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 25.7

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

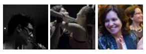
3067 readers
BY FEEDBURNER

On Facebook:

R blog…
28k likes

Like Page

1 friend likes 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 Programming for a Daily Fantasy Sports Application
- Data Scientists @ Hyderabad, Telangana, India
- Equity Research Analytics Associate @ Minneapolis, Minnesota, United States
- Data Scientist – Analytics @ Amsterdam, Noord-Holland, Netherlands
- Data Scientist – Machine Learning @ Amsterdam, Noord-Holland, Netherlands

Search & Hit Enter

Popular Searches

- heatmap
- web scraping
- maps
- hadoop
- shiny
- twitter
- boxplot
- alt=
- time series
- animation
- ggplot2
- ggplot
- how to import image file to R
- title=
- trading
- latex
- PCA
- excel
- finance
- googlevis
- RStudio
- eclipse
- quantmod
- market research
- rattle
- tutorial
- coplot
- rcmdr
- knitr
- map

Recent Posts

- How to Learn R
- Cause of Death: Melanin | Evaluating Death by Police Data
- svglite 1.0.0
- How to Run A Shiny App in the Cloud Using Tutum, Digital Ocean and Docker Containers
- Cohort analysis: Retention Rate Visualization with R
- Data Science Radar – Modeller Profile
- Fitting Generalized Regression Neural Network with Python
- Download and plot Google Trends data with R
- ezknitr: R package to avoid the typical working directory pain when using knitr
- Who’s downloading the forecast package?
- Register for Hadley Wickham’s Master R Developer Workshop – San Francisco
- The Data-Driven Weekly #1.5
- Installing

XGBoost on Ubuntu
- asdfree 2015-12-09 04:28:00
- New Course: Mapmaking in R with Choroplethr

Other sites

- SAS blogs
- Jobs for R-users
- Statistics of Israel

How to Learn R

December 10, 2015
By Tal Galili

There are tons of resources to help you learn the different aspects of R, and as a beginner this can be overwhelming. It’ s also a dynamic language and rapidly changing, so it’ s important to keep up with the latest tools and technologies.

That’ s why R-bloggers and DataCamp have worked together to bring you a learning path for R. Each section points you to relevant resources and tools to get you started and keep you engaged to continue learning. It’ s a mix of materials ranging from documentation, online courses, books, and more.

Just like R, this learning path is a dynamic resource. We want to continually evolve and improve the resources to provide the best possible learning experience. So if you have suggestions for improvement please email tal.galili@gmail.com with your feedback.

Learning Path

Getting started: The basics of R

Setting up your machine

R packages

Importing your data into R

Data Manipulation

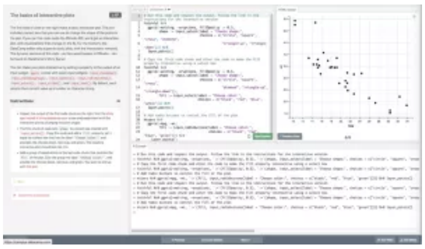
Data Visualization

Data Science & Machine Learning with R

Reporting Results in R

Next steps

Getting started: The basics of R



The best way to learn R is by doing. In case you are just getting started with R, this free introduction to R tutorial by DataCamp is a great resource as well the successor Intermediate R programming (subscription required). Both courses teach you R programming and data science interactively, at your own pace, in the comfort of your browser. You get immediate feedback during exercises with helpful hints along the way so you don’ t get stuck.

Another free online interactive learning tutorial for R is available by O’ reilly’ s code school website called try R. An offline interactive learning resource is swirl, an R package that makes if fun and easy to become an R programmer. You can take a swirl course by (i) installing the package in R, and (ii) selecting a course from the course library. If you want to start right away without needing to install anything you can also choose for the online version of Swirl.

There are also some very good MOOC’ s available on edX and Coursera that teach you the basics of R programming. On edX you can find Introduction to R Programming by Microsoft, an 8 hour course that focuses on the fundamentals and basic syntax of R. At Coursera there is the very popular R Programming course by Johns Hopkins. Both are highly recommended!

If you instead prefer to learn R via a written tutorial or book there is plenty of choice. There is the introduction to R manual by CRAN, as well as some very accessible books like Jared Lander’ s R for Everyone or R in Action by Robert Kabacoff.

Setting up your machine

You can download a copy of R from the Comprehensive R Archive Network (CRAN). There are binaries available for Linux, Mac and Windows.

Once R is installed you can choose to either work with the basic R console, or with an integrated development environment (IDE). RStudio is by far the most popular IDE for R and supports debugging, workspace management, plotting and much more (make sure to check out the RStudio shortcuts).

