



# R: fast and big strategies

Adolfo Álvarez  
Analyx

2015 Use R! Lightning talk

# Introduction

- Size is limited by RAM. Measure size by [`print\(object.size\(\), units="Mb"\)`](#), and [`pryr::object\_size\(\)`](#)
- Time is limited by your life. Measure time by [`system.time\(\)`](#), [`rbenchmark::benchmark\(\)`](#), [`microbenchmark::microbenchmark\(\)`](#), [`lineprof::lineprof\(\)`](#)



**One R Tip a Day**  
@RLangTip

 Follow

Compare the speed of several R expressions with the benchmark function: [bit.ly/Sppip1](http://bit.ly/Sppip1) #rstats

6:06 PM - 21 Oct 2013

  5  16

# 1) Write efficient code

- Vectorize
- \*apply family
- Use linear algebra
- Avoid loops



**Adolfo Álvarez**  
@adolfoalvarez

Follow

In case you missed it: Vectorization, matrix operations and readability in [#rstats](#) [adolfoalvarez.cl/vectorization-...](#)

5:42 PM - 7 Aug 2014



**Adolfo Álvarez**

**Vectorization, matrix operations, and readability**

My thoughts about R and analytics.

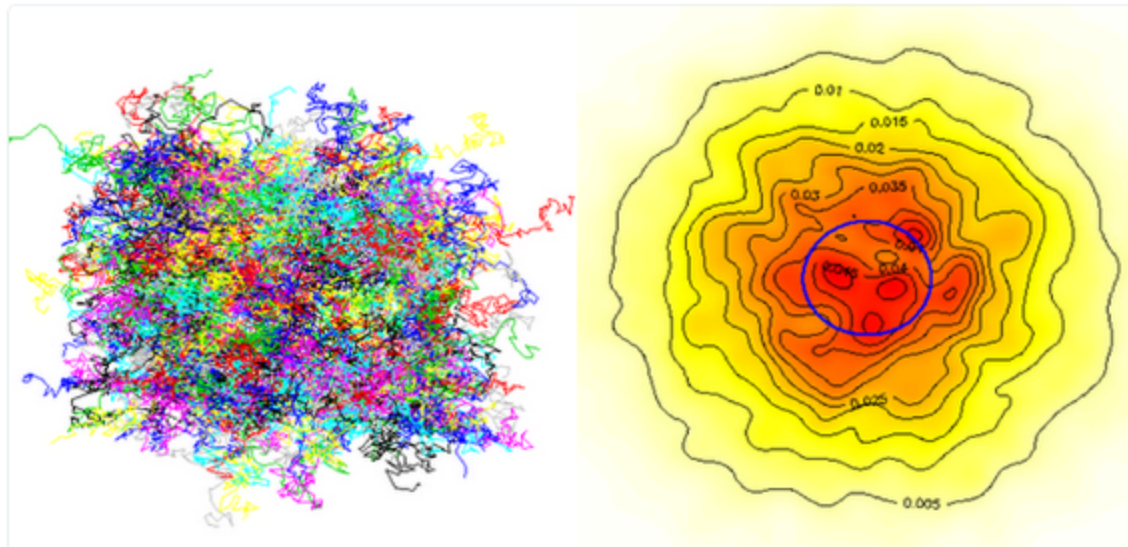
[View on web](#)



★ 1

## 2) Use modern data input and manipulation packages

- Related packages: [data.table](#) / [fread](#), [dplyr](#) / [readr](#)



**Arthur Charpentier**  
@freakonometrics



"Working with 'large' datasets, with dplyr and data.table" (in #rstats) [freakonometrics.hypotheses.org/19645](https://freakonometrics.hypotheses.org/19645)

9:59 PM - 4 May 2015 · Rennes, Bretagne, France

🔄 8 ★ 29

# 3) Connect R to most SQL and no-SQL databases

- Related packages: [dplyr](#), [RCassandra](#), [rmongodb](#), [RSQLite](#), [RPostgreSQL](#), [ROracle](#), [RMySQL](#), [monetdb.R](#), etc...



