

Opening up Clinical Trials: 7 Trends Towards Open Source

Jeremy Wildfire, Rho

PhUSE Single Day Event, 12/1/2016

Chapel Hill, NC



What is Open Source?

The [Open Source Initiative](#) has a good definition:

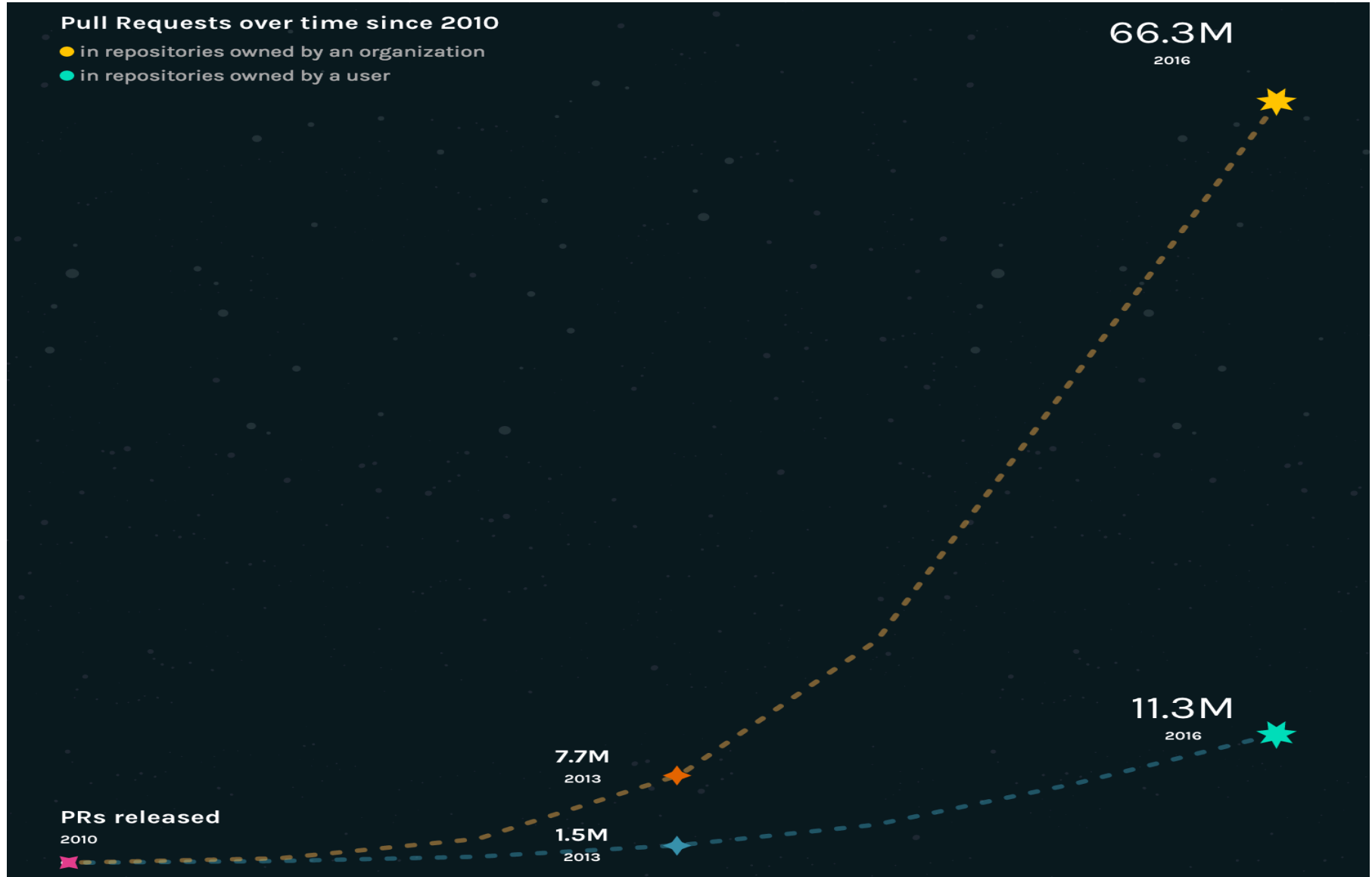
Generally, Open Source software is software that can be freely accessed, used, changed, and shared (in modified or unmodified form) by anyone.