Next to RStudio you also have Architect, and Eclipse-based IDE for R. If you prefer to work with a graphical user interface you can have a look at R-commander (aka as Rcmdr), or Deducer.

R packages

R packages are the fuel that drive the growth and popularity of R. R packages are bundles of code, data, documentation, and tests that are easy to share with others. Before you can use a package, you will first have to install it. Some packages, like the base package, are automatically installed when you install R. Other packages, like for example the ggplot2 package, won’ t come with the bundled R installation but need to be installed.

Many (but not all) R packages are organized and available from CRAN, a network of servers around the world that store identical, up-to-date, versions of code and documentation for R. You can easily install these package from inside R, using the install.packages function. CRAN also maintains a set of Task Views that identify all the packages associated with a particular task such as for example TimeSeries.

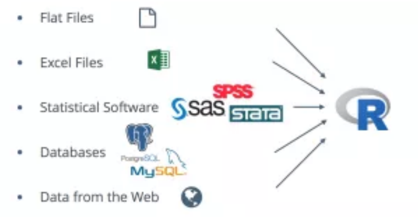
Next to CRAN you also have bioconductor which has packages for the analysis of high-throughput genomic data, as well as for example the github and bitbucket repositories of R package developers. You can easily install packages from these repositories using the devtoolspackage.

Finding a package can be hard, but luckily you can easily search packages from CRAN, github and bioconductor using Rdocumentation, inside-R, or you can have a look at this quick list of useful R packages.

To end, once you start working with R, you’ ll quickly find out that R package dependencies can cause a lot of headaches. Once you get confronted with that issue, make sure to check outpackrat (see video tutorial) or checkpoint. When you’ d need to update R, if you are using Windows, you can use the updateR() function from the installr package.

Importing your data into R

The data you want to import into R can come in all sorts for formats: flat files, statistical software files, databases and web data.



Getting different types of data into R often requires a different approach to use. To learn more in general on how to get different data types into R you can check out this online Importing Data into R tutorial (subscription required), this post on data importing, or this webinar by RStudio.

- Flat files are typically simple text files that contain table data. The standard distribution of R provides functionality to import these flat files into R as a data frame with functions such as read.table() and read.csv() from the utils package. Specific R packages to import flat files data are readr, a fast and very easy to use package that is less verbose as utils and multiple times faster (more information), and data.table’ s fread() function for importing and munging data into R (using the fread function).

- In case you want to get your excel files into R, it’ s a good idea to have a look at thereadxl package. Alternatively, there is the gdata package which has function that supports the import of Excel data, and the XLConnect package. The latter acts as a real bridge between Excel and R meaning you can do any action you could do within Excel but you do it from inside R. Read more on importing your excel files into R.

- Software packages such as SAS, STATA and SPSS use and produce their own file types. The haven package by Hadley Wickham can deal with importing SAS, STATA and SPSS data files into R and is very easy to use. Alternatively there is the foreign package, which is able to import not only SAS, STATA and SPSS files but also more exotic formats like Systat and Weka for example. It’ s also able to export data again to various formats. (Tip: if you’ re switching from SAS,SPSS or STATA to R, check out Bob Muenchen’ s tutorial (subscription required))

- The packages used to connect to and import from a relational database depend on the type of database you want to connect to. Suppose you want to connect to a MySQL database, you will need the RMySQL package. Others are for example theRpostgreSQL and ROracle package.The R functions you can then use to access and manipulate the database, is specified in another R package called DBI.

- If you want to harvest web data using R you need to connect R to resources online using API’ s or through scraping with packages like rvest. To get started with all of this, there is this great resource freely available on the blog of Rolf Fredheim.

Data Manipulation

Turning your raw data into well structured data is important for robust analysis, and to make data suitable for processing. R has many built-in functions for data processing, but they are not always that easy to use. Luckily, there are some great packages that can help you:

- The tidyr package allows you to “tidy” your data. Tidy data is data where each column is a variable and each row an observation. As such, it turns your data into data that is easy to work with. Check this excellent resource on how you can tidy your data using tidyr.

- If you want to do string manipulation, you should learn about the stringr package. The vignette is very understandable, and full of useful examples to get you started.
- dplyr is a great package when working with data frame like objects (in memory and out of memory). It combines speed with a very intuitive syntax. To learn more on dplyr you can take this data manipulation course (subscription required) and check out this handy cheat sheet.
- When performing heavy data wrangling tasks, the data.table package should be your "go-to" package. It's blazingly fast, and once you get the hang of it's syntax you will find yourself using data.table all the time. Check this data analysis course (subscription required) to discover the ins and outs of data.table, and use this cheat sheet as a reference.
- Chances are you find yourself working with times and dates at some point. This can be a painful process, but luckily lubridate makes it a bit easier to work with. Check it's vignette to better understand how you can use lubridate in your day-to-day analysis.
- Base R has limited functionality to handle time series data. Fortunately, there are package like zoo, xts and quantmod. Take this tutorial by Eric Zivot to better understand how to use these packages, and how to work with time series data in R.

