

# Navigating the R package universe

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

Spencer Graves

Julia Silge

Ludovic Vannoorenberghe



# Background

- Initial name "Navigating the R package jungle"
- Jungles -- rain forests -- are places rich in resources.
  - more than 10000 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)
  - Rdocumentation (LV)
- **Search** -- improving how users can find the tools they need and information on how to use them effectively and efficiently (SG)
- ***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 simply different names for same functionality



# Unifying packages (2)

- 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
- **jmv** (Jamovi) (Jonathan Love) attempts to integrate many common statistical tests
- **bbmle** (Ben Bolker) some integration of tools for maximum likelihood estimation
- Have I missed good examples? Let me know! (nashjc\_at\_uottawa.ca)



# Opportunities for unification?

- Principal Components / svd -- (JN and Claudia Beleites) <https://gitlab.com/nashjc/svdpls>
- nonlinear modeling -- better integration of `nls()`, packages `**nlshr**`, `**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/> and <https://github.com/nashjc/Rnavpkg/wiki>

