



R news and tutorials contributed by (573) R bloggers

- [Home](#)
- [About](#)
- [RSS](#)
- [add your blog!](#)
- [R jobs](#) ♦♦♦
- [Contact us](#)

Welcome!

Follow

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:](#)

18204 readers

BY FEEDBURNER

[On Facebook:](#)



Me gusta esta página

Sé el primero de tus amigos en indicar que te gusta esto.



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](#)

- [Quantitative Trading Analyst](#)
- [Scientific Programmer \(@Portland\)](#)
- [Data Scientist jobs at Google \(@Mountain View\)](#)
- [Quantitative Developer \(@New York\)](#)
- [P2P Lending Credit](#)

Popular Searches

- [heatmap](#)
- [web scraping](#)
- [maps](#)
- [undefined](#)
- [hadoop](#)
- [shiny](#)
- [twitter](#)
- [boxplot](#)
- [trading](#)
- [ggplot2](#)
- [finance](#)
- [animation](#)
- [time series](#)
- [LaTeX](#)
- [excel](#)
- [ggplot](#)
- [pca](#)
- [googlevis](#)
- [quantmod](#)
- [market research](#)
- [eclipse](#)
- [RStudio](#)
- [tutorial](#)
- [how to import image file to R](#)
- [knitr](#)
- [rattle](#)
- [coplot](#)
- [rcmdr](#)
- [GIS](#)
- [alt=](#)

Recent Posts

- [Visualising the predictive distribution of a log-transformed linear model](#)
- [PDQ Version 6.2.0 Released](#)
- [How R is used at Zillow to estimate housing values](#)
- [New Pacakge “docextractr” – Easily Extract Tables From Microsoft Word Docs](#)
- [RStudio adds a new Starter Plan, More Active Hours, and a Performance Boost to shinyapps.io](#)
- [RcppDE 0.1.3](#)

- [Changing the font of R base graphic plots.](#)
- [Canberra Data Miners: Seminar on Text, Knowledge and Information Extraction, by Dr Lizhen Qu \(NICTA\), Canberra, 4:30-5:30pm, Tuesday 1 Sept](#)
- [Using R To Get Data *Out Of* Word Docs](#)
- [Predicting Titanic deaths on Kaggle IV: random forest revisited](#)
- [Modern Honey Network Machinations with R, Python, phantomjs, HTML & JavaScript](#)
- [Examining Email Addresses in R](#)
- [Pricing Game](#)
- [Where Does the S&P 500 Stand?](#)
- [Analysing longitudinal data: Multilevel growth models \(I\)](#)

Other sites

- [SAS blogs](#)
- [Statistics of Israel](#)
- [Jobs for R-users](#)

Rth: a Flexible Parallel Computation Package for R

June 17, 2014

By [matloff](#)

Like Share 82

(This article was first published on [Mad \(Data\) Scientist](#), and kindly contributed to [R-bloggers](#))

I've been mentioning here that I'll be discussing a new package, [Rth](#), developed by me and Drew Schmidt, the latter of [pbdR](#) fame. It's now ready for use! In this post, I'll explain what goals **Rth** has, and how to use it.

Platform Flexibility

The key feature of **Rth** is in the word *flexible* in the title of this post, which refers to the fact that **Rth** can be used on two different kinds of platforms for parallel computation: multicore systems and Graphics Processing Units (GPUs). You all know about the former—it's hard to buy a PC these days

that is not at least dual-core—and many of you know about the latter. If your PC or laptop has a somewhat high-end graphics card, this enables extremely fast computation on certain kinds of problems. So, whether have, say, a quad-core PC or a good NVIDIA graphics card, you can run **Rth** for fast computation, again for certain types of applications. And both multicore and GPUs are available in the Amazon EC2 cloud service.

Rth Quick Start

Our [Rth home page](#) tells you the GitHub site at which you can obtain the package, and how to install it. (We plan to place it on CRAN later on.)

Usage is simple, as in this example:

```
library(Rth)
Loading required package: Rcpp
> x <- runif(10)
> x
[1] 0.21832266 0.04970642 0.39759941 0.27867082 0.01540710
0.15906994
[7] 0.65361604 0.95695404 0.25700848 0.94633625
> sort(x)
[1] 0.01540710 0.04970642 0.15906994 0.21832266 0.25700848
0.27867082
[7] 0.39759941 0.65361604 0.94633625 0.95695404
> rthsort(x)
[1] 0.01540710 0.04970642 0.15906994 0.21832266 0.25700848
0.27867082
[7] 0.39759941 0.65361604 0.94633625 0.95695404
```

Performance

So, let's see how fast we can sort. We'll try R's built-in sort (with the default method, Quicksort), and then try **Rth** with 2 cores and then 4.

```
> system.time(sort(x))
user system elapsed
18.866 0.209 19.144
> system.time(rthsort(x,nthreads=2))
user system elapsed
5.763 0.739 3.949
> system.time(rthsort(x,nthreads=4))
user system elapsed
8.798 1.114 3.724
```

I ran this on a 32-core machine, so I could have tried even more threads, though typically one reaches a point at which increasing the number of cores actually slows things down.

