



OPEN SOURCE

LEADERSHIP SUMMIT

FEBRUARY 14-16, 2017 | LAKE TAHOE, CA



OPEN SOURCE
LEADERSHIP SUMMIT
FEBRUARY 14-16, 2017 | LAKE TAHOE, CA

R Consortium

First Year Update

David Smith, *R Consortium / Microsoft*



What is R?

Most widely used data analysis software

- Used by 2M+ data scientists, statisticians and analysts

Most powerful statistical programming language

- Flexible, extensible and comprehensive for productivity

Create beautiful and unique data visualizations

- As seen in New York Times, The Economist and FlowingData

Fills the Data Science talent gap

- New graduates prefer R

Thriving open-source community

- Leading edge of analytics research



A brief history of R

1993 Research project in Auckland, NZ

- Ross Ihaka and Robert Gentleman

1995 Released as open-source software

- Generally compatible with the "S" language

1997 R core group formed

2003 R Foundation formed in Austria

2007 Revolution Analytics founded

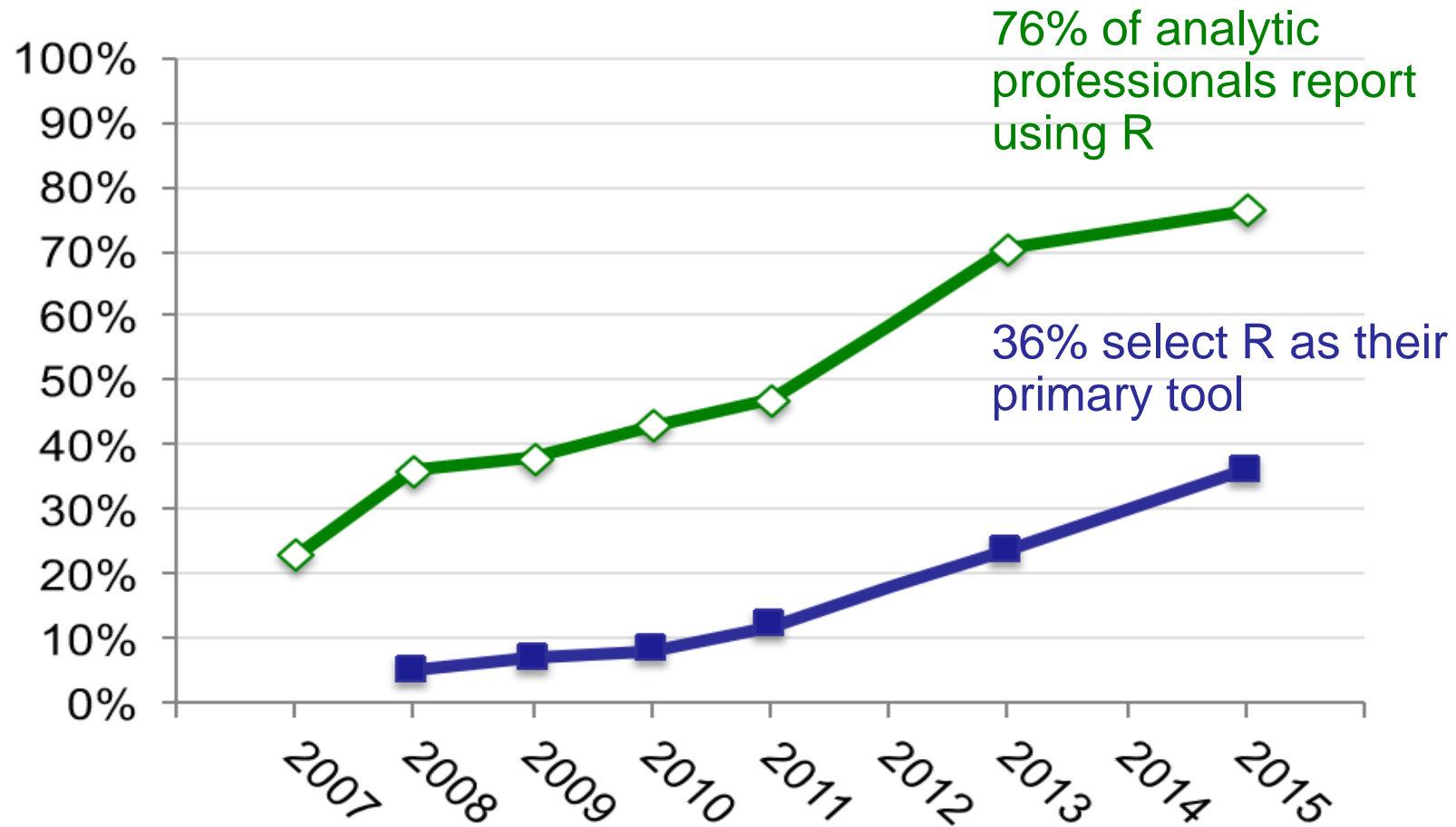
... Followed by Oracle, Rstudio, HP, IBM ,
Microsoft ...



