Navigating the R Package Universe

John C. Nash
Telfer School of Management
University of Ottawa
nashjc@uottawa.ca

Julia Silge
Stack Overflow
julia.silge@gmail.com

Spencer Graves
EffectiveDefense.org
spencer.graves@effectivedefense.org

Ludovic Vannoorenberghe
DataCamp
ludo@datacamp.com

Background

- Initial name "Navigating the R package jungle"
- Jungles -- rain forests -- are places rich in resources.
  - more than 10 000 packages in CRAN
  - many vignettes and Blogs
  - more stuff in Bioconductor, Github, and other collections
- Resources are often difficult to find
- Forest is usually hard to navigate

Some responses to the challenge

- **Wrappers** -- packages that **unify** the call to a number of resources for a common set of tasks (JN)
- **Task Views** -- **Guidance** on resources and how their development, timeliness and accessibility can be improved (JS)
- **Search** -- improving how users can find the tools they need and information on how to use them effectively and efficiently
  - The “sos” package (SG)
  - Rdocumentation (LV)
  - RStudio CRANsearcher
- **MORE** – you!

Unifying packages

Best seen via an example: "optimization" (function minimization)

- optim(), nlm() and nlminb() in base R
- quite large number of individual packages: BB, dfoptim, Rcgmin, Rvmmin, Rtnmin, lbfgs, lbfgs3, trust, trustOptim, nloptr, minqa, powell, and others
- MANY and DIFFERENT calling sequences
- MANY control parameters, some with same name but different function, others with different names for same functionality

Unifying packages

Response: package **optimrx** (prev. **optimx**)

- function optimr() uses optim() calling sequence with more choices for "method="
- ongoing development
- extra functions opm(), multistart(), polyopt()

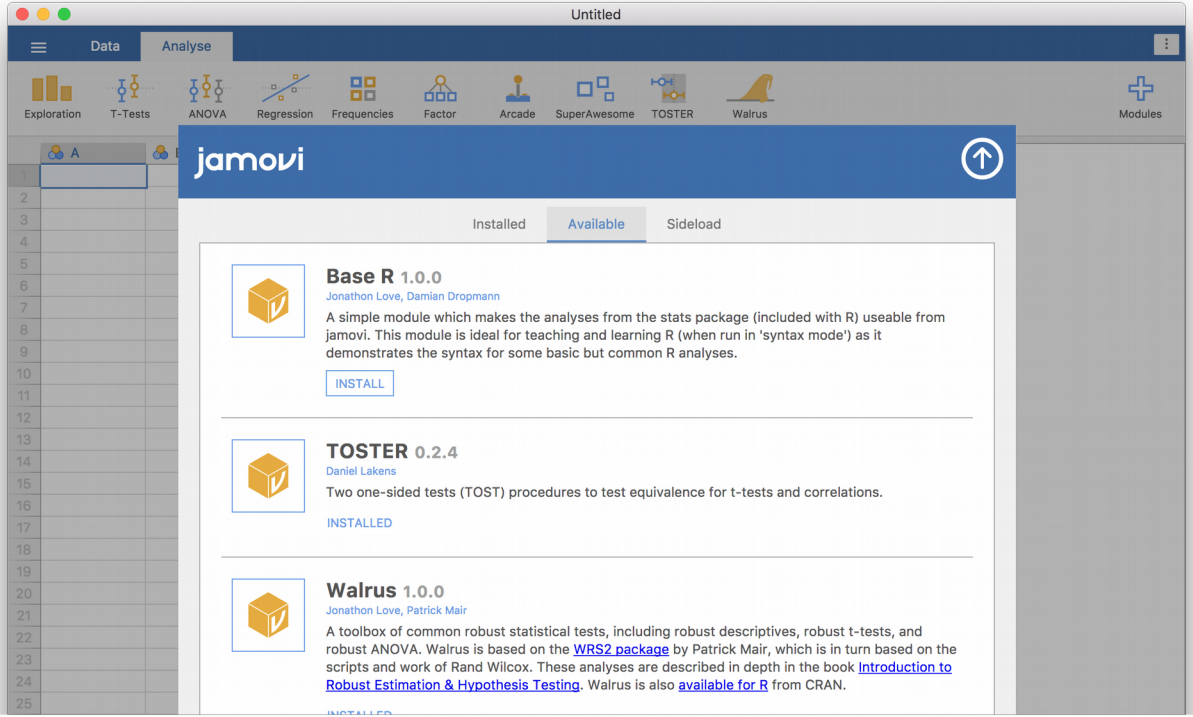
Other unification efforts

- **gloptim** (Hans Werner Borchers) global / stochastic optimization
- **bbmle** (Ben Bolker) some integration of tools for maximum likelihood estimation
- **jmv** (Jamovi) (Jonathan Love) attempts to integrate many common statistical tests
- Have I missed good examples? Let me know! (nashjc _at_ uottawa.ca)

Other unification efforts

jamovi is a graphical spreadsheet for R

(Jonathan Love will be presenting at 11:36 on the thursday)



Opportunities for unification?

