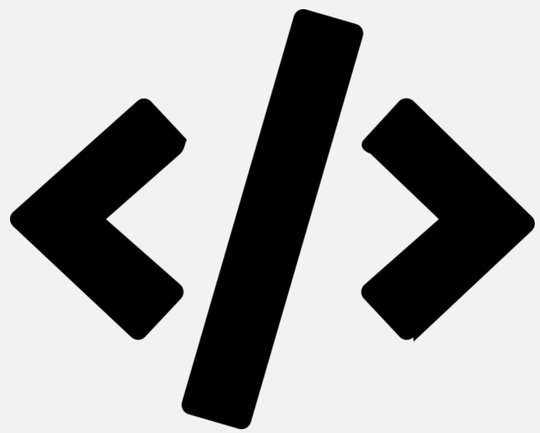
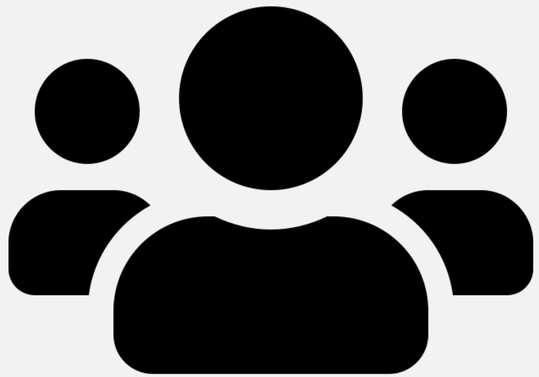
ROpenSci

ropensci.org @ropensci

Stefanie Butland
@stefaniebutland



Software



Community

Peer review

rOpenSci by the numbers

5 staff

1 postdoc

2 Bioconductor pkgs

192 CRAN pkgs

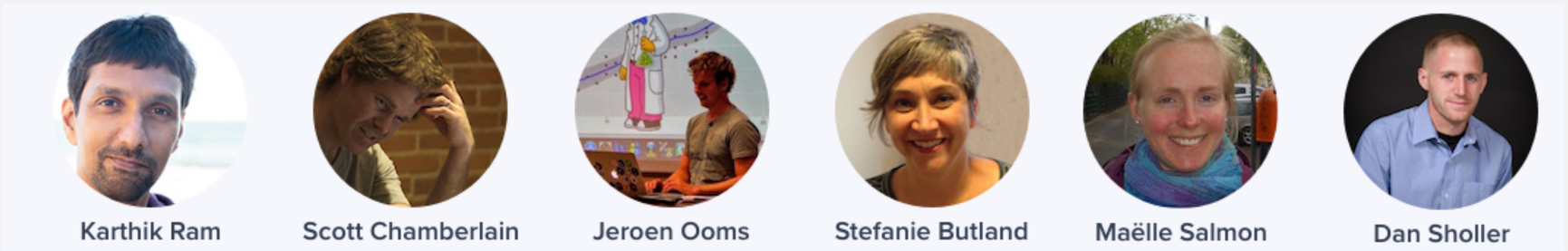
287 total pkgs

~ 500 code contributors

> 500 citations

LOTS! of community members

1 awesome open system for peer review of software



Lack of reproducibility is quite widespread in applied computational research

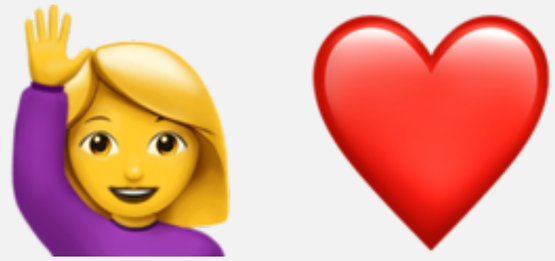
The extent to which code would actually build with reasonable effort is quite low