7 Trends Towards Open Source

(relevant to medical research in the last 10 years or so)

1. Mature Open Source Ecosystem
2. Data Science
3. Regulatory Guidance Updates
4. Reproducible Research
5. Slow (but accelerating) Industry Adoption
6. R can't be ignored
7. Open Source and SAS

Trend 1: Mature Open Source Ecosystem



Link: [State of the Octoverse](#) (aka github)



Hello World

🕒 10 minute read

Example 1 – GitHub Guides

The **Hello World** project is a time-honored tradition in computer programming. It is a simple exercise that gets you started when learning something new. Let's get started with GitHub!

You'll learn how to:

- Create and use a repository
- Start and manage a new branch
- Make changes to a file and push them to GitHub as commits
- Open and merge a pull request

Intro

[What is GitHub?](#)

[Create a Repository](#)

[Create a Branch](#)

[Make a Commit](#)

[Open a Pull Request](#)

[Merge Pull Request](#)

What is GitHub?

GitHub is a code hosting platform for version control and collaboration. It lets you and others work together on projects from anywhere.

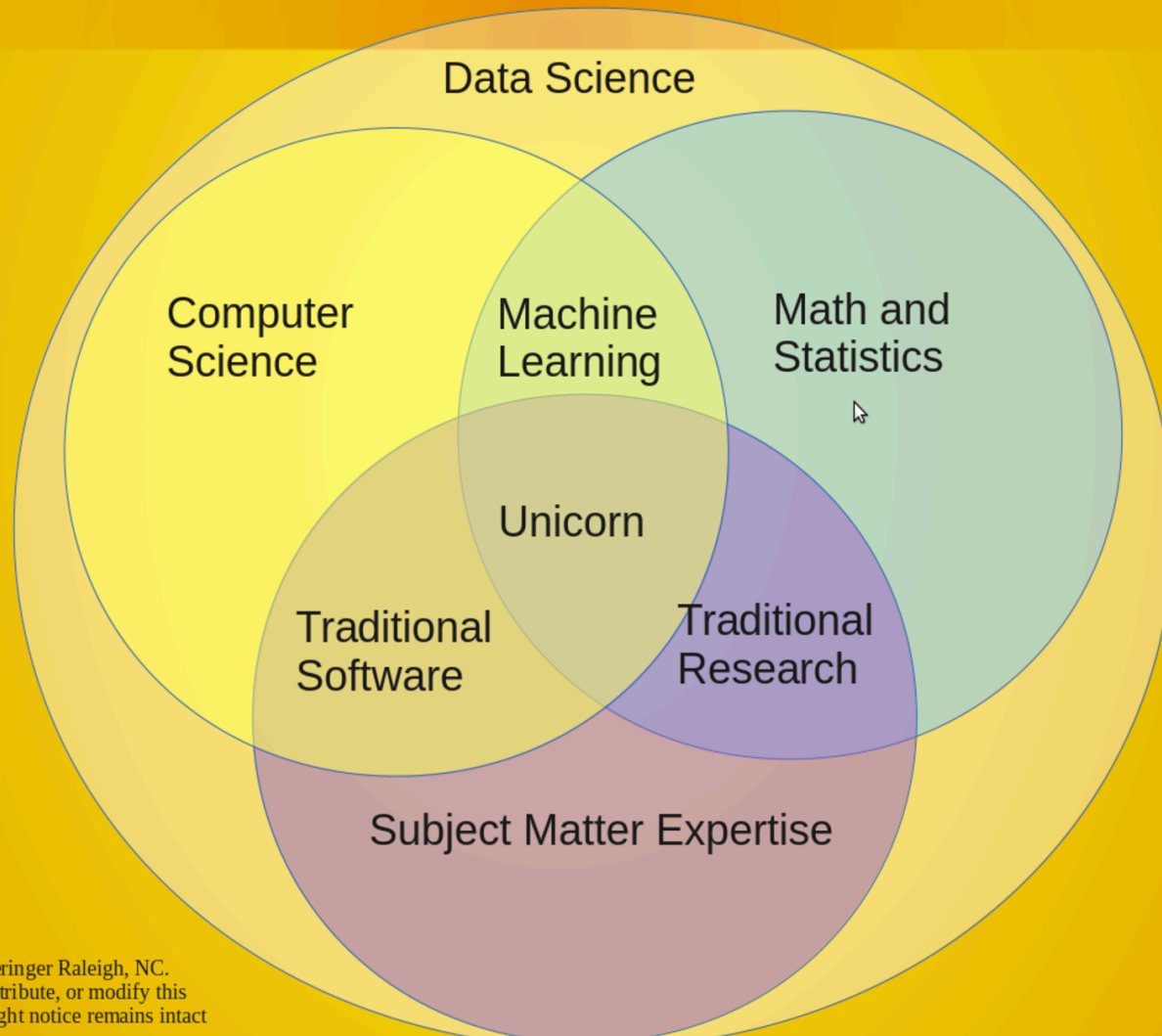
This tutorial teaches you GitHub essentials like *repositories*, *branches*, *commits*, and *Pull Requests*. You'll create

