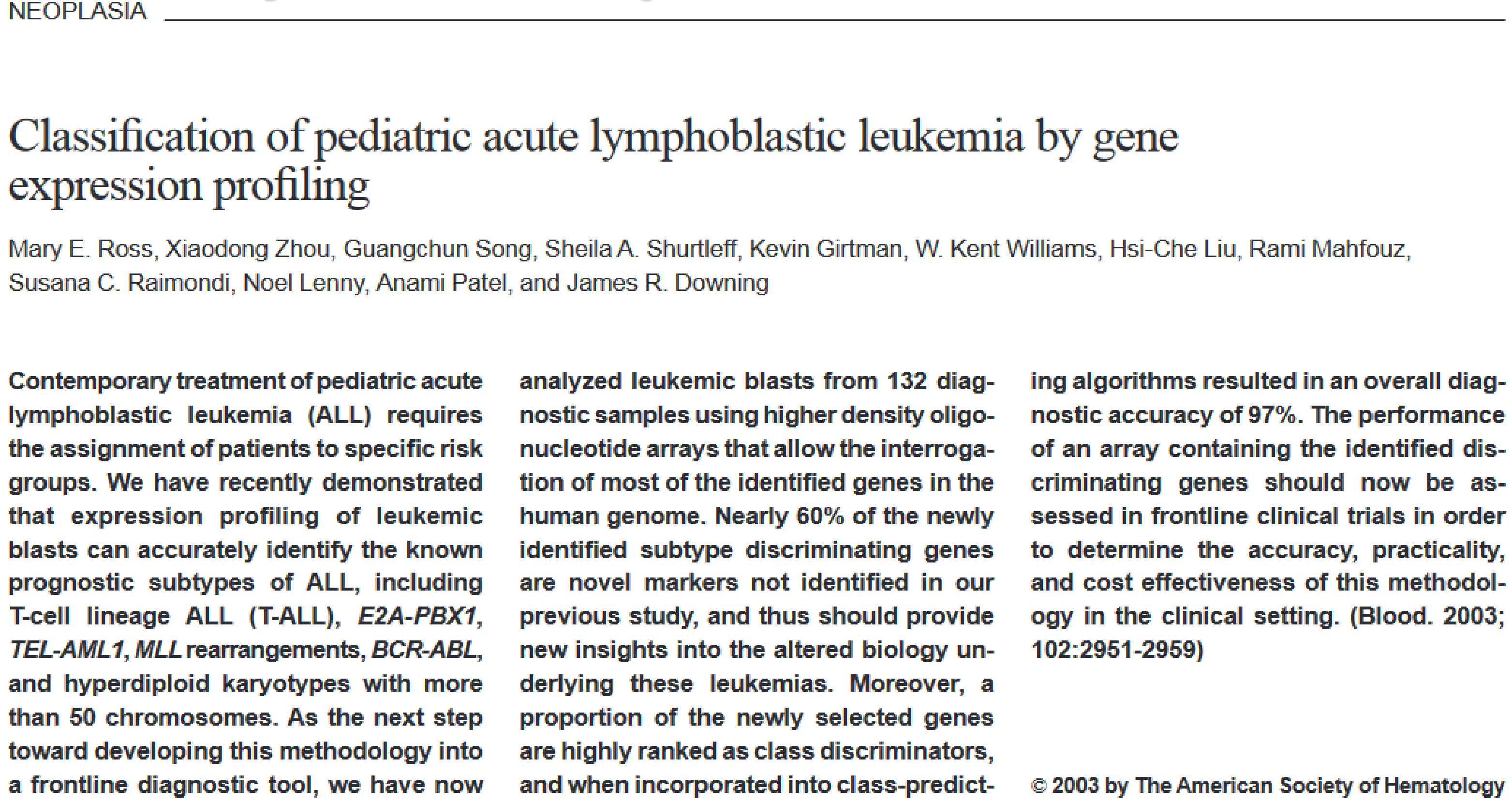
Creating #useGalaxy workflows is easier than you think. If you like reproducibility, you'll like workflows!

The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, the Department of Biology and at Johns Hopkins University and the Computational Biology Program at Oregon Health & Science University.

This instance of Galaxy is utilizing infrastructure generously provided by the CyVerse at the Texas Advanced Computing Center, with support from the National Science Foundation.

