

R USER CONFERENCE IN KOREA 2019 GO TO COMMUNITY!

📅 10월 25일 (금) 📍 한국마이크로소프트(더케이타워 11층)

한국R사용자커뮤니티(R Korea)

The R Consortium, Community and Ecosystem

Joseph B. Rickert

10/2/2019

What this talk is about

- R: The Big Picture
- The R Ecosystem
- The R Consortium
 - What it does
 - How it works
 - Some Accomplishments
 - Why Join?
 - How everyone can get involved



R is more than a language!

- The world's largest open repository for statistical knowledge
- Contains reference implementations for almost every important statistical algorithm
- The dominant language for statistical computing
- A gateway to other computational resources
- **A vibrant community with a stable ecosystem**

Community means people who:

1. self-identify
2. interact frequently
3. contribute and recognize the contributions of others

R's Trajectory

1976: John Chambers, Bell Labs begins work on S

1993: Gentleman & Ihaka, U of Auckland begins work on R

1995: R - open-source

1997: R Core Group formed

CRAN starts: 12 pkgs 2000: R 1.0.0 released

2001: Bioconductor Project

2003: R Foundation formed

2004: First useR! conf

2009: NY Times article on R

2015: The R Consortium

2016: Tidyverse

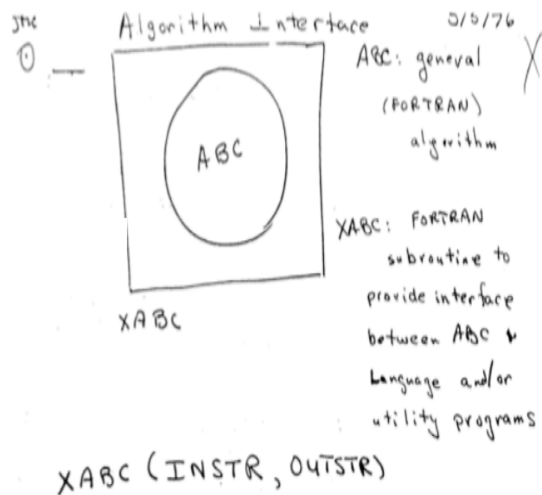
2018: R/Pharma

2018: R/Medicine

2019: CRAN - 15K pkgs



S was conceived as an interface



John Chambers' famous diagram from May 1976 indicates the intention to design a software

interface to call an arbitrary FORTRAN subroutine, ABC, by wrapping it in some simplified calling syntax: XABC().

"It was always understood that R is meant to build on a base of computational tools. R relies on the ability of functions to communicate with, and exchange objects with other software."

John Chambers, Extending R, CRC Press (2016)

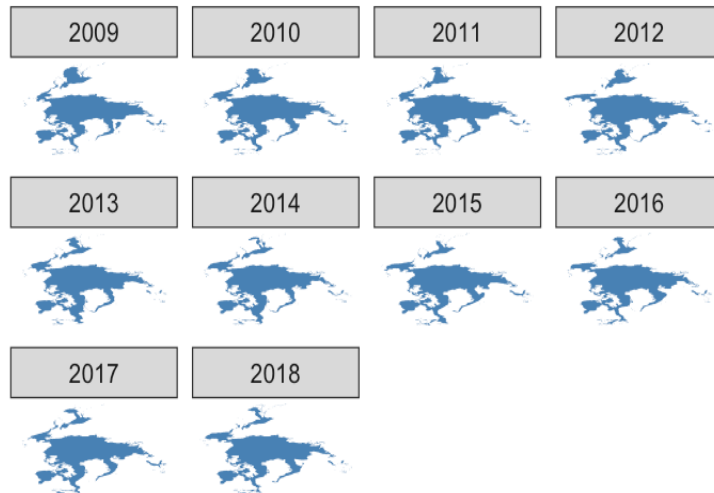
Characteristics of the R Language

- R is an interpreted scripting language
- Base R has a relatively small footprint
- The majority of growth and innovation comes from contributed packages
- **Features such as non-standard evaluation make it a good choice for “Design Specific Languages”**
- R is a functional, object based language
 - Everything that exists in R is an object
 - Everything that happens in R is a function call
 - Interfaces to other software are part of R

