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 Gentlemen

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

76% of analytic professionals report using R

36% select R as their primary tool
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.

R Foundation member

Platinum Member ($100k)

Infrastructure Steering Committee
- Grant review / awards
- Working groups

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 Quatar) 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


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 Sharanovich, 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](https://github.com/rstats-db/DBItest)
Community Projects
.. 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

<table>
<thead>
<tr>
<th>Language Rank</th>
<th>Types</th>
<th>Spectrum Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. C</td>
<td>📊💻📱</td>
<td>100.0</td>
</tr>
<tr>
<td>2. Java</td>
<td>📊📱</td>
<td>98.1</td>
</tr>
<tr>
<td>3. Python</td>
<td>📊💻</td>
<td>98.0</td>
</tr>
<tr>
<td>4. C++</td>
<td>📊💻📱</td>
<td>95.9</td>
</tr>
<tr>
<td>5. R</td>
<td>📊💻</td>
<td>87.9</td>
</tr>
<tr>
<td>6. C#</td>
<td>📊📱</td>
<td>86.7</td>
</tr>
<tr>
<td>7. PHP</td>
<td>📊📱</td>
<td>82.8</td>
</tr>
<tr>
<td>8. JavaScript</td>
<td>📊📱</td>
<td>82.2</td>
</tr>
<tr>
<td>9. Ruby</td>
<td>📊💻</td>
<td>74.5</td>
</tr>
<tr>
<td>10. Go</td>
<td>📊💻</td>
<td>71.9</td>
</tr>
</tbody>
</table>
Thank You

David M Smith
Director, R Consortium
R Community Lead, Microsoft
davidsmi@microsoft.com
@revodavid