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](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/)
- [https://github.com/nashjc/Rnavpkg/wiki](https://github.com/nashjc/Rnavpkg/wiki)
Packages on CRAN

How many packages are available on CRAN?
Only packages that are still available

Number of available packages

0 2500 5000 7500 10000
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 different kinds of tasks
- It can be difficult to find what you are looking for
- Assessing quality can be a challenge
<table>
<thead>
<tr>
<th>package</th>
<th>published</th>
<th>dl_last_month</th>
<th>stars</th>
<th>tidyverse_happy</th>
<th>has_tests</th>
<th>vignette</th>
<th>last_commit</th>
<th>last_issue_closed</th>
<th>contributors</th>
<th>depends_count</th>
<th>reverse_count</th>
</tr>
</thead>
<tbody>
<tr>
<td>arsenal</td>
<td>2017-03-10</td>
<td>236</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ascii</td>
<td>2011-09-29</td>
<td>1806</td>
<td>19</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>compareGroups</td>
<td>2017-03-14</td>
<td>929</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>conformat</td>
<td>2017-06-18</td>
<td>1252</td>
<td>4</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>descetable</td>
<td>2017-05-15</td>
<td>494</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>DT</td>
<td>2016-08-09</td>
<td>352</td>
<td>206</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>31</td>
</tr>
<tr>
<td>expss</td>
<td>2017-04-09</td>
<td>222</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>ezsummary</td>
<td>2016-07-11</td>
<td>168</td>
<td>30</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.7</td>
<td>0</td>
</tr>
<tr>
<td>formatable</td>
<td>2016-08-05</td>
<td>2753</td>
<td>299</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>gmodels</td>
<td>2015-07-22</td>
<td>2966</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>gtable</td>
<td>2016-02-26</td>
<td>22569</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>Hmisc</td>
<td>2017-05-02</td>
<td>124933</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>126</td>
</tr>
<tr>
<td>htmlTable</td>
<td>2017-01-26</td>
<td>81902</td>
<td>33</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>huxtable</td>
<td>2017-05-16</td>
<td>645</td>
<td>36</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>janitor</td>
<td>2017-05-06</td>
<td>1572</td>
<td>191</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>knitr</td>
<td>2017-05-18</td>
<td>1237499</td>
<td>1365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>67</td>
<td>1</td>
</tr>
<tr>
<td>kableExtra</td>
<td>2017-06-25</td>
<td>2148</td>
<td>71</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>pander</td>
<td>2015-11-23</td>
<td>9884</td>
<td>182</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>pixiedust</td>
<td>2017-05-07</td>
<td>511</td>
<td>118</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>ReporteRs</td>
<td>2017-01-09</td>
<td>3738</td>
<td>205</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>stargazer</td>
<td>2015-07-14</td>
<td>12178</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>tableone</td>
<td>2015-11-11</td>
<td>1272</td>
<td>33</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>tables</td>
<td>2017-01-03</td>
<td>5051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>tangram</td>
<td>2017-05-03</td>
<td>137</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>texreg</td>
<td>2017-03-03</td>
<td>4313</td>
<td>7</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>xtable</td>
<td>2016-02-05</td>
<td>150002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>97</td>
</tr>
<tr>
<td>ztable</td>
<td>2015-02-15</td>
<td>670</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
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 packages now?
# R Package Discovery

How do you currently discover and learn about R packages? 

<table>
<thead>
<tr>
<th>Method</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email lists such as r-help, r-packages, or r-pkg-devel</td>
<td>15%</td>
</tr>
<tr>
<td>General search websites such as Google and Yahoo</td>
<td>57%</td>
</tr>
<tr>
<td>R-specific search websites such as METACRAN (<a href="http://www.r-pkg.org">www.r-pkg.org</a>) or Rdocumentation (<a href="https://www.rdocumentation.org/">https://www.rdocumentation.org/</a>)</td>
<td>11%</td>
</tr>
<tr>
<td>R packages built for search such as the sos package</td>
<td>2%</td>
</tr>
<tr>
<td>CRAN Task Views</td>
<td>22%</td>
</tr>
<tr>
<td>Your personal network, such as colleagues and professors</td>
<td>41%</td>
</tr>
<tr>
<td>Conferences, meet-ups, or seminars</td>
<td>24%</td>
</tr>
<tr>
<td>Books, textbooks, or journal articles (JSS, JOSS, R-Journal)</td>
<td>32%</td>
</tr>
<tr>
<td>Social media such as blogs, R-bloggers, Twitter, Slack, or GitHub contacts</td>
<td>79%</td>
</tr>
<tr>
<td>Other (send ideas to @julaslige on Twitter!)</td>
<td>4%</td>
</tr>
</tbody>
</table>
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**

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

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

See also: vignette('sos')

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

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

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

**See also**: "writeFindFn2xls(x)"

<table>
<thead>
<tr>
<th>Id</th>
<th>Package</th>
<th>Count</th>
<th>MaxScore</th>
<th>TotalScore</th>
<th>Date</th>
<th>Title and Link</th>
<th>Version</th>
<th>Author</th>
<th>Maintainer</th>
<th>helpPages</th>
<th>vignette</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>taxize</td>
<td>5</td>
<td>283</td>
<td>1224</td>
<td>2017-01-18 00:00:00</td>
<td>Taxonomic Information from Around the Web</td>
<td>0.8.4</td>
<td>Scott Chamberlain [aut, cre], Eduard Szoeck [aut], Zachary Foster [aut], Carl Boettiger [ctb], Karthik Ram [ctb], Ignasi Bartomeus [ctb], John Baumgartner [ctb], James O'Donnell [ctb], Jari Oksanen [ctb]</td>
<td>Scott Chamberlain</td>
<td>109</td>
<td>3: taxize_infotable, name_cleaning, taxize_vignette</td>
<td><a href="https://github.com/ropensci/taxize">https://github.com/ropensci/taxize</a></td>
</tr>
<tr>
<td>2</td>
<td>rsunlight</td>
<td>3</td>
<td>229</td>
<td>558</td>
<td>2016-12-20 16:58:44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SocialMediaLab</td>
<td>3</td>
<td>49</td>
<td>96</td>
<td>2017-06-03 07:49:17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>rvertnet</td>
<td>2</td>
<td>326</td>
<td>359</td>
<td>2016-12-20 18:56:50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Online search and discovery tool to navigate all packages on CRAN, BioConductor and GitHub
RDocumentation: 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
RDocumentation: How does it work?

Navigating R Packages | useR! Brussels
RDocumentation: What?

1. Search
   - Fast
   - Precise
   - Relevant

2. Content
   - Up to date
   - Easy to browse
   - Exhaustive
   - Older versions are browsable

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

Navigating R Packages | useR! Brussels
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
   ◆ What 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
BREAKOUT SESSIONS

- Unification: John Nash in Plenary
- Search: Spencer Graves in 3.02
- Guidance: Julia Silge in 2.02

https://github.com/nashjc/Rnavpkg