**Carl Boettiger**  
@cboettig



Great introduction to the awesomeness that is [#rstats](#) tools for modern NoSQL databases from [@rOpenSci ropensci.org/blog/2015/05/2...](#)

6:45 PM - 20 May 2015

← ↻ 7 ★ 11

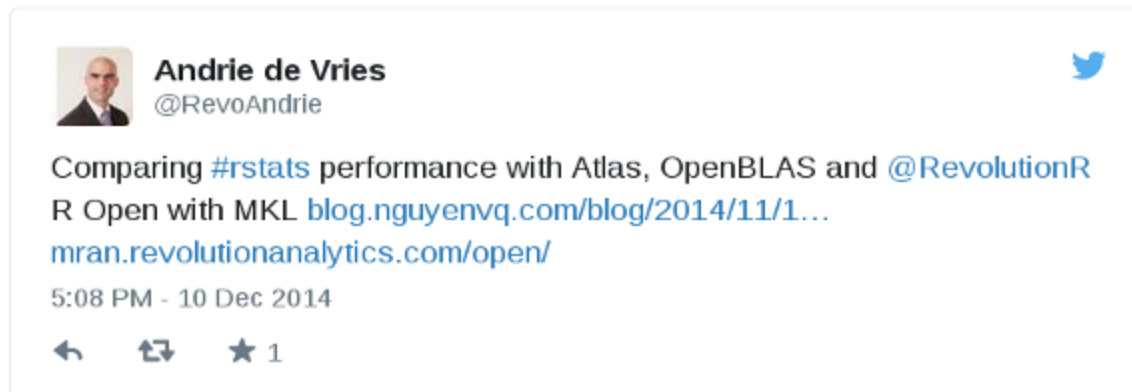
## 4) Include C++ code to reduce execution time.

- Related packages: [Rcpp, RcppArmadillo, RcppEigen, RInside](#)



# 5) Connect R to different Basic Linear Algebra Subprograms routines.

- [Blas, OpenBlas, Atlas, Intel MKL, ...](#)



A screenshot of a Twitter post. The user is Andrie de Vries (@RevoAndrie). The tweet text reads: "Comparing #rstats performance with Atlas, OpenBLAS and @RevolutionR R Open with MKL [blog.nguyenvq.com/blog/2014/11/1...](http://blog.nguyenvq.com/blog/2014/11/1...) [mran.revolutionanalytics.com/open/](http://mran.revolutionanalytics.com/open/)". The tweet is dated 5:08 PM - 10 Dec 2014 and has 1 star. The tweet is enclosed in a light blue rounded rectangle.

## 6) Try another open or proprietary implementations of the R language.

- [Revolution R](#), [TERR](#), [Oracle R](#), [pqR](#), [FastR](#), [Renjin](#), [Riposte](#).



A screenshot of a tweet from Adolfo Álvarez (@adolfoalvarez) dated 10:54 AM on 11 Nov 2014. The tweet contains a link to a blog entry titled "The free, open, and proprietary flavors of R" with the hashtag #rstats. Below the tweet, there is a preview of the blog post, including the title "The free, open, and proprietary flavors of R" and the beginning of the text: "In this fifth entry of my blog about R and analytics, I write about the origins of R as a free software project, an analysis of the several licenses used by packages in CRAN (including the use of...". To the right of the text is a small image of a white building with a window. At the bottom of the tweet are icons for reply, retweet (3), and favorite (4).

 **Adolfo Álvarez**  
@adolfoalvarez

New entry of my blog: "The free, open, and proprietary flavors of R" #rstats  
[adolfoalvarez.cl/the-free-open-...](http://adolfoalvarez.cl/the-free-open-...)

10:54 AM - 11 Nov 2014

 **Adolfo Álvarez**

**The free, open, and proprietary flavors of R**

In this fifth entry of my blog about R and analytics, I write about the origins of R as a free software project, an analysis of the several licenses used by packages in CRAN (including the use of...