Classification of pediatric acute lymphoblastic leukemia by gene expression profiling
Mary E. Ross, Xiaodong Zhou, Guangchun Song, Sheila A. Shurtleff, Kevin Girtman, W. Kent Williams, Hsi-Che Liu, Rami Mahfouz, Susana C. Raimondi, Noel Lenny, Anami Patel, James R. Downing
Blood 2003 102:2951-2959; doi:10.1182/blood-2003-01-0338

Example Data: Affymetrix chip 22282 ProbeIDs and 132 Samples



NEOPLASIA

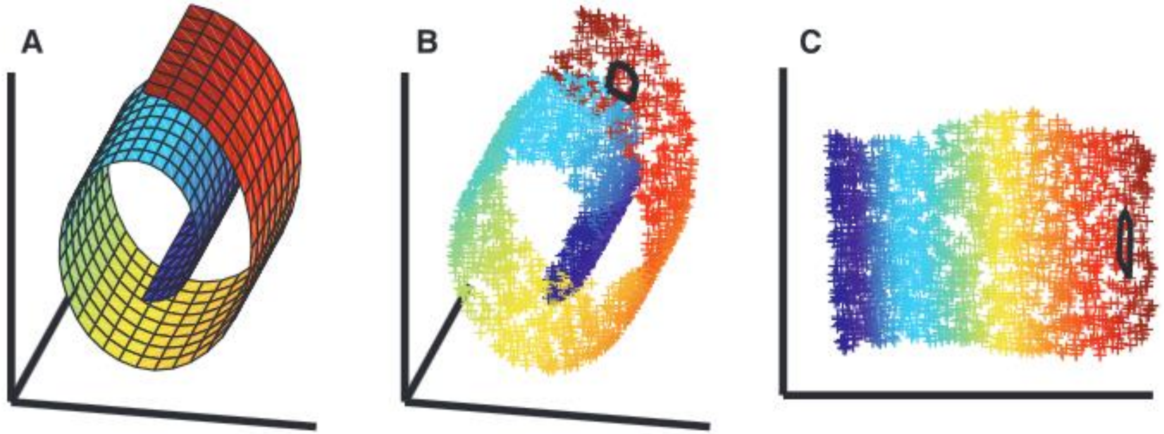
Classification of pediatric acute lymphoblastic leukemia by gene expression profiling

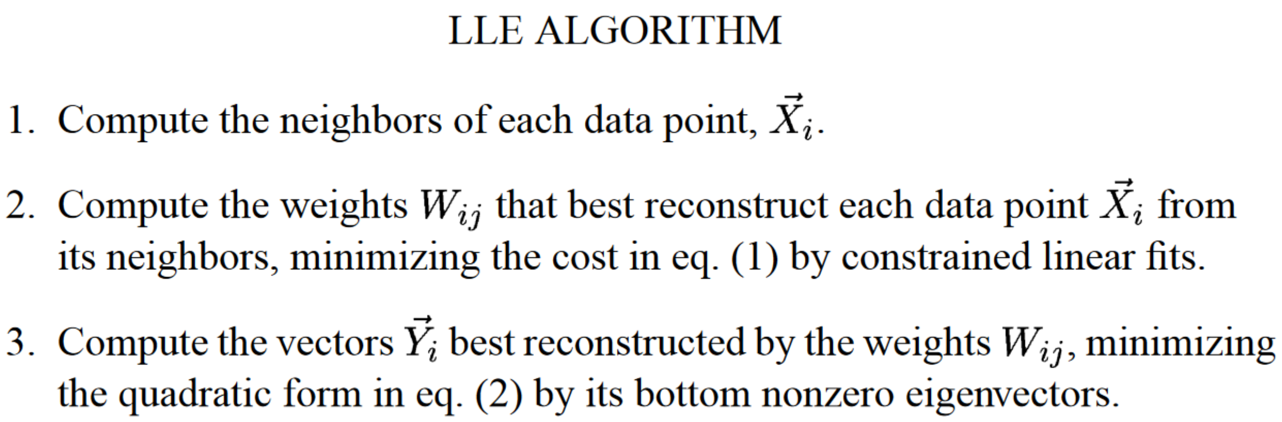
Mary E. Ross, Xiaodong Zhou, Guangchun Song, Sheila A. Shurtleff, Kevin Girtman, W. Kent Williams, Hsi-Che Liu, Rami Mahfouz, Susana C. Raimondi, Noel Lenny, Anami Patel, and James R. Downing