Collaboration: The Ordinary Magic of R

```
urls <- seaiceurls(mo = 'Apr', pole = 'N')[1:10]
out <- lapply(urls, seaice)
names(out) <- seq(2009, 2018, 1)
df <- ldply(out)

ggplot(df, aes(long, lat, group = group)) +
  geom_polygon(fill = "steelblue") +
  theme_ice() +
  facet_wrap(~.id)
```



rnoaa: 'NOAA' Weather Data from R

Client for many 'NOAA' data sources including the 'NCDC' climate 'API' at <https://www.ncdc.noaa.gov/cdo-web/webservices/v2>, with functions for each of the 'API' endpoints: data, data categories, data sets, data types, locations, location categories, and stations. In addition, we have an interface for 'NOAA' sea ice data, the 'NOAA' severe weather inventory, 'NOAA' Historical Observing 'Metadata' Repository ('HOMR') data, 'NOAA' storm data via 'IBTrACS', tornado data via the 'NOAA' storm prediction center, and more.

Version: 0.9.0

Imports: utils, [crul](#) (≥ 0.7.0), [lubridate](#), [dplyr](#), [tidyr](#), [tidyselect](#), [ggplot2](#), [scales](#), [XML](#), [xml2](#), [jsonlite](#), [rappdirs](#), [gridExtra](#), [tibble](#), [isdparsr](#) (≥ 0.2.0), [geonames](#), [hoardr](#) (≥ 0.5.2)

Suggests: [roxygen2](#) (≥ 6.1.1), [testthat](#), [knitr](#), [taxize](#), [ncdf4](#), [raster](#), [leaflet](#), [rgdal](#), [sf](#), [rmarkdown](#), [purrr](#), [ggmap](#), [ropenq](#), [vcr](#) (≥ 0.2.2)

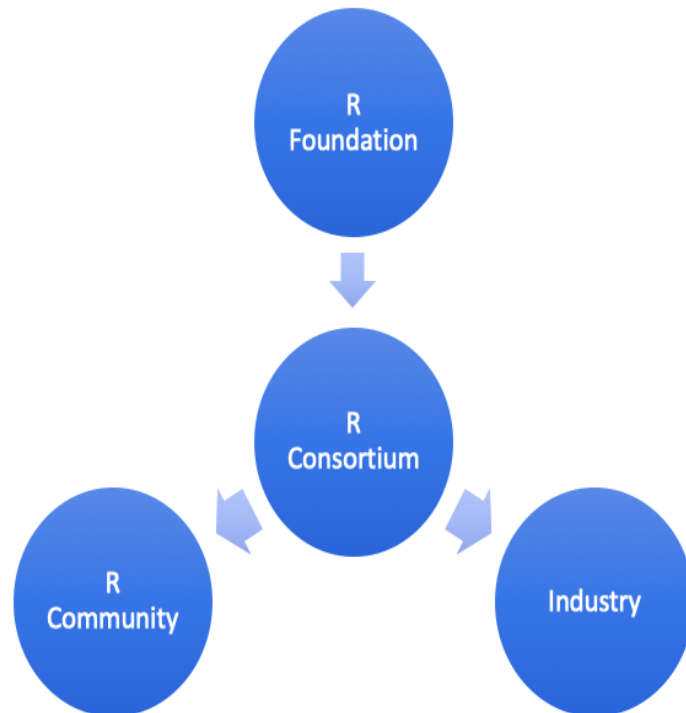
Published: 2019-09-26

Author: Scott Chamberlain [aut, cre], Brooke Anderson [ctb], Maëlle Salmon [ctb], Adam Erickson [ctb], Nicholas Potter [ctb], Joseph Stachelek [ctb], Alex Simmons [ctb], Karthik Ram [ctb], Hart Edmund [ctb]

The R Ecosystem

- The [R Language](#)
- The [R Foundation](#) (Austrian Foundation that controls R)
- The [R Core Group](#) (~ 20 R Foundation members that make commits to R)
- [CRAN](#): the central repository
- [Bioconductor](#): Project for high-throughput genomic analysis
- The R Community: Developers, users, [RUGS](#), [R-Ladies](#), conferences etc.
- The [R Consortium](#)

R Consortium



Non Profit Corp

- Founded in 2015
- Under the Linux Foundation
- Governed by:
- Board of Directors
- Technical committee (ISC)

Mission

- Support the R Foundation
- Promote the R language
- Develop and fund Infrastructure