The cogniscenti out there will notice immediately that we obtained a speedup of far more than 2 while using only 2 cores. This obviously is due to use of different algorithms. In this instance, the difference arises from a different sorting algorithm being used in Thrust, a software system on top of which **Rth** runs. (See the **Rth** home page for details on Thrust.)

Rth is an example of what I call Pretty Good Parallelism (an allusion to Pretty Good Privacy). For certain applications it can get you good speedup on two different kinds of common platforms (multicore, GPU). Like most parallel computation systems, it works best on very regular, “embarrassingly parallel” problems. For very irregular, complex apps, one may need to resort to very detailed C code to get a good speedup.

Platforms

As mentioned, the code runs on top of Thrust, which runs on Linux, Mac and Windows OSs. Also, it uses **Rcpp**, which is cross-platform as well.

In other words, **Rth** should run under all three OSs. However, so far it has been tested only on Linux and Mac platforms. It should work fine on Windows, but neither of us has ready access to such a machine, so it hasn't been tested there yet.

Necessary Programming Background

As seen above, the **Rth** functions are just R code, hence usable by anyone familiar with R. No knowledge of Thrust, C++, GPU etc. is required.

However, you may wish to write your own **Rth** functions. In fact, we hope you can contribute to the package! For this you need a good knowledge of C++, which is what Thrust is written in.

What Functions Are Available, And What Might Be Available?

Currently the really fast operations available in Rth are: sort/order/rank; distance computation; histograms; and contingency tables. These can be used as foundations for developing other functions. For example, the parallel distance computation can be used to write code for parallel k-means clustering, or for kernel-based nonparametric multivariate density estimation. Some planned new functions are listed on the home page.

Conclusion

Give **Rth** a try! Let us know about your experiences with it, and again, code contributions would be highly welcome.

I plan to devote some of my future blog posts here to other topics in parallel computation. Much of the material will come from my forthcoming book, *Parallel Computation for Data Science*.

Comments: 13

Related

Comments: 13

Rth: a Flexible Parallel Computation Package for R.

In "R. bloggers"

Parallel Programming with GPU's and R.

by Norman Matloff

You've heard that graphics processing units — GPU's — can bring big increases in

In "R. bloggers"

Add a Comment

Intermediate High-Performance Computing in R.

In "R. bloggers"

82

Like

Share

To **leave a comment** for the author, please follow the link and comment on his blog: [Mad\(Data\) Scientist](#).

[R-bloggers.com](#) offers [daily e-mail updates](#) about [R](#) news and [tutorials](#) on topics such as: visualization ([ggplot2](#), [Boxplots](#), [maps](#), [animation](#)), programming ([RStudio](#), [Sweave](#), [LaTeX](#), [SQL](#), [Eclipse](#), [git](#), [hadoop](#), [Web Scraping](#)) statistics ([regression](#), [PCA](#), [time series](#), [trading](#)) and more...

If you got this far, why not **subscribe for updates** from the site?

Choose your flavor: [e-mail](#), [twitter](#), [RSS](#), or [facebook](#)...

Like Share  82

Comments are closed.

Top 3 Posts from the past 2 days

- [In-depth introduction to machine learning in 15 hours of expert videos](#)
- [Installing R packages](#)
- [RStudio adds a new Starter Plan, More Active Hours, and a Performance Boost to shinyapps.io](#)

Search & Hit Enter

Top 9 articles of the week

1. [5 New R Packages for Data Scientists](#)
2. [In-depth introduction to machine learning in 15 hours of expert videos](#)
3. [Installing R packages](#)
4. [Using apply, sapply, lapply in R](#)
5. [Data Manipulation with dplyr](#)
6. [R courses on basic R, advanced R, statistical machine learning with R, text mining with R, spatial modelling with R and R package building](#)
7. [Importing Data Into R – Part Two](#)
8. [Scatterplots](#)
9. [Basics of Histograms](#)

Sponsors



Highland Statistics Ltd

Zero Inflated Models & GLMM

Beginner's Guide to GAM

Beginner's Guide to GLM & GLMM

Beginner's Guide to GAMM





 **DataCamp**
Learn R Interactively



#ODSC
SAN FRANCISCO
NOV. 14TH - 15TH
www.opendatascicon.com

[LEARN MORE](#)

Save **25%** on
R Books

 **CRC Press**
Taylor & Francis Group



Use
Promo
Code
CZP40

Plus Free
Shipping

**Data Analysis
Modeling**
Cloud Solution



[Free Sign Up!](#)



STATWORX

Consulting
Schulung
Data Mining

[Mehr erfahren](#)



**Try the FASTEST ML
for R**




[Click for a Free Trial](#)

 **YOTTAMINE
ANALYTICS**

Become a
Certified Data Scientist


[ENROLL NOW](#)

simplilearn



DATA FESTIVAL
SEPT 17TH-24TH • N.E.R.D CTR, CAMBRIDGE

[LEARN MORE](#)





EARN YOUR MASTER'S IN
DATA SCIENCE **online.**

Berkeley
UNIVERSITY OF CALIFORNIA

LEARN MORE



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.

Copyright © 2015 [R-bloggers](#). All Rights Reserved. [Terms and Conditions](#) for this website