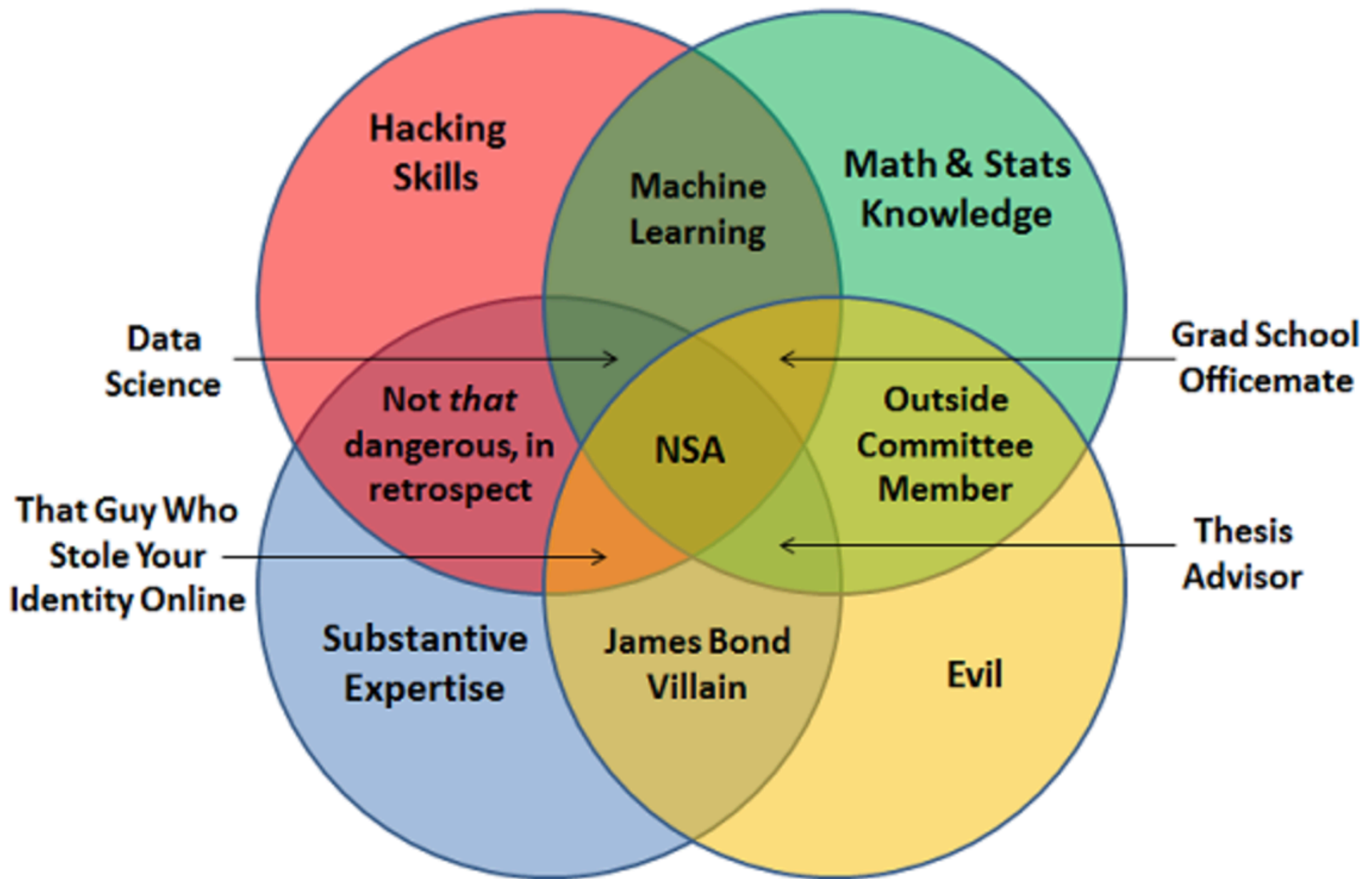
Links: [Rho GitHub](#) & [GitHub Guides](#)

GitHub's Pull Request workflow, a popular way to create and review code.