< 20%

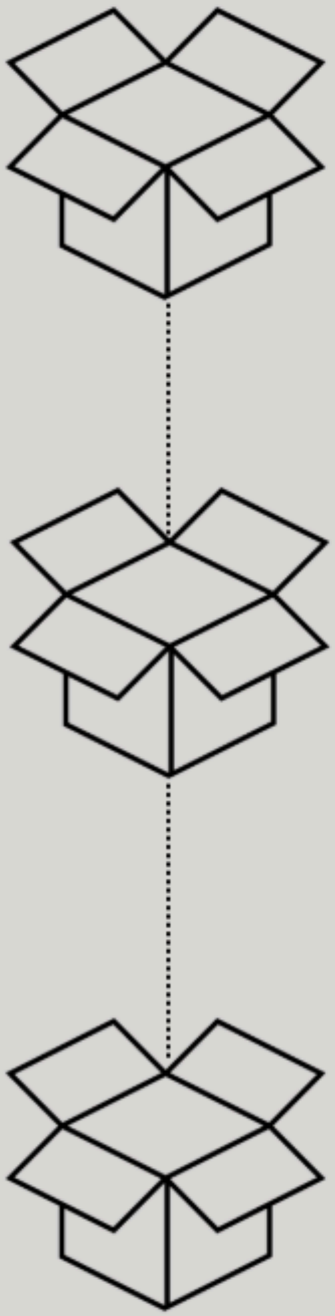
Collberg et al 2014


find R tools for your research here
they will work

We need to create a **culture** around peer reviewing our research software



academic peer review of research publications?



🤔 Pre-submission inquiry

Fit based on our criteria

Peer-review

evaluate the package for usability, quality, and style based on our guidelines

Acceptance 💯

Packages are badged and added to our system

A software review thread

“This type of review where the reviewers actively help you as well as objectively evaluating your work is a revelation”

-Rory Nolan

benmarwick commented 15 days ago — Member

Yes, thanks, I'm happy. You've done lots of work to address the concerns in my review, that's excellent to see. The package is much more accessible now, and it's easier to see how to get started using it.

1

wlandau-lilly commented 15 days ago — Member

Thank you, **@benmarwick**! I am so glad you think the changes are making a difference.

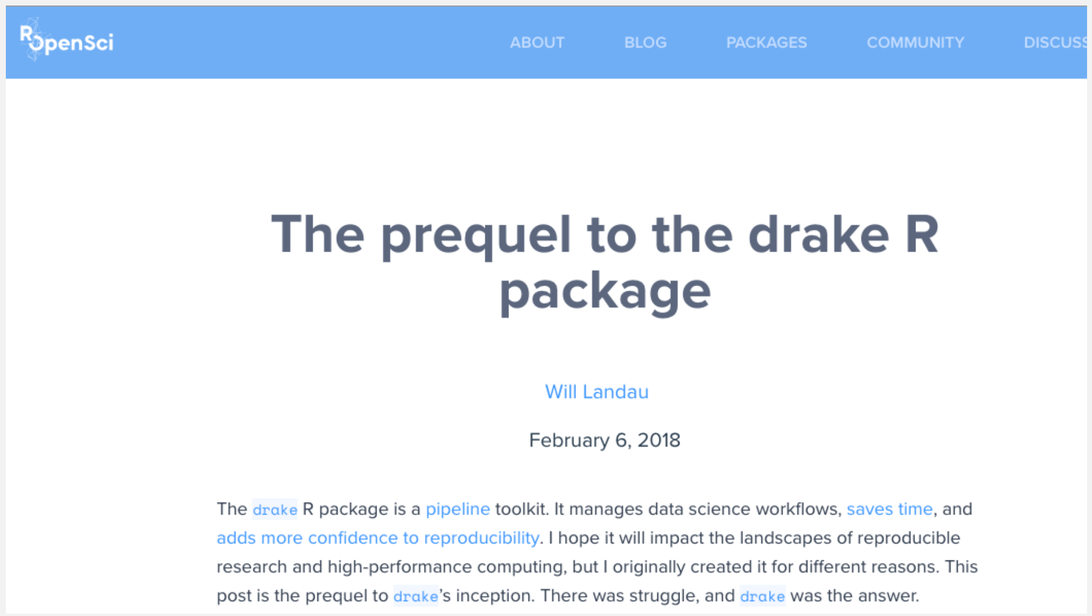
By the way, I just updated the review summaries in light of wlandau-lilly/drake#195.

gothub commented 15 days ago — Member

All of the points I mentioned have been fully addressed with clear and complete documentation, and I have no other issues, so thumbs up from me.

I have one remaining question that is not really an issue for the review, but I'm just curious about. If i were to develop a workflow with drake, what drake specific artifacts would need to be preserved for a researcher to reproduce my analysis (other than the R scripts and data that may have existed before even starting to build the drake workflow)? I'm assuming that it's simply a CSV from the workflow itself, e.g. 'my_plan.csv' from the basic example. If more than that is required then explaining that in the docs or having an export function would be a good idea.