R Consortium Members

R Foundation Member



Platinum Members



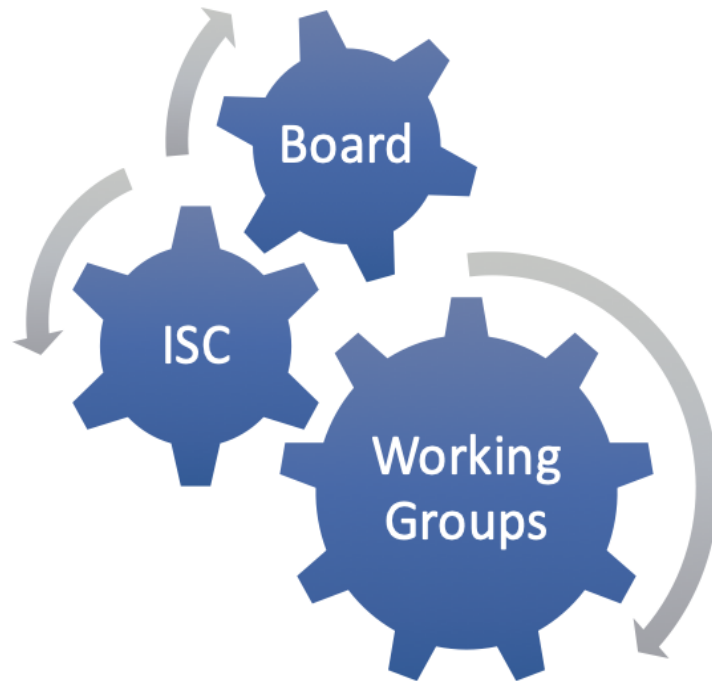
Gold Members



Silver Members



R Consortium Mechanics



- Board of Directors
 - Marketing Committee
- Infrastructure Steering Committee
 - Top Level Projects
 - Funded Projects
 - Working Groups

R Consortium Activities



Event Sponsorship 2019: \$60K



Top Level Project: R-Hub



Documentation

What is R-hub?
What is this website?
R-hub docs' privacy policy
Package builder
How to use?
R-hub webform or rhub package?
What R-hub platform(s)?
Email validation, why and how?
What if you are not the package maintainer?
R-hub vs CRAN platforms
How to build R packages for distributing them?
Secret API key, token, etc.?
How are dependencies installed?
System requirements
How to reproduce a bug uncovered by R-hub?
Local debugging with Docker
CRAN source code mirror

| What is R-hub?

R-hub is a collection of services to help R package development, free for all members of the community thanks to support from the [R Consortium](#).

Its most prominent service is the [package builder](#) that allows building and checking any R package on several platforms and R versions, with submission happening via a webform or an R client.