Trend 2: Data Science

Data Science Venn Diagram v2.0





Link: [Battle of the Data Science Venn Diagrams](#)

Hopkins/Coursera

Data Science Specialization

About This Specialization

Courses

Creators

FAQs

Data Science
Specialization

Enroll

Starts Nov 28

Financial Aid is available for learners who cannot afford the fee. [Learn more and apply.](#)



10 courses

Follow the suggested order or choose your own.



Projects

Designed to help you practice and apply the skills you learn.



Certificates

Highlight your new skills on your resume or LinkedIn.

Courses



Beginner Specialization. No prior experience required.

COURSE 1

The Data Scientist's Toolbox

Current session: Nov 28 — Jan 1.

Commitment

1-4 hours/week

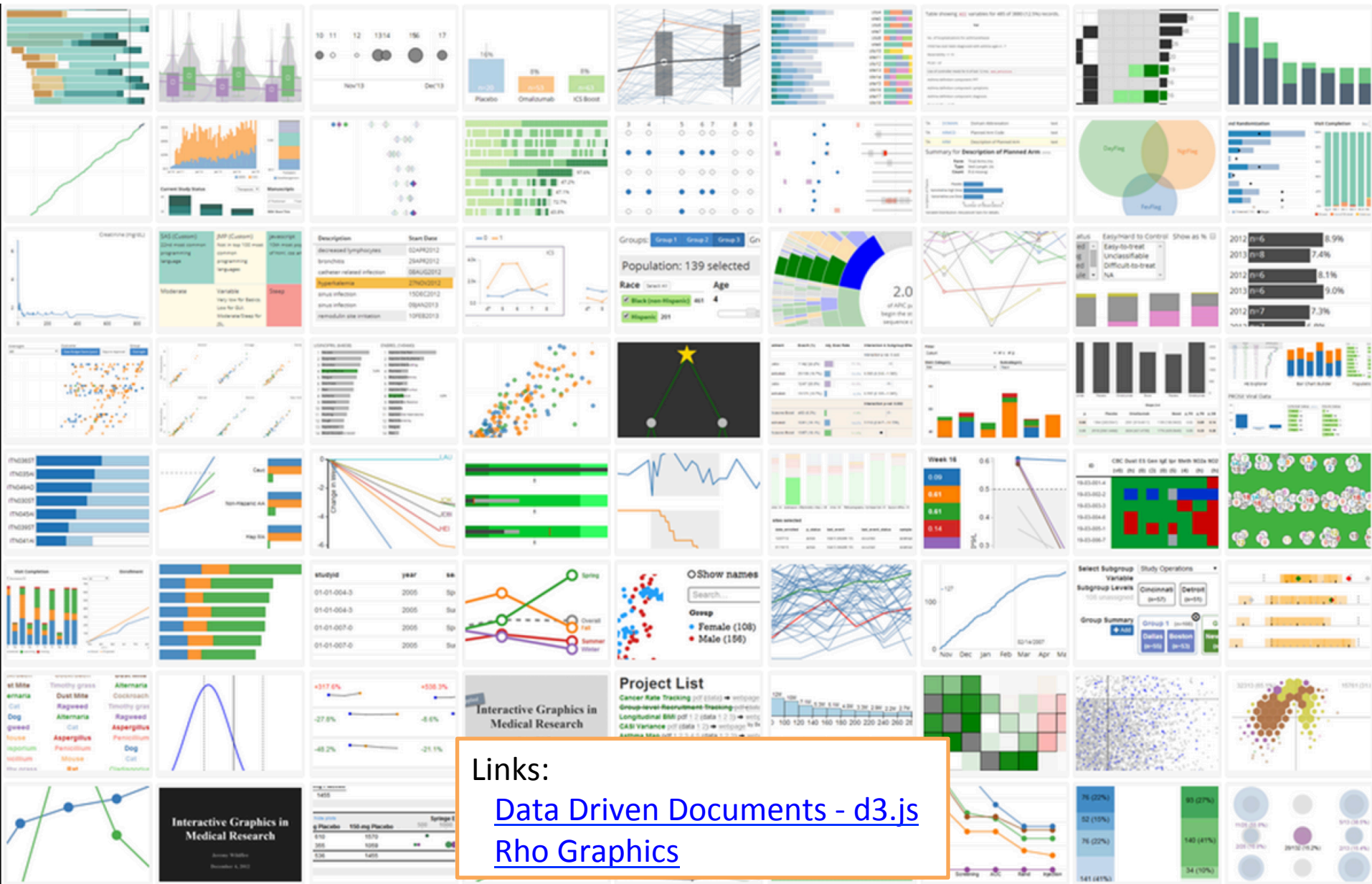
Link: [Coursera Data Science Specialization](#)