- Principal Components / svd -- (JN and Claudia Beleites) https://gitlab.com/nashjc/svdpls
- Nonlinear modeling -- better integration of nls(), packages **nlsr**, **nls2** and **minpack.lm**, though the gains may be small
- Are there opportunities to simplify or streamline the user experience with database access? With data manipulation and display (plyr, dplyr, tables, others)?

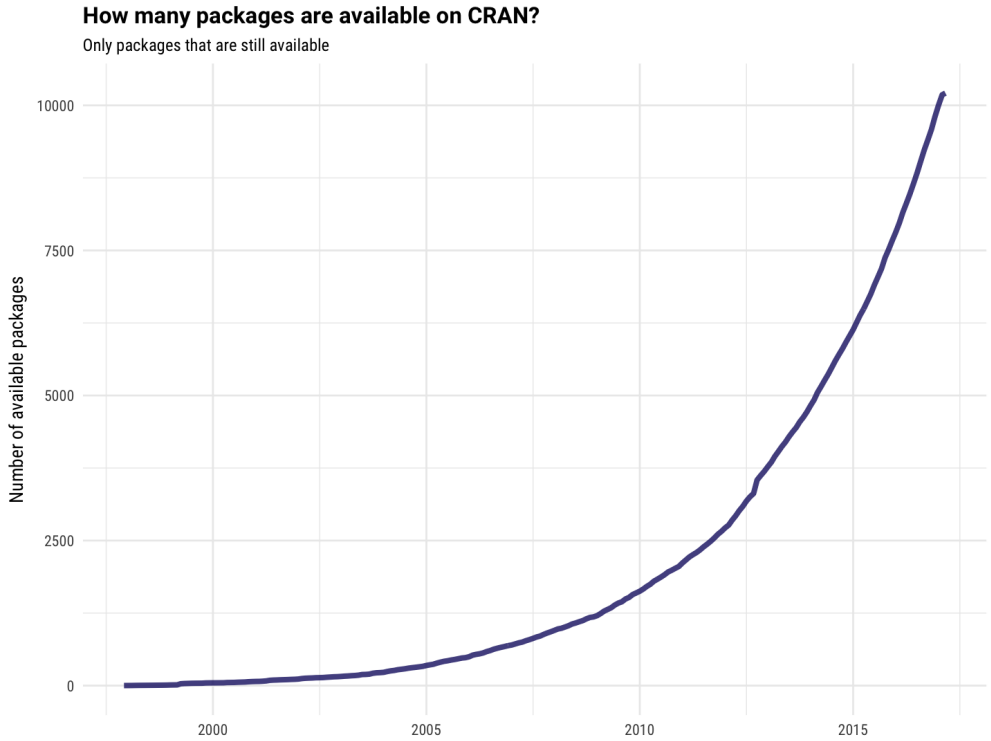
Opportunities to highlight or conceal packages

- Do we need to see a list of all packages as a default in CRAN?
- Lists by task or application?
- Lists by "popularity" of call? (Paul Gilbert 2piQA)
- Hide "infrastructure" packages from general users
- Omit some "junk" from the streamlined lists
- Note that such lists can be external to CRAN, i.e., wrappers

Where to now?