[View on web](#)

← ↻ 3 ★ 4



## 7) Work in parallel

- Related packages: [snow](#), [parallel](#), [doParallel](#), [doSNOW](#)



**Jason H. Moore, Ph.D**  
@moorej



#datascience RT @Rbloggers: How-to go parallel in R – basics + tips  
[wp.me/pMm6L-o5T](http://wp.me/pMm6L-o5T) #rstats

3:58 PM - 20 Feb 2015

← ↻ 8 ★ 8

# 8) Use the power of your GPU

- Related packages: [gputools](#), [gmatrix](#), [Rth](#)



# 9) Compile your code to make it faster

- Related packages: [compiler](#)



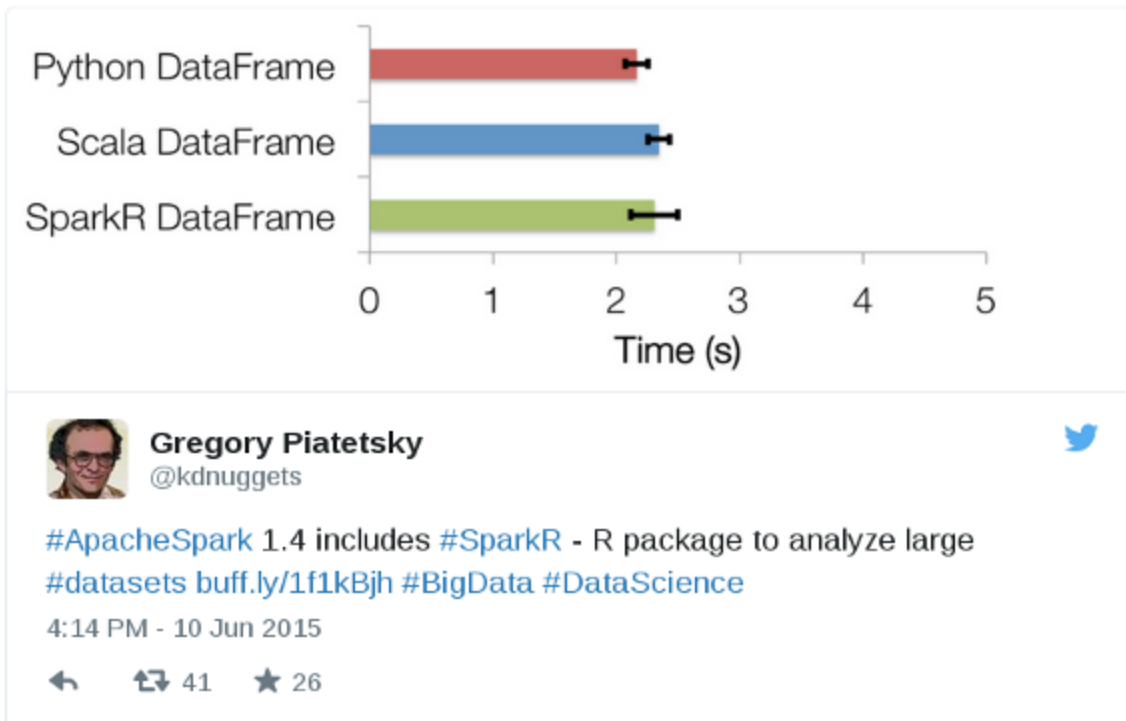
# 10) Use big data specific packages:

- [ff](#), [bigmemory](#), [pbdR](#), [biglm](#), [bigrf](#), [RevoScaleR](#),...