Repositories with the most forks

	jtleek/datasharing	38,020
	octocat/Spoon-Knife	28,738
	rdpeng/ProgrammingAssignment2	27,499
	twbs/bootstrap	15,669
	tensorflow/tensorflow	14,070
	rdpeng/ExData_Plotting1	12,458
	udacity/frontend-nanodegree-resume	11,553
	LarryMad/recipes	10,229
	barryclark/jekyll-now	10,070
	angular/angular.js	9,334

Link: [State of the Octoverse](#) (aka github)

Trend 2.1: Interactive Data Visualization

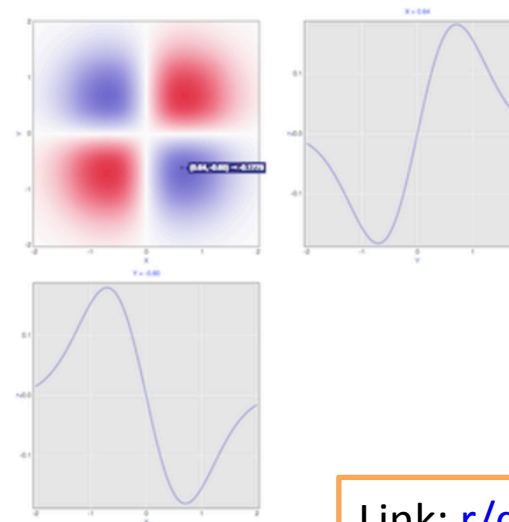
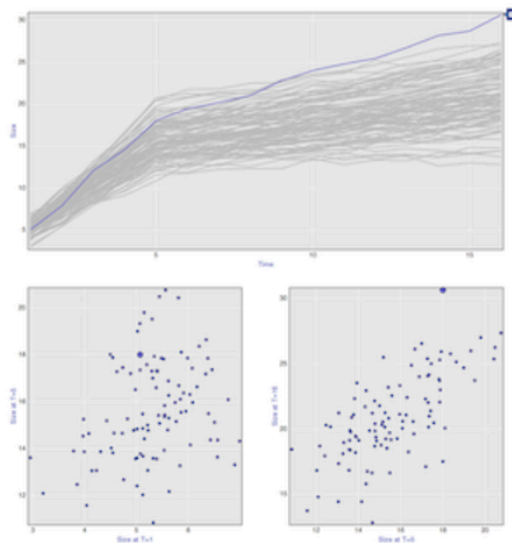
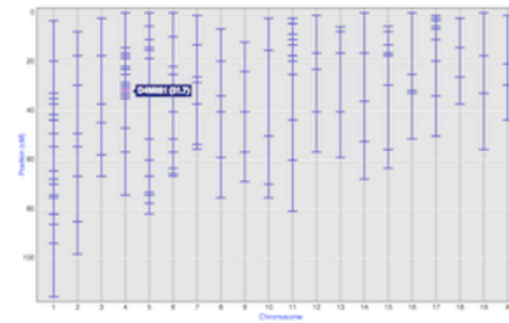
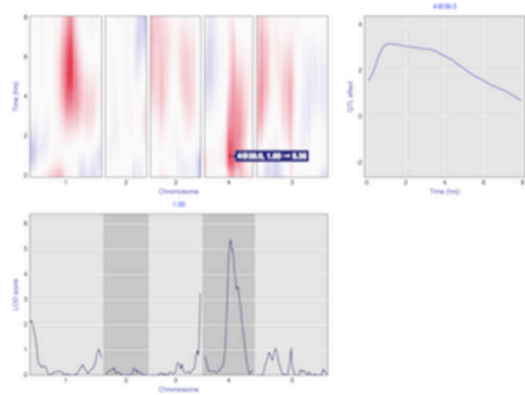
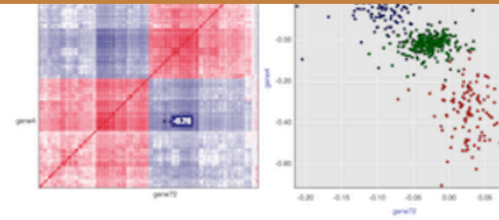
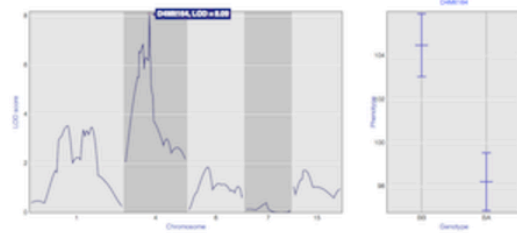