- Form groups to identify opportunities in unification, guidance or search
- Encourage/start projects to actually try out ideas
- Note Google Summer of Code and R Foundation initiatives
- https://github.com/nashjc/Rnavpkg/
- https://github.com/nashjc/Rnavpkg/wiki

Packages on CRAN




many library
such install
wow
very cran
much packages

GOOD NEWS

BAD NEWS

What this means for R users

- There are many resources out there for many differents kinds of tasks
- It can be difficult to find what you are looking for
- Assessing quality can be a challenge

package	published	dl_last_month	stars	tidyverse_happy	has_tests	vignette	last_commit	last_issue_closed	contributors	depends_count	reverse_count
arsenal	2017-03-10	236		✓	✓					2	0
ascii	2011-09-29	1806	19		✗	✓		64.8	0	3	1
compareGroups	2017-03-14	929			✗					5	0
condformat	2017-05-18	1252	4	✓	✓		0.6	0.6	4		0
desctable	2017-05-15	494	26	✓	✗		0.8	0.8	1	1	0
DT	2016-08-09	35232	206	✓	✗		2.8	0.1	6		31
expss	2017-04-09	222	7		✗		2		3	3	0
ezsummary	2016-07-11	168	30	✓	✓			11.7	0	1	0
formattable	2016-08-05	2753	299		✓		9.5	2	6	1	2
gmodels	2015-07-22	29699			✗	✓				1	7
gtable	2016-02-26	225699			✓	✓				1	31
Hmisc	2017-05-02	124033	67		✗	✓	0.3	0.6	0	4	126
htmlTable	2017-01-26	80092	33	✓	✓		4.5	0.5	4		9
huxtable	2017-05-18	645	36	✓	✓		0.7	0.7	4		0
janitor	2017-05-06	1572	191	✓	✓		0.6	0.6	4	1	0
knitr	2017-05-18	237499	1365	✓	✗		0.3	0	87	1	94
kableExtra	2017-05-25	2149	71	✓	✓		0	0	2	1	0
pander	2015-11-23	9884	182		✗		2.9	0.4	16	1	26
pixiedust	2017-05-07	511	118	✓	✓		0.6	0.1	7	1	2
ReporteRs	2017-01-09	3739	205	✓	✓	✓	0	0	3	2	5
stargazer	2015-07-14	12178			✗	✓					2
tableone	2015-11-11	1272	33		✓		19.2	15.6	19		0
tables	2017-01-03	5051			✗					2	3
tangram	2017-05-03	137			✓					6	0
texreg	2017-03-03	4313	7		✗	✓	1.7	0.1	0	1	4
xtable	2016-02-05	150002			✗					1	97
ztable	2015-02-15	670			✓					1	1

JS

What this means for package developers

- Most R users are open to trying out new packages
- There are so many packages that it can be difficult to connect with *your* audience

What is a CRAN Task View?

Let's check out the Task Views right now

Let’s look at this Shiny app from Mikhail Popov

CRAN Task Views (CTV), the UNOFFICIAL VERSIONS

Sometimes package developers and users put together Task Views on their own

Check out Ben Marwick's archeology CTV or Thomas Leeper's open data CTV

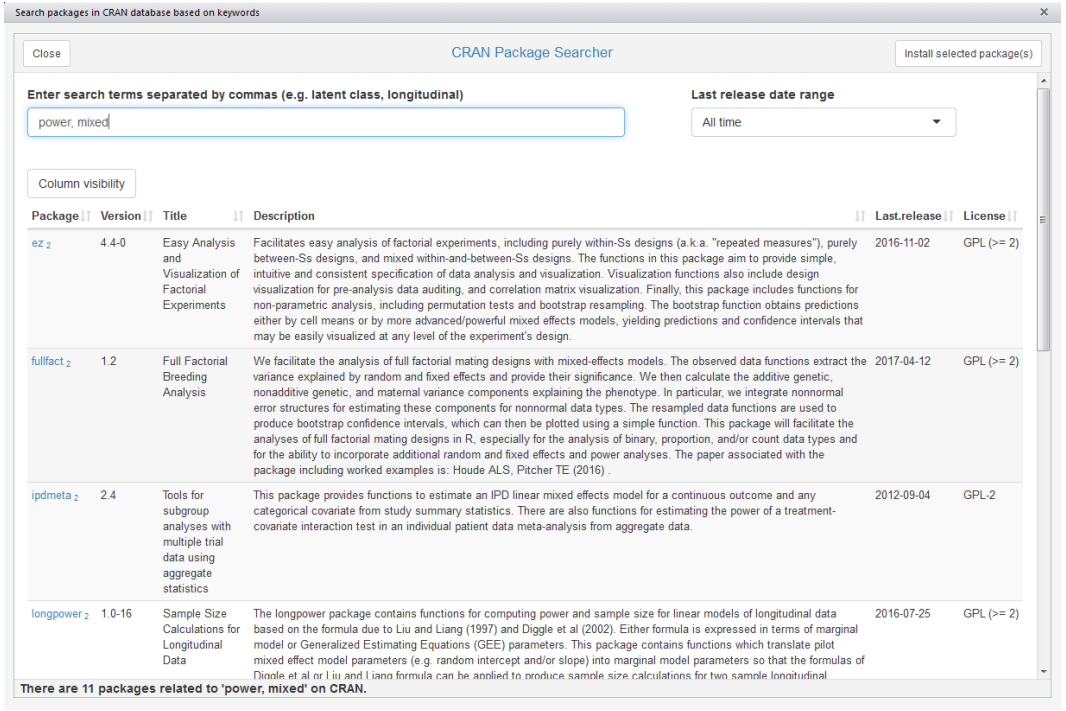
What this means for Task View maintainers