Photo credit: Robert Gentleman

R Usage Growth

Rexer Data Miner Survey, 2007-2015



Technical Capability

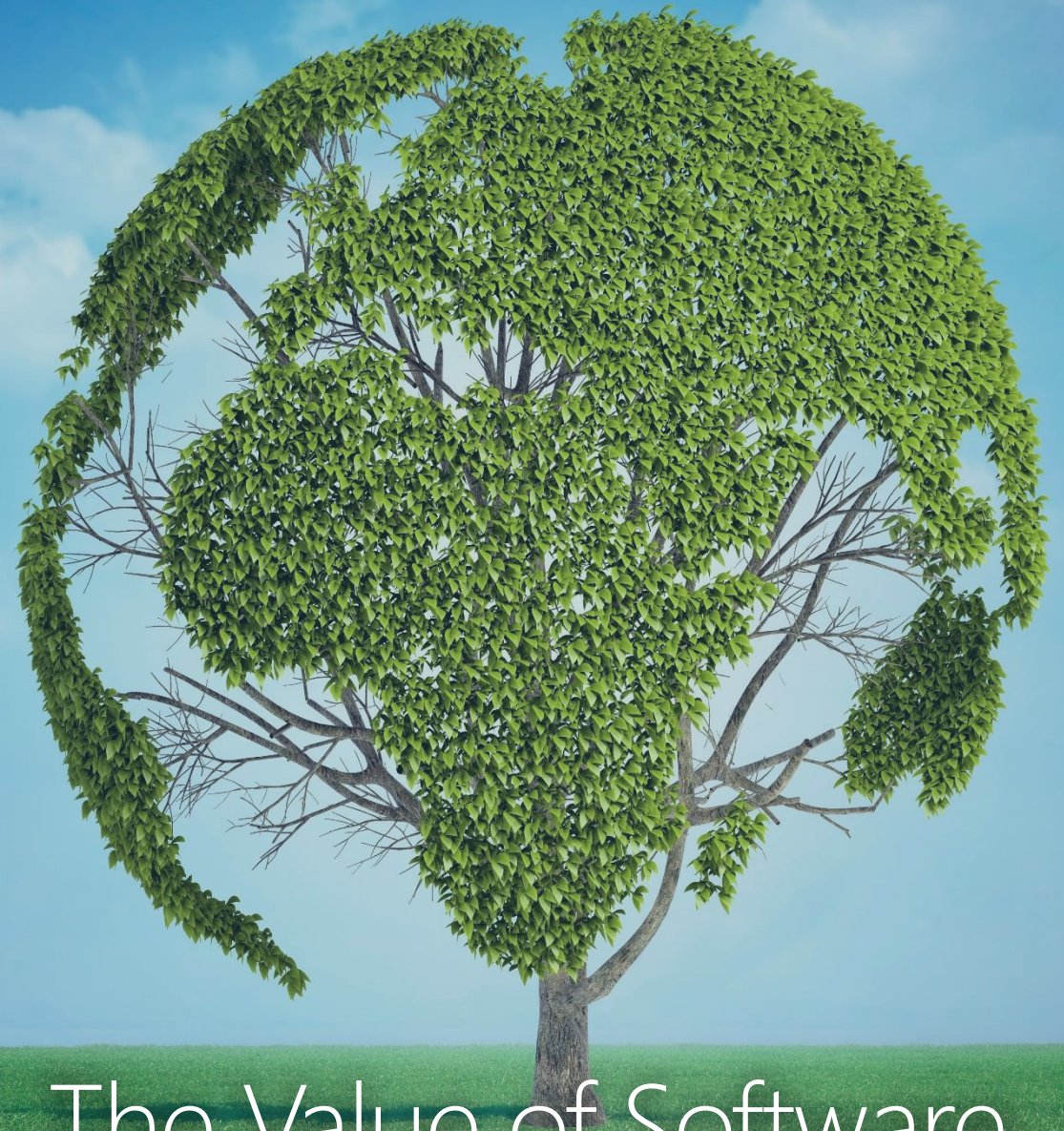
Governance

Vendor Support

Community Resources

Talent Pool

Legal Clarity



The Value of Software
lies in its ecosystem

Why a Linux Foundation Project?

- Easy to get started
- Experience
- Staff & Infrastructure
- International recognition
- Trust
- 501(c)(6)
- Overhead



A non-profit trade organization supporting the R Community.
Founded June 2015.



Goals

- Promote growth and development of R as leading platform for data science
- Support the R Foundation
- Fund projects to support R and community
- Foster growth of R community
- Enable use of R in commercial environments
- Foster collaboration between companies

Structure

Board of Directors

- Governance, budget, etc.