1

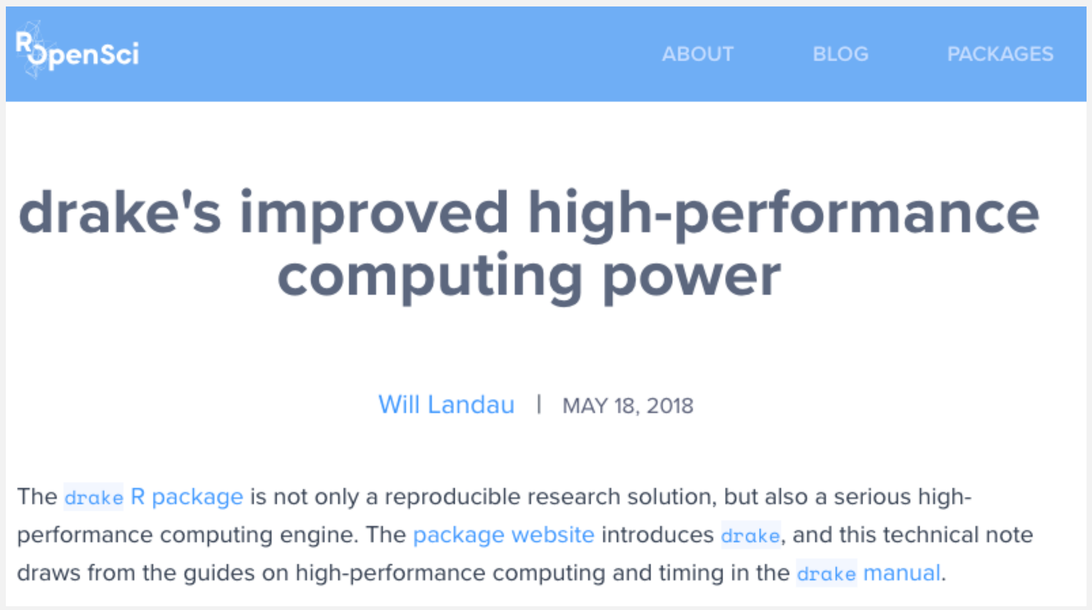


The prequel to the drake R package

Will Landau

February 6, 2018

The `drake` R package is a pipeline toolkit. It manages data science workflows, saves time, and adds more confidence to reproducibility. I hope it will impact the landscapes of reproducible research and high-performance computing, but I originally created it for different reasons. This post is the prequel to `drake`'s inception. There was struggle, and `drake` was the answer.



drake's improved high-performance computing power

Will Landau | MAY 18, 2018

The `drake` R package is not only a reproducible research solution, but also a serious high-performance computing engine. The package website introduces `drake`, and this technical note draws from the guides on high-performance computing and timing in the `drake` manual.

Preface

I Building Your Package

1 Packaging Guide

1.1 Package name and metadata

1.2 Package API

1.3 Function and argument naming

1.4 Code Style

1.5 README

1.6 Documentation

1.7 Documentation website

1.8 Authorship

1.9 Testing

1.10 Examples

1.11 Package dependencies

1.12 Recommended scaffolding

1.13 Miscellaneous CRAN gotchas

1.14 Further guidance

2 Continuous Integration Best Practices

2.1 Why use continuous integration …

2.2 How to use continuous integrati…

2.3 Which continuous integration se…

2.4 Test coverage

rOpenSci Packages: Development, Maintenance, and Peer Review

rOpenSci onboarding editorial team: Scott Chamberlain, Anna Krystalli, Lincoln Mullen, Karthik Ram, Noam Ross, Maëlle Salmon

2018-08-03

Preface

This book has the ambition to become a guide for maintainers of rOpenSci packages, in particular, people who volunteer to submit a package to onboarding.

The first section of the book presents our guidelines for building and testing your package.

The second section is dedicated to onboarding: what it is, our policies, and specific guides for authors, editors and reviewers.

The third and last section features our best practice for nurturing your package once it has been onboarded: how to collaborate with other developers, how to document releases, how to promote your package and how to leverage GitHub as a development platform. The third section also features a chapter for anyone wishing to start contributing to rOpenSci packages.

We hope that you'll find the guide useful and clear, and welcome your suggestions in the issue tracker of the book. Happy R packaging!

The rOpenSci editorial team.

Review for us

ropensci.org/onboarding

... you are qualified as a potential package reviewer if you have some appreciation for what makes your favourite packages useful.