If you want to have a general overview of data manipulation with R, you can read more in the book Data Manipulation with R or see the Data Wrangling with R video by RStudio. In case you run into troubles with handling your data frames, check 15 easy solutions to your data frame problems.

Data Visualization

One of the things that make R such a great tool is its data visualizations capabilities. For performing visualizations in R, ggplot2 is probably the most well known package and a must learn for beginners! You can find all relevant information to get you started with ggplot2 onhttp://ggplot2.org/ and make sure to check out the cheatsheet and the upcomming book. Next to ggplot2, you also have packages such as ggvis for interactive web graphics (seetutorial (subscription required)), googleVis to interface with google charts (learn to re-create this TED talk), Plotly for R, and many more. See the task view for some hidden gems, and if you have some issues with plotting your data this post might help you out.

In R there is a whole task view dedicated to handling spatial data that allow you to create beautiful maps such as this famous one:

To get started look at for example a package such as ggmap, which allows you to visualize spatial data and models on top of static maps from sources such as Google Maps and Open Street Maps. Alternatively you can start playing around with maptools, choroplethr, and thetmap package. If you need a great tutorial take this Introduction to visualising spatial data in R.

You'll often see that visualizations in R make use of all these magnificent color schemes that fit like a glove on the graph/map/… If you want to achieve this for your visualizations as well, then deepen yourself into the RColorBrewer package and ColorBrewer.

One of the latest visualizations tools in R is HTML widgets. HTML widgets work just like R plots but they create interactive web visualizations such as dynamic maps (leaflet), time-series data charting (dygraphs), and interactive tables (DataTables). There are some very nice examples of HTML widgets in the wild, and solid documentation on how to create your own one (not in a reading mode: just watch this video).

If you want to get some inspiration on what visualization to create next, you can have a look at blogs dedicated to visualizations such as FlowingData.

Data Science & Machine Learning with R

There are many beginner resources on how to do data science with R. A list of available online courses:

- Andrew Conway’ s Introduction to statistics with R (subscription required)
- Data Analysis and Statistical Inference
- Data Analysis for life sciences
- Data Science Specialization by Johns Hopkins

Alternatively, if you prefer a good read:

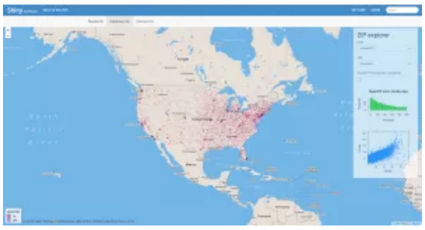
- Practical Data Science With R
- R for Data Science (upcomming, see progress)
- A Survival Guide to Data Science with R

Once your start doing some machine learning with R, you will quickly find yourself using packages such as caret, rpart and randomForest. Luckily, there are some great learning resources for these packages and Machine Learning in general. If you are just getting started,this guide will get you going in no time. Alternatively, you can have a look at the booksMastering Machine Learning with R and Machine Learning with R. If you are looking for some step-by-step tutorials that guide you through a real life example there is the Kaggle Machine Learning course or you can have a look at Wiekvoet’ s blog.

Reporting Results in R

R Markdown is an authoring format that enables easy creation of dynamic documents, presentations, and reports from R. It is a great tool for reporting your data analysis in a reproducible manner, thereby making the analysis more useful and understandable. R markdown is based on knitr and pandoc. With R markdown, R generates a final document that replaces the R code with its results. This document can be in an html, word, pfd, ioslides, etc. format. You can even create interactive R markdown documents using Shiny. This 4 hour tutorial on Reporting with R Markdown (subscription required) get’ s you going with R markdown, and in addition you can use this nice cheat sheet for future reference.

Next to R markdown, you should also make sure to check out Shiny. Shiny makes it incredibly easy to build interactive web applications with R. It allows you to turn your analysis into interactive web applications without needing to know HTML, CSS or Javascript. RStudio maintains a great learning portal to get you started with Shiny, including this set of video tutorials (click on the essentials of Shiny Learning Roadmap). More advanced topics are available, as well as a great set of examples.