Infrastructure Steering Committee

- Grant review / awards
- Working groups

R Foundation member

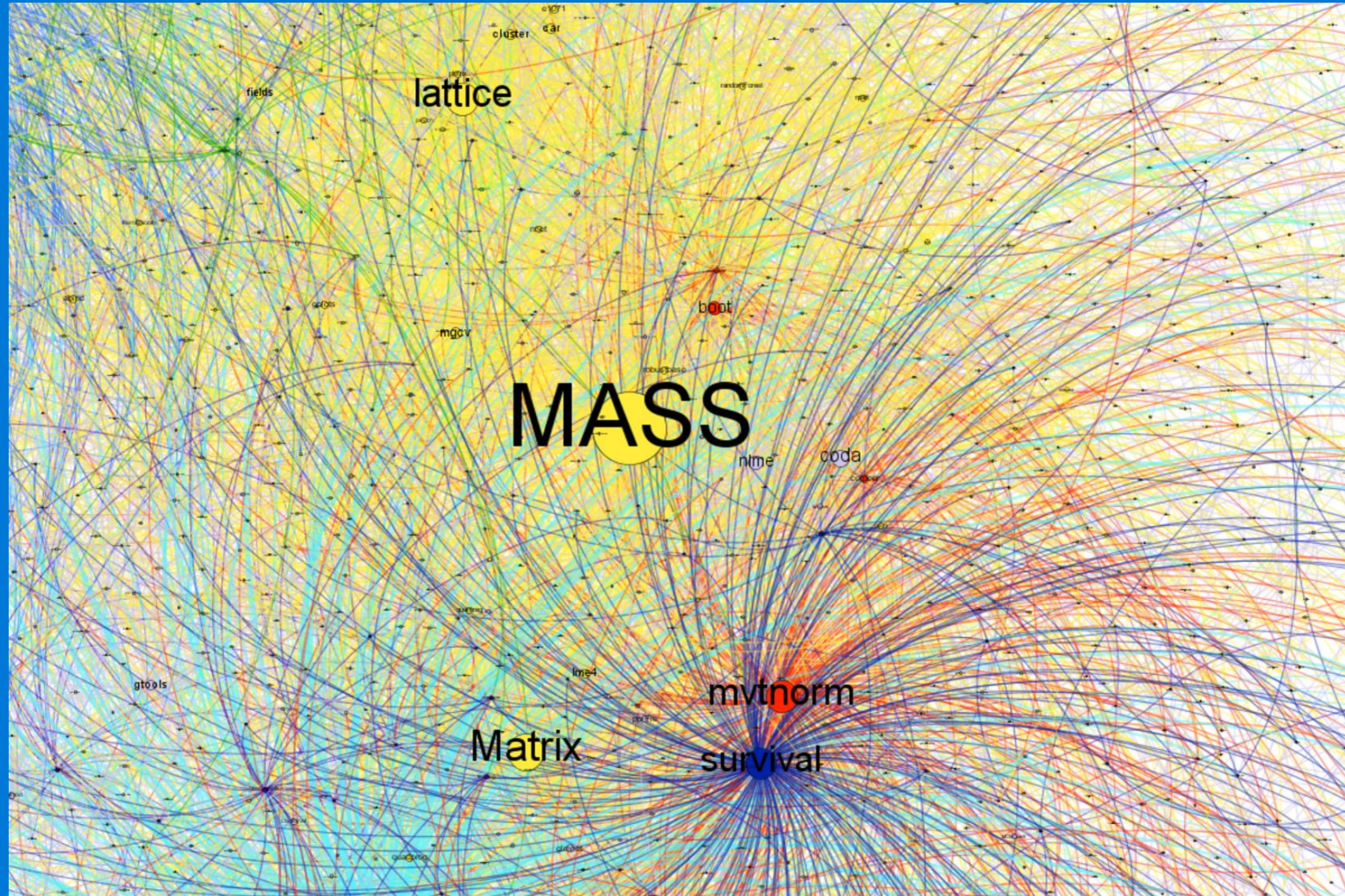
Platinum Member (\$100k)

Gold Member (\$50k)

Silver Member (\$25k / \$10k)

ISC Projects

- Infrastructure
- Education
- Documentation
- Production use of R
- Package ecosystem
- Grants fund developers



ISC Working Groups

What they are:

- Projects for exploring new technology
- Forums for achieving consensus
- The mechanism for organizing and executing large collaborative projects