- Making (and keeping!) a Task View useful can be a challenge
- Task Views vary in how helpful and up-to-date they are
- Could more CTVs move to being maintained on GitHub or a Wiki?
- Two possible approaches for CTVs + GitHub
  - Editing a markdown file and using makefiles to get to XML
  - Editing XML and using a pretty simple script to get to markdown

How do people find p ackages now?

CRANsearcher

RStudio add-in to search CRAN packages based on keyword(s)



CRANsearcher

Install from CRAN: `install.packages("CRANsearcher")`

Functionality

- Search CRAN database based on keyword(s)
    - Searches the package name, title, and **description**
- Filter by most recent release date
- Link to websites to learn more
- Install selected package(s) with the click of a button

CRANsearcher

GitHub: https://github.com/RhoInc/CRANsearcher

Authors:

Agustin Calatroni, Rho, Inc.

Becca Krouse, Rho, Inc.

Are you an R User?

Are you a package developer?

sos::findFn

- Searches the “RSiteSearch” database for matches in help pages
- Sorts the results to put first the package with the most matches
- writeFindFn2xls to produce a package summary
    - required installing packages locally to get some of the information needed
    - Not well known

sos::findFn

- library(sos)
- findFn('your search term')
  - development version opens two web pages for
    - help pages
    - packages

findFn Results

SG

call: "x <- findFn(string = 'your search term')"

For a summary by package, see: "packageSum(x,...)"

See also: vignette('sos')

Id	Count	MaxScore	TotalScore	Package	Function	Date	Score	Description and Link
1	5	283	1224	taxize	get_boldid	2017-02-14 21:56:52	283	Get the BOLD (Barcode of Life) code for a search term.
2	5	283	1224	taxize	get_ubioid-defunct	2017-02-14 21:56:52	257	Get the uBio id for a search term
3	5	283	1224	taxize	get_tolid	2017-02-14 21:56:52	252	Get the OTT id for a search term
4	5	283	1224	taxize	get_tsn	2017-02-14 21:56:52	252	Get the TSN code for a search term.
5	5	283	1224	taxize	eol_search	2017-02-14 21:56:52	180	Search for terms in EOL database.
6	3	229	558	rsunlight	os_billsearch	2016-12-20 16:58:44	229	Search OpenStates bills.

Package Summary

SG

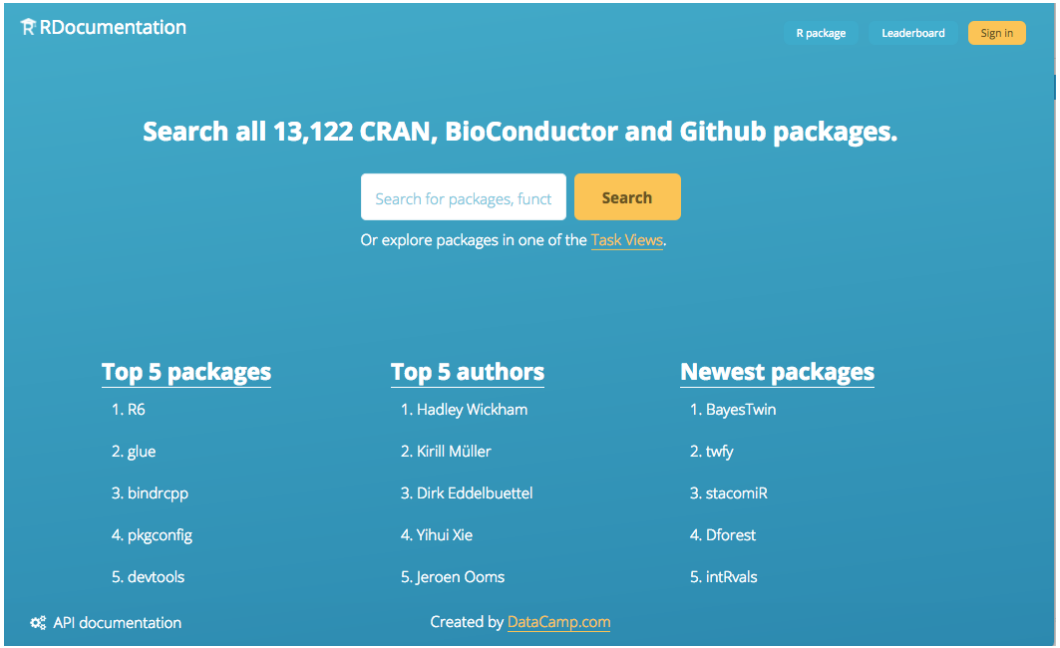
call: "x <- findFn(string = 'your search term')"