What R-hub is not

- A replacement of CRAN

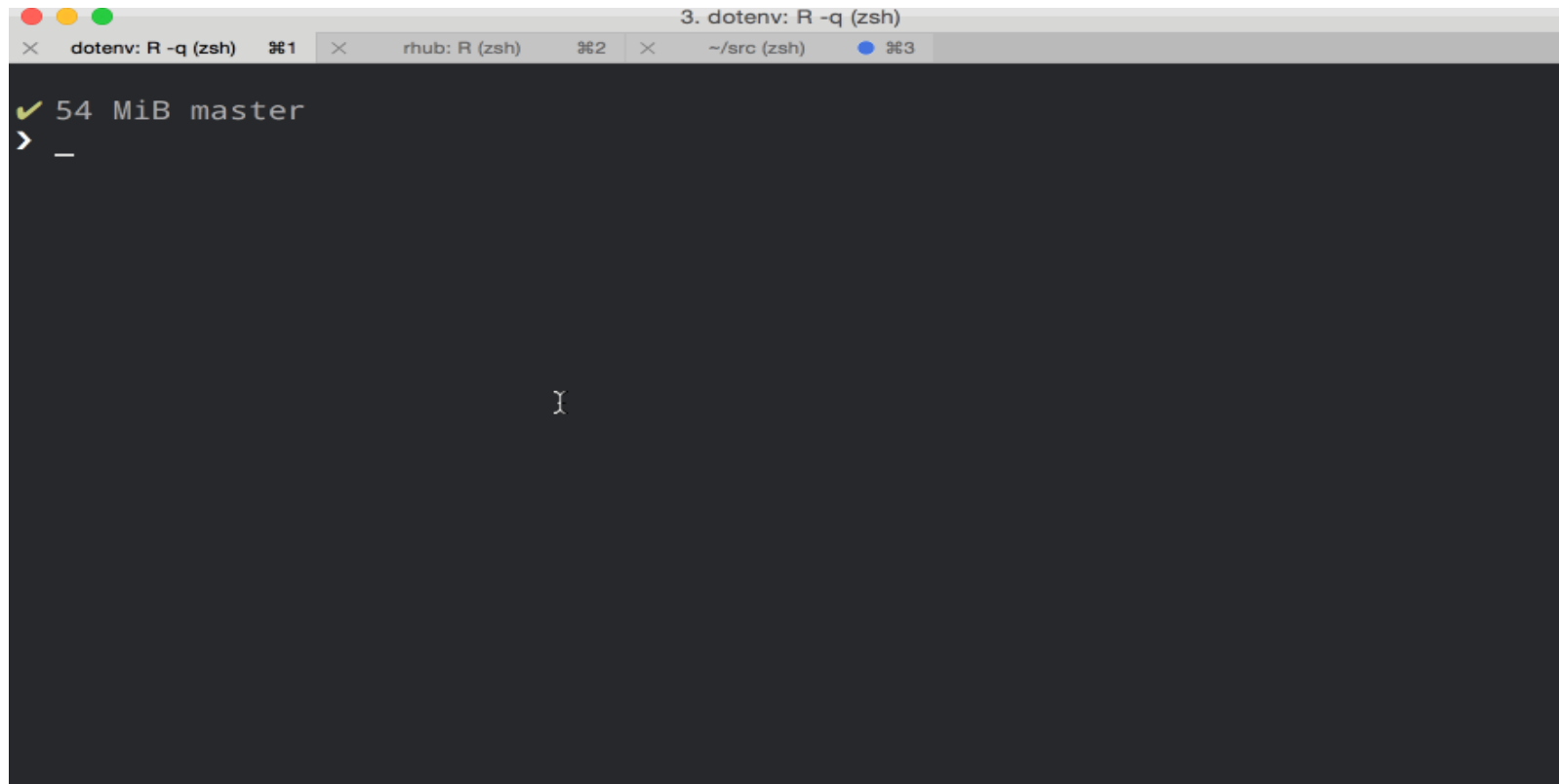
| What is this website?

You are now reading R-hub documentation website, currently a work in progress. The website aims at documenting all R-hub services, of which the table of content on the left gives a good overview.

You can find the source of this website [on GitHub](#) where [your feedback is welcome](#). In particular, do get in touch if you find something unclear, or wish a topic were added. By contributing to this project you agree to [its code of conduct](#). Thank you!

Except where otherwise noted, content on this site is licensed under the [CC-BY license](#).

Using R-Hub



A terminal window with a dark background and light gray text. The window title bar shows three colored buttons (red, yellow, green) and the text "3. dotenv: R -q (zsh)". Below the title bar, there are three tabs: "dotenv: R -q (zsh) 1", "rhub: R (zsh) 2", and "~/src (zsh) 3". The main content of the terminal shows a yellow checkmark followed by "54 MiB master", a prompt ">", and a cursor positioned after a hyphen "_".

```
3. dotenv: R -q (zsh)
x dotenv: R -q (zsh) 1 x rhub: R (zsh) 2 x ~/src (zsh) 3
✓ 54 MiB master
> _
```