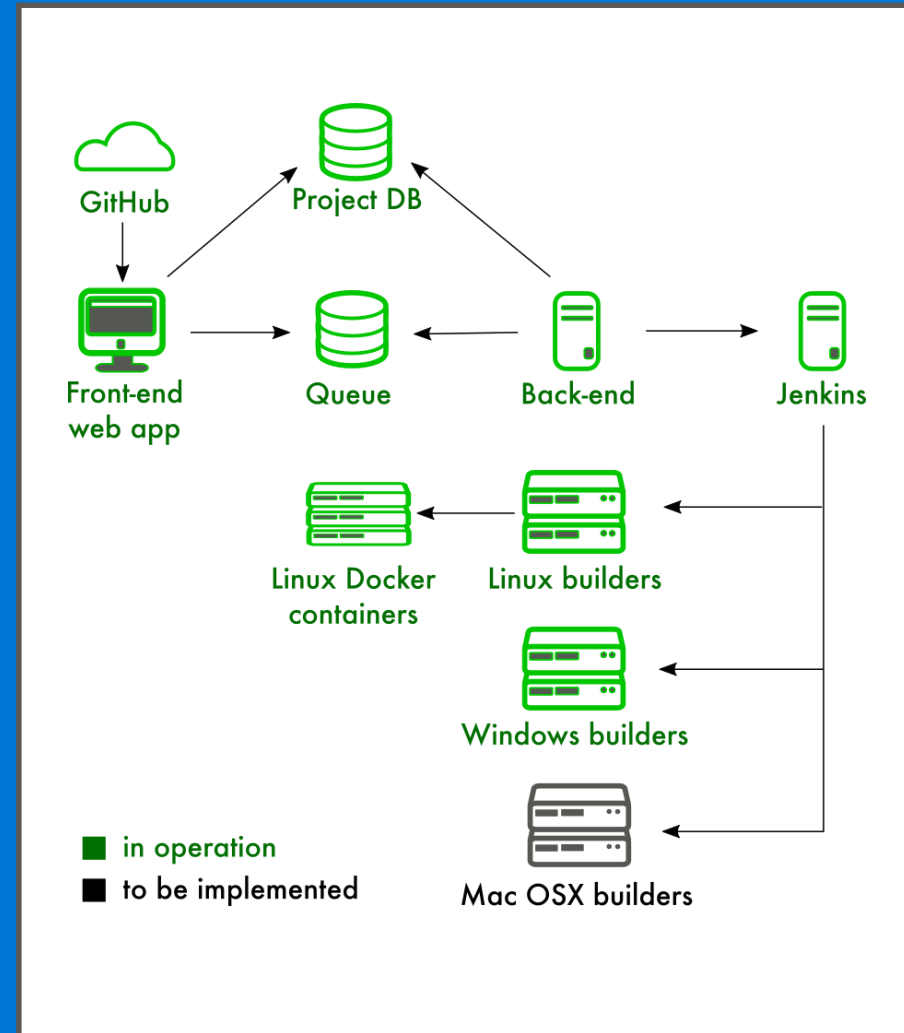
Benefits:

- Sponsored by the R Consortium
- Receive attention from the R Foundation
- Visible to the greater R Community
- Receive administrative support from the R Consortium

Infrastructure Projects

R-hub: R package building service

- Services that ease all steps the R package development process.
- Make these services free for all members of the community.
- Allow community contributions to R-hub itself.
- Make CRAN maintainers' work easier by pre-testing CRAN package submissions.
- \$85,000, Jan 2016 — Feb 2017



R Localization Proposal (RL10N):

- Richie Cotton (Weill Cornell Medicine in Qatar) and Thomas Leeper (The London School of Economics)
- Majority of R packages in English only
- RL10N project will make it easier for R developers to include translations in their own packages
- Plan:
 - Improve msgtools package
 - New package to adapt MTurkR for managing translation
 - New package to adapt translateR for automated translations

Code Coverage Tool for R

- Develop a tool for R that determines code coverage upon execution of a test suite
- Improve software quality
- Promoting the use of code coverage more systematically within the R ecosystem

https://wiki.r-consortium.org/view/Code_Coverage_Tool_for_R

Working Group Members

- Shivank Agrawal, Oracle
- Chris Campbell, Mango Solutions
- Santosh Chaudhari, Oracle
- Karl Forner, Quartz Bio
- Jim Hester, RStudio
- Mark Hornick, Oracle – Group Leader
- Chen Liang, Oracle
- Willem Ligtenberg, Open Analytics
- Andy Nicholls, Mango Solutions
- Vlad Sharanhovich, Oracle
- Tobias Verbeke, Open Analytics
- Qin Wang, Oracle
- Hadley Wickham, RStudio – ISC Sponsor