Contemporary treatment of pediatric acute lymphoblastic leukemia (ALL) requires the assignment of patients to specific risk groups. We have recently demonstrated that expression profiling of leukemic blasts can accurately identify the known prognostic subtypes of ALL, including T-cell lineage ALL (T-ALL), *E2A-PBX1*, *TEL-AML1*, *MLL* rearrangements, *BCR-ABL*, and hyperdiploid karyotypes with more than 50 chromosomes. As the next step toward developing this methodology into a frontline diagnostic tool, we have now analyzed leukemic blasts from 132 diagnostic samples using higher density oligonucleotide arrays that allow the interrogation of most of the identified genes in the human genome. Nearly 60% of the newly identified subtype discriminating genes are novel markers not identified in our previous study, and thus should provide new insights into the altered biology underlying these leukemias. Moreover, a proportion of the newly selected genes are highly ranked as class discriminators, and when incorporated into class-predicting algorithms resulted in an overall diagnostic accuracy of 97%. The performance of an array containing the identified discriminating genes should now be assessed in frontline clinical trials in order to determine the accuracy, practicality, and cost effectiveness of this methodology in the clinical setting. (Blood. 2003; 102:2951-2959)

Locally Linear Embedding LLE

S. T. Roweis and L. K. Saul,
Nonlinear Dimensionality Reduction by Locally Linear Embedding,
Science Vol 290, 2000

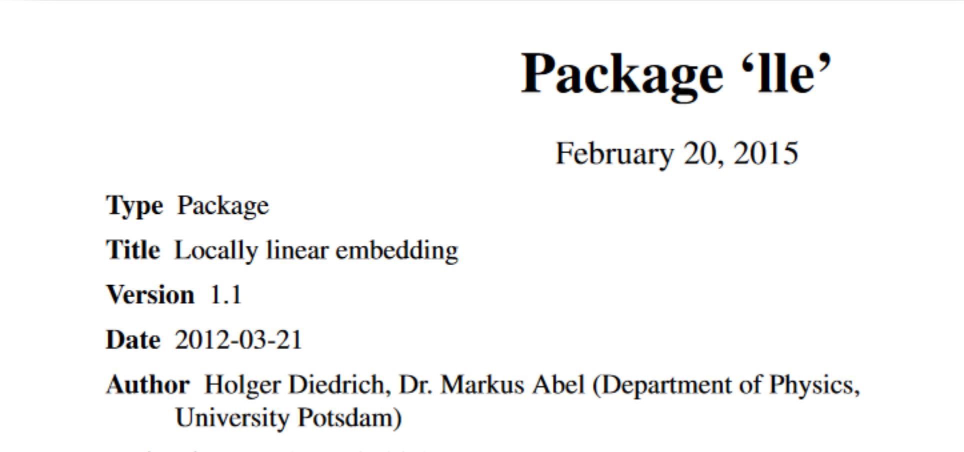




LLE ALGORITHM

1. Compute the neighbors of each data point, $\vec{X}_i$.
2. Compute the weights W_{ij} that best reconstruct each data point $\vec{X}_i$ from its neighbors, minimizing the cost in eq. (1) by constrained linear fits.
3. Compute the vectors $\vec{Y}_i$ best reconstructed by the weights W_{ij}, minimizing the quadratic form in eq. (2) by its bottom nonzero eigenvectors.