Next steps

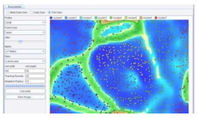
After spending some time writing R code (and you became an R-addict), you’ ll reach a point that you want to start writing your own R package. Hilary Parker from Etsy has written a short tutorial on how to create your first package, and if you’ re really serious about it you need to readR packages, an upcoming book by Hadley Wickham that is already available for free on the web.

Once you become more fluent in writing R syntax (and consequently addicted to R), you will want to unlock more of its power (read: do some really nifty stuff). In that case make sure to check out RCPP, an R package that makes it easier for integrating C++ code with R, orRevoScaleR (start the free tutorial).

Finally, if you want to start learning on the inner workings of R and improve your understanding of it, the best way to get you started is by reading Advanced R.

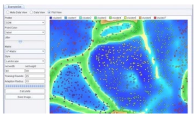
Like Share 1.3k Tweet Share 33

Related



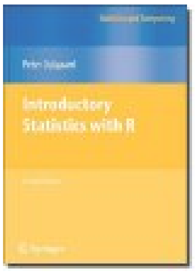
Tips on installing R extension for Rapidminer on Mac OS X

Similar post



Tips on installing R extension for Rapidminer on Mac OS X

Similar post



Getting staRted with R.

Similar post

If you got this far, why not subscribe for updates from the site? Choose your flavor: e-mail, twitter, RSS, or facebook...

Like Share 1.3k Tweet Share 33

Comments are closed.

Top 3 Posts from the past 2 days

- How to Learn R
- Scatterplots
- Installing R packages

Search & Hit Enter

Top 9 articles of the week

1. How to Learn R
2. In-depth introduction to machine learning in 15 hours of expert videos
3. Installing R packages
4. Scatterplots
5. Using apply, sapply, lapply in R
6. Computing and visualizing PCA in R
7. How to Make a Histogram with Basic R
8. Adding a legend to a plot
9. Box-plot with R – Tutorial

Sponsors

plotly
Free, open-source D3.js charting
for RStudio and Shiny

50% OFF
COUPON: BOOST50
ONLY TILL 14th DECEMBER
Become a
Certified Data Scientist
ENROLL NOW
simplilearn

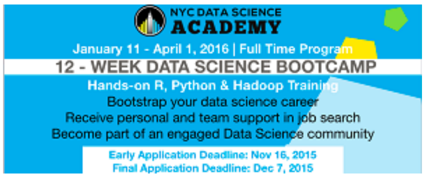
R training
R consulting
QUANTIDE
knowledge from data

R Studio
open source & enterprise ready
professional software for R

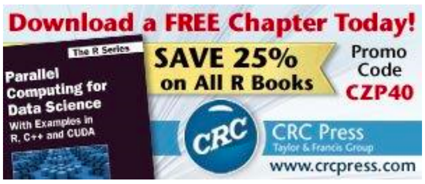
Highland Statistics Ltd
Zero Inflated Models & GLMM
Beginner's Guide to GAM
Beginner's Guide to GLM & GLMM
Beginner's Guide to GAMM

Werden Sie zum Expe[R]ten mit der
R-Akademie von
eoda
daten ▸ wissen ▸ nutzen
Beratung | Software
Training | Lösungen

STATISTICS
VIEWS
Bringing Statistics Together


NYC DATA SCIENCE
ACADEMY
January 11 - April 1, 2016 | Full Time Program
12 - WEEK DATA SCIENCE BOOTCAMP
Hands-on R, Python & Hadoop Training
Bootstrap your data science career
Receive personal and team support in job search
Become part of an engaged Data Science community
Early Application Deadline: Nov 16, 2015
Final Application Deadline: Dec 7, 2015

#ODSC
BOSTON MAY 20-22
TICKETS ON SALE NOW



Statistik-Hilfe gesucht?
Promovierte Statistiker unterstützen Sie
Jetzt unverbindlich anfragen
NOVUSTAT

Contact us if you wish to help support R-bloggers, and place your banner here.

Jobs for R users

- R Programming for a Daily Fantasy Sports Application
- Data Scientists @ Hyderabad, Telangana, India
- Equity Research Analytics Associate @ Minneapolis, Minnesota, United States
- Data Scientist – Analytics @ Amsterdam, Noord-Holland, Netherlands
- Data Scientist – Machine Learning @ Amsterdam, Noord-Holland, Netherlands
- Data Scientist @ Framingham, Massachusetts, United States
- Quantitative Analysis Specialist @ New York, United States

Search & Hit Enter

Full list of contributing R-bloggers

R-bloggers was founded by Tal Galili, with gratitude to the R community.
Is powered by WordPress using a bavotasan.com design.