Top Level Project: RUGS

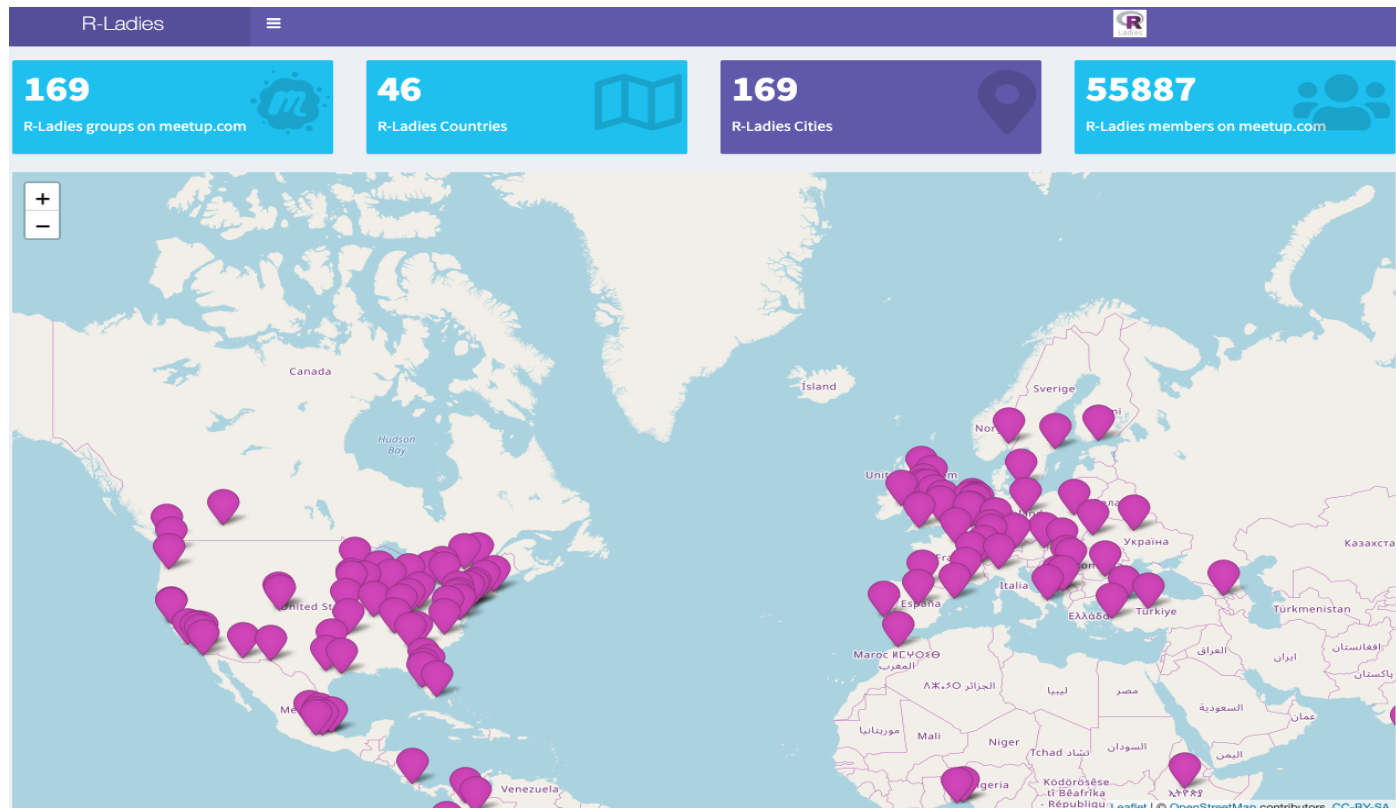
R User Group and Small Conference Support Program



Funds R user groups worldwide and makes conference grants for up to \$1K. So far in 2019:

- 23 Conference grants awarded
- 86 User groups funded in 32 countries
- 50,000 members

Top Level Project: R-Ladies Global



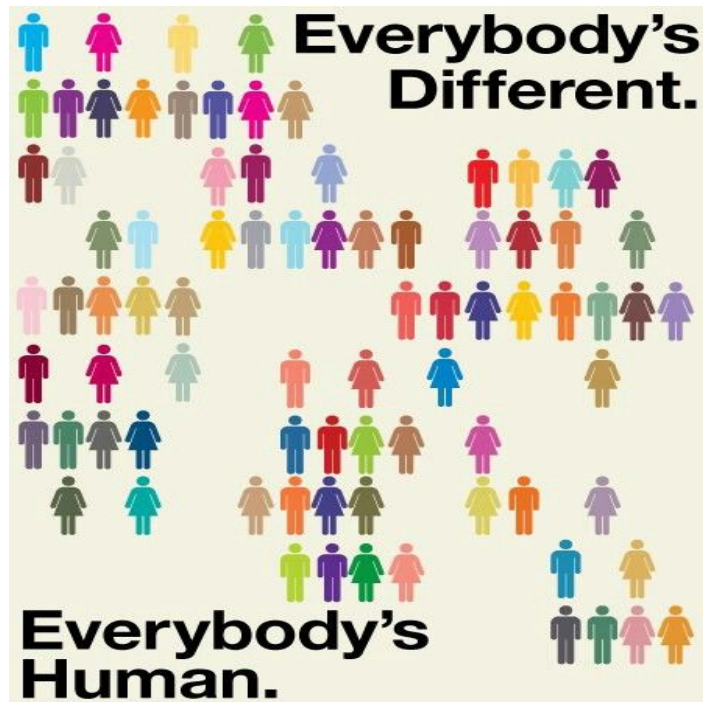
R-Ladies is a worldwide organization whose mission is to promote gender diversity in the R Community.

A Force Behind R!