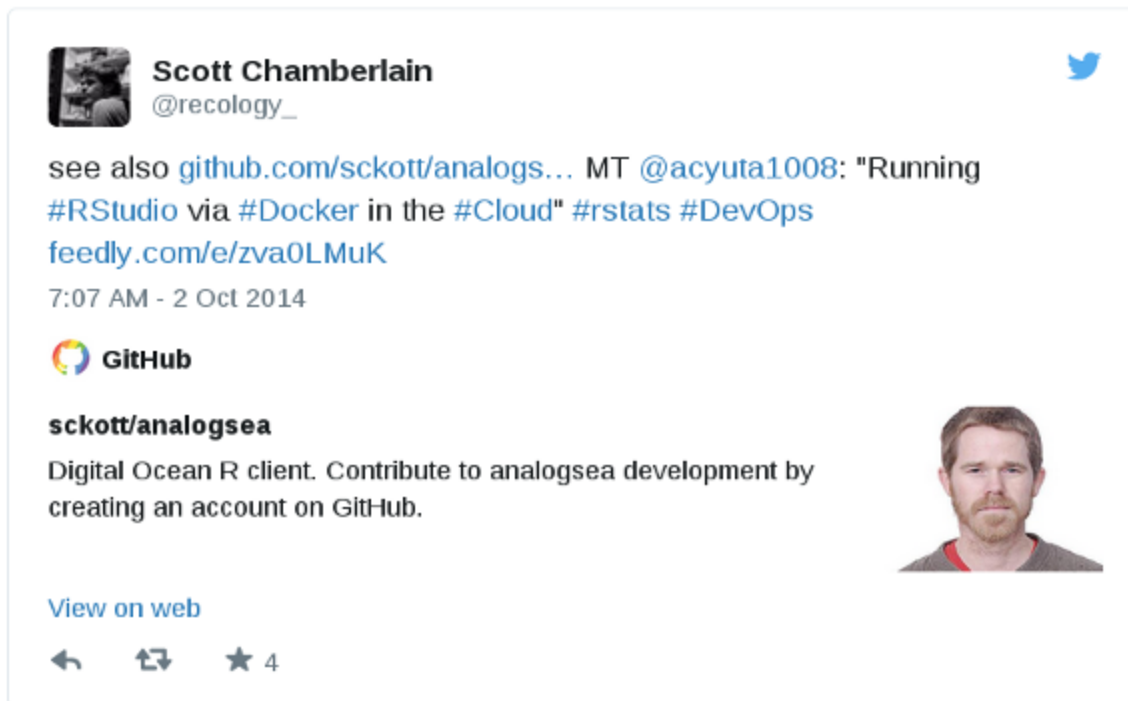
# 11) Connect R to modern big data environments:



- [RHadoop](#), [SparkR](#)



# 12) Set up R cloud instances.

- [Digital Ocean](#), [Amazon](#), [Azure](#), etc.
- Related packages: [cloudyr](#)



 **Scott Chamberlain**  
@recology\_ 


see also [github.com/sckott/analogsea](https://github.com/sckott/analogsea) MT @acyuta1008: "Running #RStudio via #Docker in the #Cloud" #rstats #DevOps  
[feedly.com/e/zva0LMuK](https://feedly.com/e/zva0LMuK)

7:07 AM - 2 Oct 2014




 **GitHub**

**sckott/analogsea**

Digital Ocean R client. Contribute to analogsea development by creating an account on GitHub.

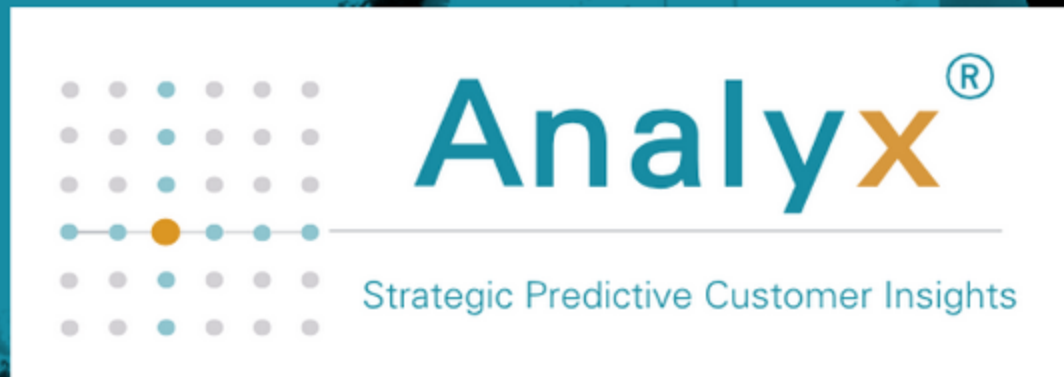


[View on web](#)

   4

*Thank you and keep contact!*

[www.adolfoalvarez.cl](http://www.adolfoalvarez.cl)  
[@adolfoalvarez](mailto:@adolfoalvarez)



[www.analyx.com](http://www.analyx.com)