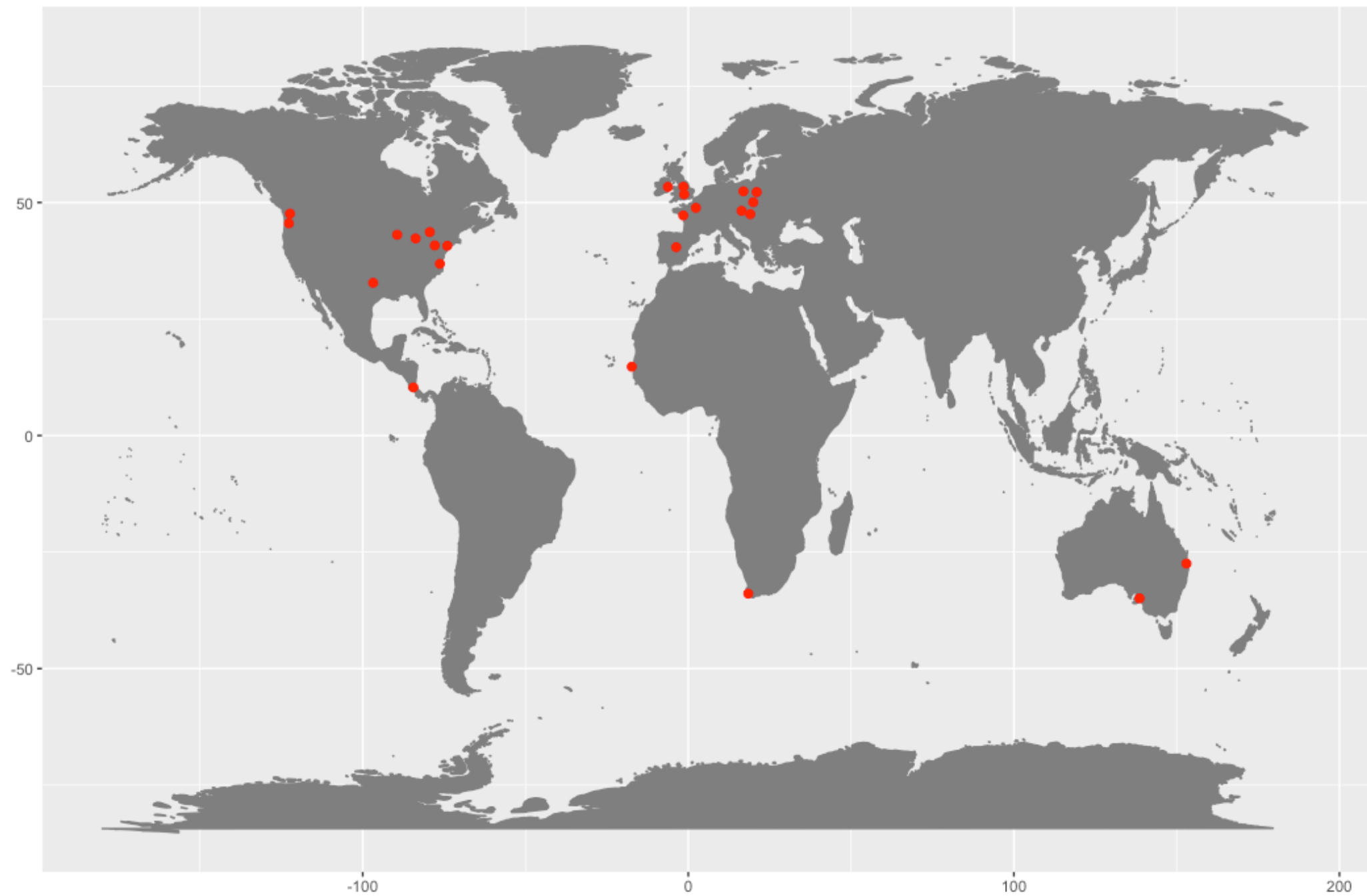
Improving Database Interface (DBI):

- Kirill Müller (ETH Zürich)
- Improve database access in R so that porting code is simplified and less prone to error
- Plan:
 - Create a DBI specification, centralized test and boiler plate for DBI backends
 - Improve existing DBI backends to adhere to the standard
 - Focus on RMySQL, RPostgres and RSQLite
- <https://github.com/rstats-db/DBItest>

Community Projects



R Consortium Supported RUGs 1-26-17





.. to over **35** R-Ladies Chapters!!