Top Level Project: (RCDI-WG) Diversity and Inclusion

Mission: broadly consider how the R Consortium can best encourage and support diversity and inclusion practices



Active Tasks:

- Revise CoC for Events
- Conference Organizers Checklists
- Guidelines for Recruiting Speakers
- R in Africa
- Establish diversity grants for international travel

Funded Projects

- Catalyzing R-hub adoption: \$46,050: Maëlle Salmon
- Conference Management System: \$32,000: Heather Turner
- Tracking of R Consortium Activities: \$5,250: Ubah Benaiah
- Tools & Templates for Teaching Materials: \$10,000: F. Michonneau
- An Earth backend for testing stars: \$5,000: Edzer Pebesma
- Editorial assistance for the R Journal: \$50,000: Di Cook
- Forwards Workshops for Women and Girls: \$25,000: Di Cook
- Future Minimal API: Spec & Test Suite: \$10,000: H. Bengtsson
- Joint Profiling of Native and R Code: \$11,000: Kirill Müller
- Licensing R Guidelines and tools: \$6,000: Colin Fay

Funded Projects (continued)

- Maintaining DBI: 2\$6,500: Kirill Müller
- SatRdays: \$10,000 Gergely Daroczi and Steph Locke
- serveRless: \$10,000: Christoph Bodner
- Spatial data management: \$20,000: E. Blondel
- Formulae for Linear Mixed Models: \$6,000: E. Tanaka
- Next-gen text layout in grid and ggplot2: \$25,000: C. Wilke
- R infrastructure on Windows and MacOS: \$62,400: J. Ooms
- Preserving Algorithmic Knowledge: \$4,000US: John Nash
- Platform for missing values methods: \$10,000: J. Josse & N. Tierney
- R Validation Hub: \$4,000: Lyn Taylor
- Refactoring the SWIG R module: \$10,000: Richard Beare

R Consortium Working Groups

Active:

- Census Working Group: Tools for working with census data
- R Community Diversity and Inclusion Working Group:
- R / Medicine: Clinicians and other medical professionals collaborating on events and advocacy of R
- R / Pharma: R users in the pharmaceutical industry collaborating on events and advocacy of R
- R Validation Hub: Standard for validating packages for the regulated Pharmaceutical industry

Concluded:

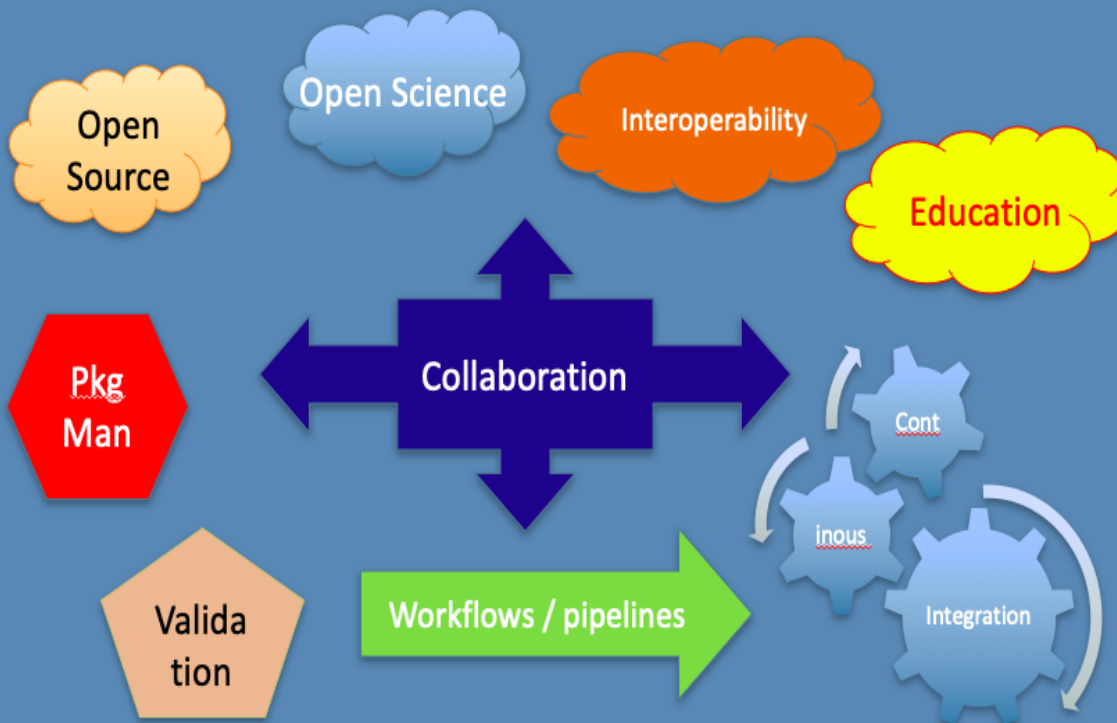
- Code Coverage: Ddevelop code coverage tools - Future-proof native APIs for R: Assess the current API - histoRicalg: Document and test older, essential Fortran and C code - Unified Framework for Distributed Computing