- Miles McBain

ropensci.org/packages/

Workflow
- drake

Visualization
- plotly
- visdat
- skimr

Image manipulation
- magick
- ijtiff

Unlocking text and data
- unrtf
- pdftools
- tabulizer
- suppdata

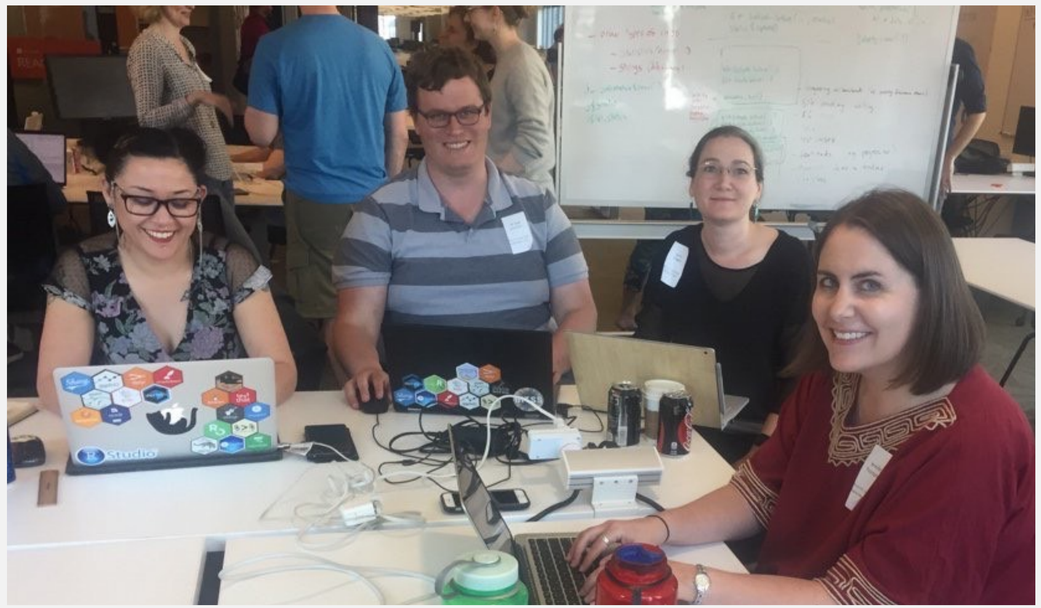
Genomic data
- genbankr
- rentrez
- rsnps
- cregulome

Patents
- patentsview

NLP
- googleLanguageR

“I often need to calculate percentiles, z scores, and other measures of growth in maternal & child health research. There are some SAS macros out there and a couple of R packages...[but they] don't have all of the measures I need...There are other measures that are just a data table of LMS parameters in a PDF. Ideally these methods would be available all in one place in an R package!”

Monica Gerber
Biostatistician,
Mass General



rOpenSci unconf18

Charles T. Gray
W. Kyle Hamilton
Jenny Draper
Jennifer Thompson

Maternal Child Health Toolbox

percentiles and z-scores based on growth charts


```
      cid height weight sex bmiz       waz   dist_
250 64464   68.1    6.8   1   NA -1.847064 2.15236
```

cid	sex	agemos	weight	height	headcir
51755	1	18	12.10	80.10	NA
51840	1	25	13.30	89.10	NA
51962	2	2	4.50	56.60	38.80
52066	2	8	9.50	72.50	NA
52089	1	32	14.20	93.60	NA
52300	1	2	6.90	63.40	41.90

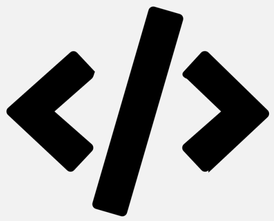
Outcomes


github.com/ropenscilabs/mchtoolbox (experimental!)


collateral learning


www.monicagerber.com/2018/07/roundup-of-growth-chart-packages/

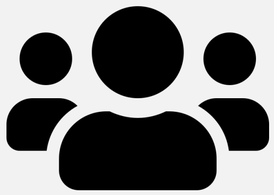

ropensci.org/blog/2018/07/05/mchtoolbox/



ropensci.org/packages/

github.com/ropensci/onboarding

ropensci.github.io/dev_guide/



ropensci.org/community/

Holly Kirk 12:04 AM

SOMEONE ASKED ME AT A CHRISTMAS PARTY WHAT HAS BEEN A HIGHLIGHT OF 2018. I SAID JOINING THE ROPENSCI COMMUNITY. AND THEN FELT BOTH SUPER PROUD AND A BIT EMBARRASSED/NERDY AT THE SAME TIME.

❤️ 5 🙌 3 1 1

ROpenSci

ropensci.org @ropensci

Stefanie Butland
@stefaniebutland