Title, etc., are available on installed packages. To get more, use "installPackages(x,...)"

See also: "writeFindFn2xls(x)"

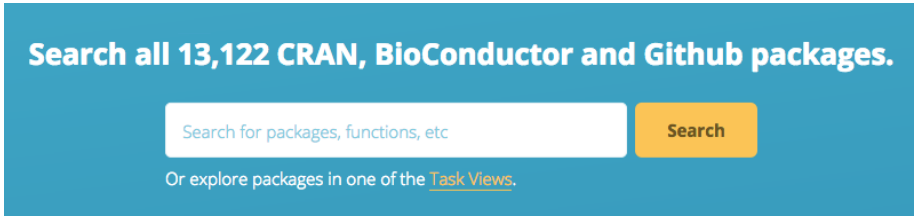
Id	Package	Count	MaxScore	TotalScore	Date	Title and Link	Version	Author	Maintainer	helpPages	vignette	URL
1	taxize	5	283	1224	2017-01-18 00:00:00	Taxonomic Information from Around the Web	0.8.4	Scott Chamberlain [aut, cre], Eduard Szoecs [aut], Zachary Foster [aut], Carl Boettiger [ctb], Karthik Ram [ctb], Ignasi Bartomeus [ctb], John Baumgartner [ctb], James O'Donnell [ctb], Jari Oksanen [ctb]	Scott Chamberlain	109	3: taxize_infotable, name_cleaning, taxize_vignette	https://github.com/ropensci/taxize
2	rsunlight	3	229	558	2016-12-20 16:58:44							
3	SocialMediaLab	3	49	96	2017-06-03 07:49:17							
4	rvertnet	2	326	359	2016-12-20 18:56:50							

Online search and discovery tool to navigate all packages on CRAN, BioConductor and GitHub



DataCamp

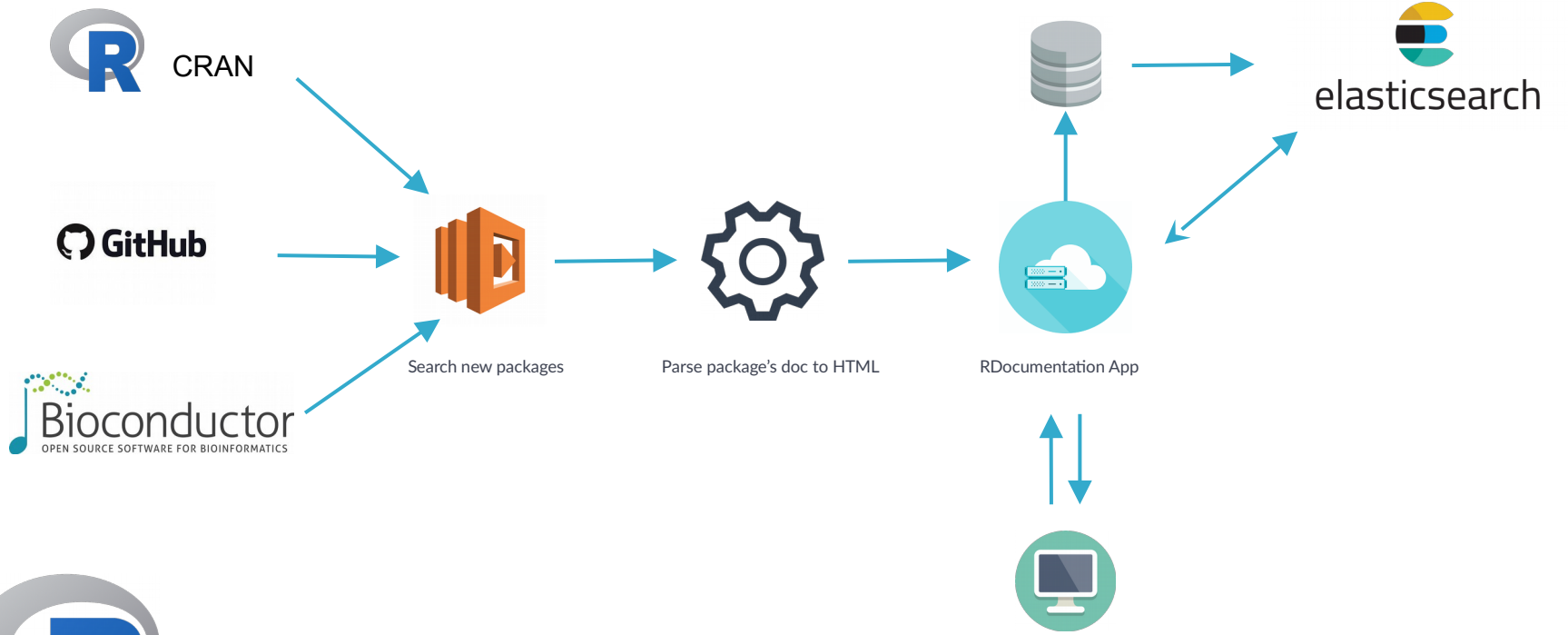
Why did we build this ?