R package lle



Package ‘lle’

February 20, 2015

Type Package

Title Locally linear embedding

Version 1.1

Date 2012-03-21

Author Holger Diedrich, Dr. Markus Abel (Department of Physics, University Potsdam)

Maintainer Holger Diedrich `<holgerdiedrich@gmx.net>`

Stochastic Neighbor Embedding

Original distances

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)},$$

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}.$$

Original SNE

Distances in reduced space

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}}.$$

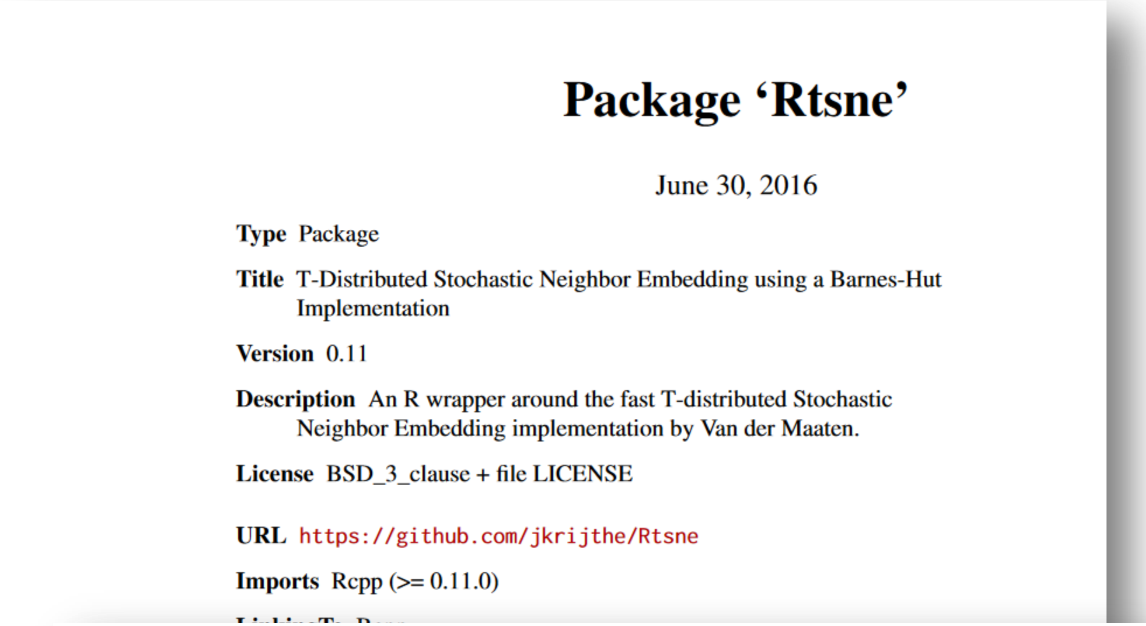
tSNE

The cost function C is given by

$$C = \sum_i KL(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}},$$

Kullback Leibler
Relative Entropy

Stochastic Neighbor Embedding (SNE) and tSNE



Package ‘Rtsne’

June 30, 2016

Type Package

Title T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation

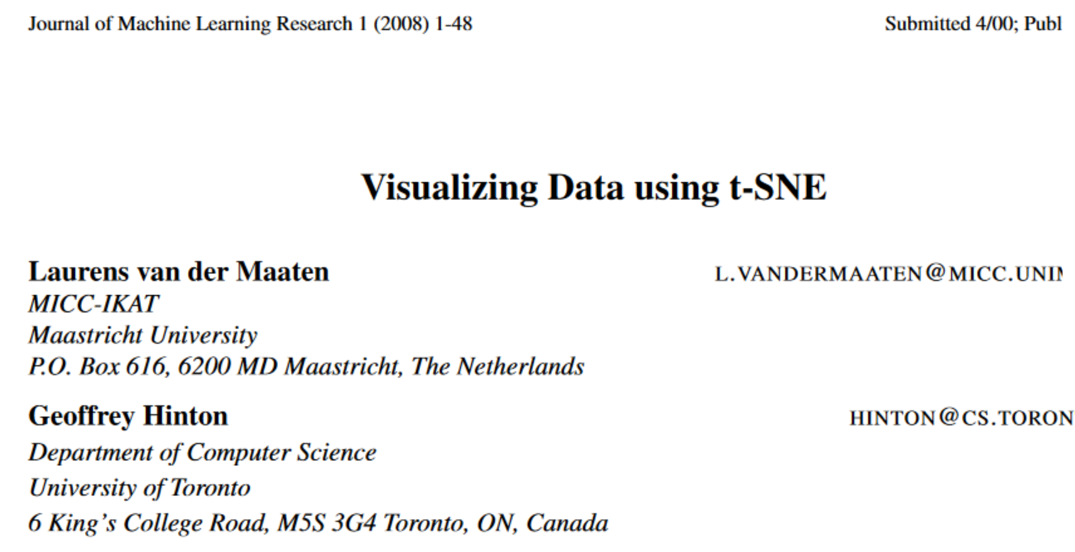
Version 0.11

Description An R wrapper around the fast T-distributed Stochastic Neighbor Embedding implementation by Van der Maaten.