~4000

R-Ladies Meetup
Members

~6470

Twitter Followers

~100

Events



Lessons and Outcomes

Challenges

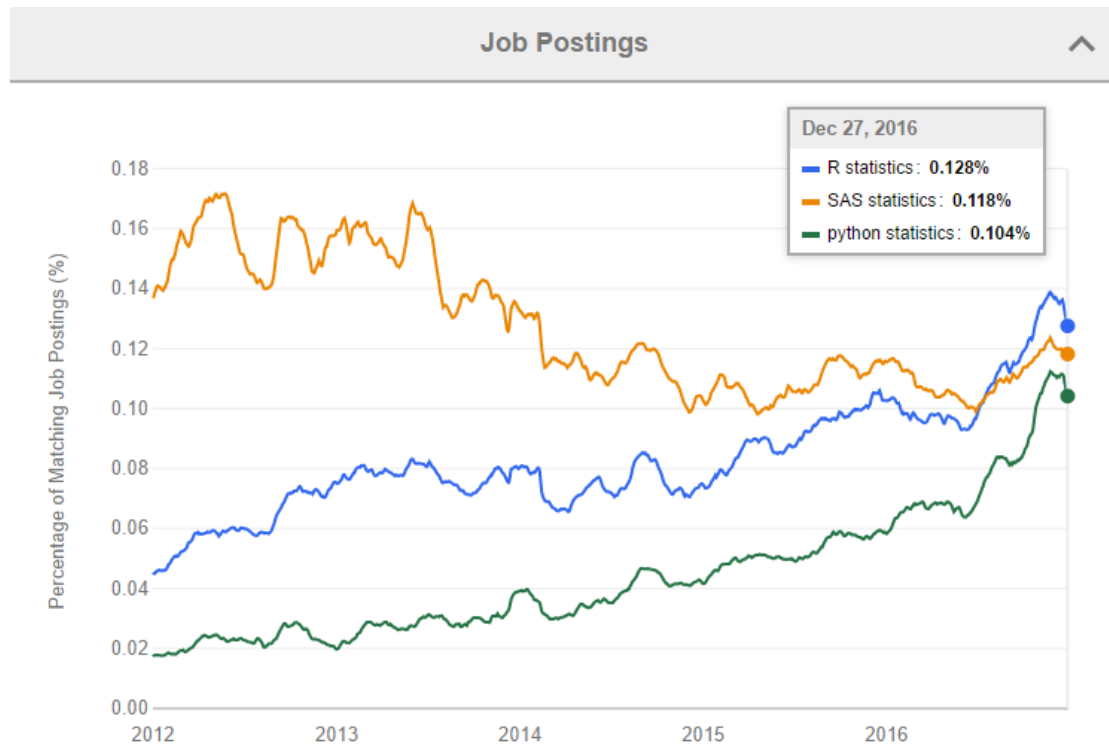
- Relationship between R Foundation and R Consortium
- Value proposition for members
- Disbursement of funds

Results

- 13 corporate members (4 gold)
- 25 funded projects / working groups
- Close collaboration with R Foundation
- Community approval and participation
- A secure, long-term foundation for R






















R: The #1 software for Data Science

... and #5 amongst general-purpose programming languages



Language Popularity

IEEE Spectrum Top Programming Languages, 2016

Language Rank	Types	Spectrum Ranking
1. C	  	100.0
2. Java	  	98.1
3. Python	 	98.0
4. C++	  	95.9
5. R		87.9
6. C#	  	86.7
7. PHP		82.8
8. JavaScript	 	82.2
9. Ruby	 	74.5
10. Go		71.9



Thank You

David M Smith

Director, R Consortium

R Community Lead, Microsoft

davidsmi@microsoft.com

@revodavid



OPEN SOURCE LEADERSHIP SUMMIT

FEBRUARY 14-16, 2017 | LAKE TAHOE, CA