- > 13,000 packages
- Make it easy to just **find** what you need
- Documentation is written by experts for experts
  - Provide a user-friendly, welcoming interface for R beginners
  - Community-driven documentation via examples
- Central documentation repository (CRAN/BioC/GitHub)
- Find older versions of packages

DataCamp

How does it work ?

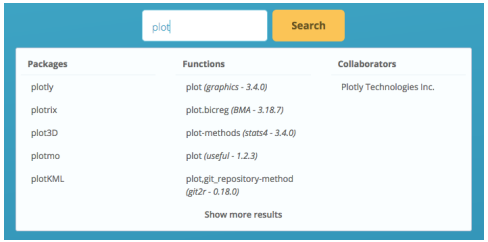


RDocumentation: What ?

DataCamp

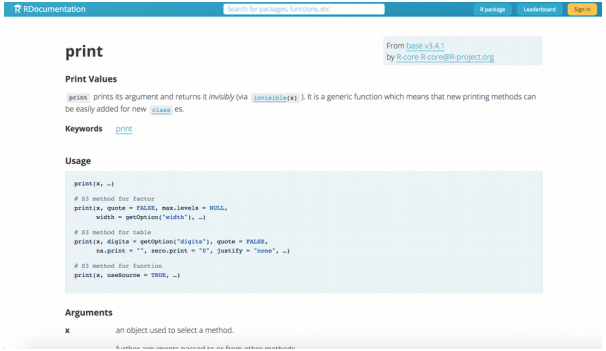
1. Search

- Fast
- Precise
- Relevant
  - Use download stats to assess package quality



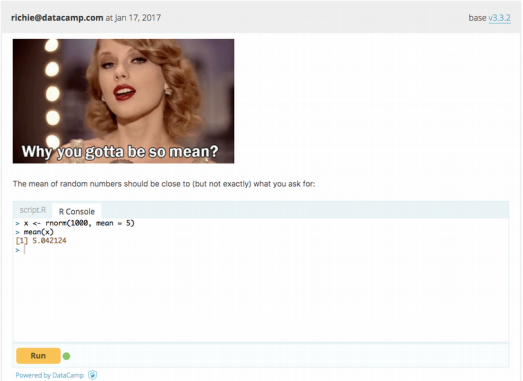
2. Content

- Up to date
- Easy to browse
- Visually clear
- Older versions are browsable



3. Community

- Community members can post high-quality, interactive examples.
- Leaderboard



Where is this going ?

- Browse source code
- RDocs Light
  - "Hover" widget to include minimal version of the doc on any website
- Increase community engagement
  - Some contributors post examples but that's not enough
  - Want would ***you*** want to see ?

RDocumentation is completely *open-source!*

Wanna help ? Feel free to contribute and post new issues/ideas on:

- https://github.com/datacamp/RDocumentation-app

Thank you

- Heather Turner
- Tobias Verbeke
- useR 2017 organizers

BREAKOUT SESSIONS