License BSD_3_clause + file LICENSE

URL https://github.com/jkrijthe/Rtsne

Imports Rcpp (>= 0.11.0)

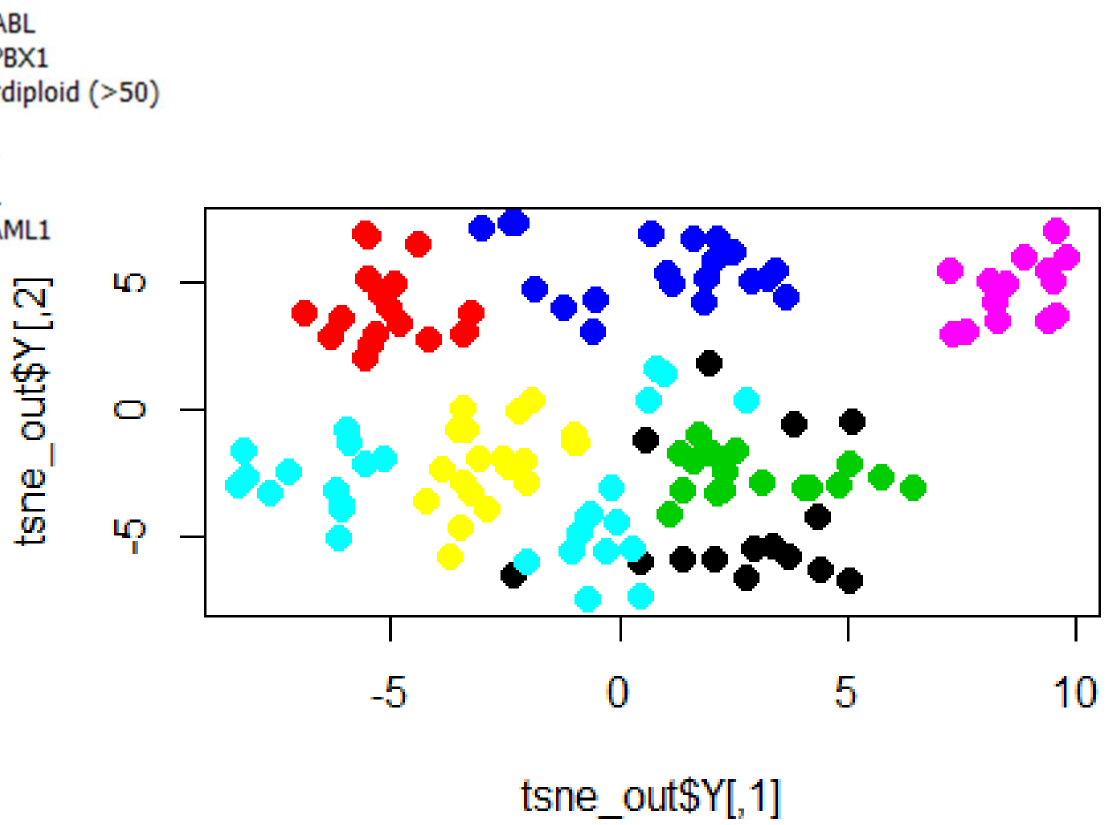


Journal of Machine Learning Research 1 (2008) 1-48 Submitted 4/00; Publ

Visualizing Data using t-SNE

Laurens van der Maaten L.VANDERMAATEN@MICC.UNI
MICC-IKAT
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands

Geoffrey Hinton HINTON@CS.TORONTO.EDU
Department of Computer Science
University of Toronto
6 King’s College Road, M5S 3G4 Toronto, ON, Canada



Copenhagen 4th-8th of December 2017

MISPCAMP 2017

Modelling of Immune system and Pathogens

mispcamp.org

The Institut Pasteur International Network for Data Analysis and the
Centre of Excellence for Personalized Medicine at Rigshospitalet