R / Pharma

Some Dominant R / Pharma Themes

Tools

- Shiny
- R Markdown
- GitHub
- Drake
- Docker
- Jenkins
- [RStudio Cloud](#)
- HPC
- Bioconductor
- Arrow
- [Plotly](#)



R Validation Hub



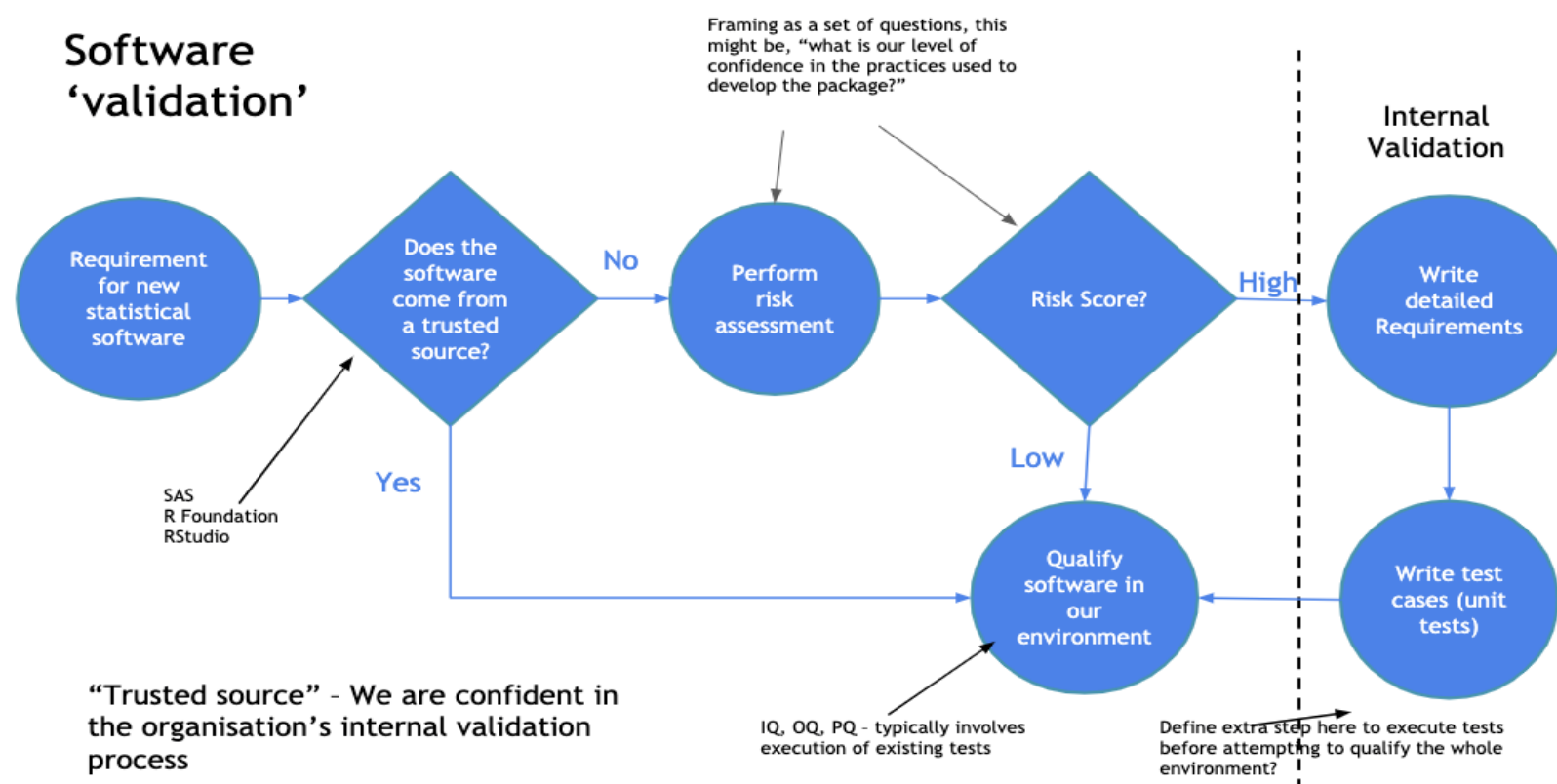
R Validation Hub Goals:

- Validate R packages for use in the Pharmaceutical industry
- Provide guidelines for assessing package risk
- Be a platform for sharing tests
- Provide links to relevant QA information
- Be free to use

R Validation Hub Roadmap

- Process and Communication
 - Publish website
 - White paper
 - Tools at pilot implementation stage
- Validation / Qualification Suite
 - Release metrics package
 - Shiny app and reporting tools for risk assessment
 - Test suite & test execution tool
 - Provide an example deployment
 - Criteria for package inclusion in Pharma Repository
- Build Pharma R Repository

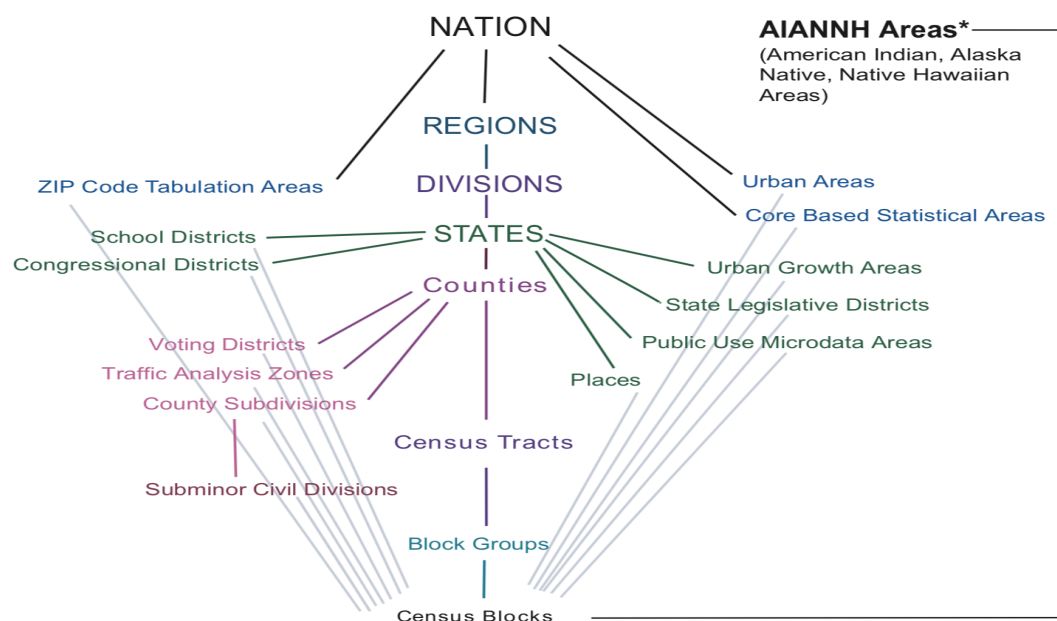
Risk Assessment Workflow



Census Working Group

- Work in cooperation with the U.S. Census Bureau to facilitate the use of Census data among R users
- [A Guide to Working with US Census Data in R](#)

Standard Hierarchy of Census Geographic Entities



R / Medicine



[Photo by Samuel Zeller on Unsplash](#)

An annual conference and a year round effort to help promote R and provide education to clinicians and others involved in the analytical aspects of Medicine.

- [R/Medicine](#)
- [R/Medicine Workshops](#)
- [Statistics in Glaucoma](#)
- [Analyzing the HIV pandemic](#)

Why Should a Company Join the R Consortium?

"We see R as the future of statistical analysis because of its flexibility and the strong active community behind it."

Alun Bedding: Director of Biostatistics at Genentech & R Consortium Board Member

- Supporting R is a strategic investment
- Joining protects your current investment
- Have a say in how infrastructure develops
- Share infrastructure support costs



How Can Individuals Get Involved?

- Join or start a working group
- Submit a proposal for funding
- Join or start a RUG or R-Ladies group
- Submit blog posts to the R Consortium

Thank You



Contact information:

- joseph.rickert@rstudio.com
- @RstudioJoe
- This presentation is available on [RStudio Cloud: https://bit.ly/2qDrtOZ](https://bit.ly/2qDrtOZ)