Links:

Data Driven Documents - d3.js

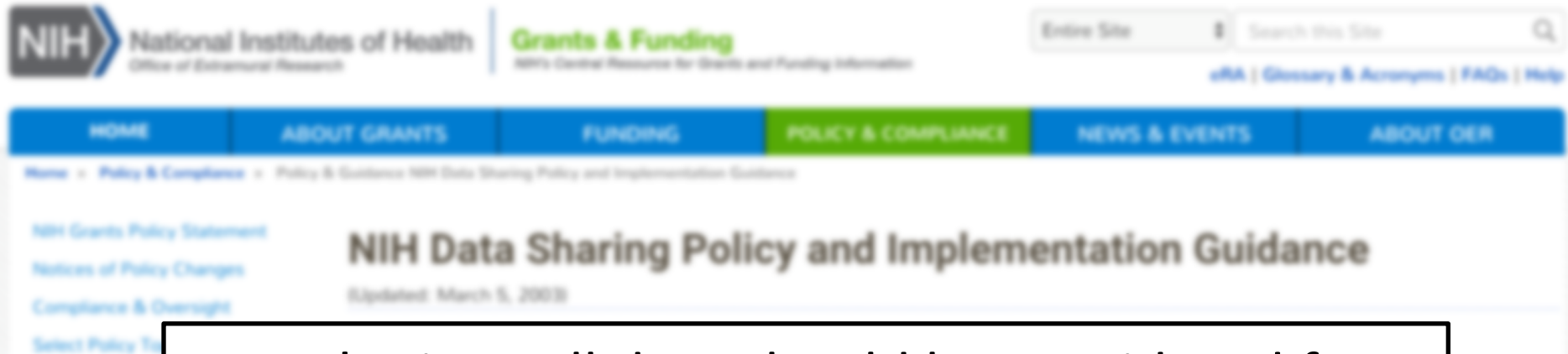
Rho Graphics

Example 2 – qtlcharts (Karl Broman, UW-Madison)



Link: r/qtlcharts

Trend 3: Regulatory Guidance Updates



In NIH's view, all data should be considered for data sharing. **Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data.**

Link: [NIH Data Sharing Policy](#) (2003)

Introduction

1 - Objectives

2 - Scope and Applicability

3 - Three-Step Software Solutions Analysis

4 - Government-Wide

5 - Open Source Software

6 - Exceptions to Government-Wide

7 - Implementation

Appendix A - Definitions

Appendix B - Glossary

Appendix C - Acronyms

Appendix D - References

Appendix E - History

M-16-21

MEMORANDUM FOR THE HEADS OF DEPARTMENTS AND AGENCIES

FROM:

Tony Scott

This policy also establishes a pilot program that requires agencies, when commissioning new custom software, **to release at least 20 percent of new custom-developed code as Open Source Software (OSS)** for three years, and collect additional data concerning new custom software to inform metrics to gauge the performance of this pilot.

For Industry

[Home](#) > [For Industry](#) > [Data Standards](#) > [Study Data Standards](#)

Study Data Standards

[Study Data Standards
Regulatory Submission
Statement](#)

[Position on Use of
Tools](#)

[Data Standards for
and Collaboration](#)

[Janus Clinical Trials Repository
\(JCTR\)](#)

[Statistical Software Clarifying
Statement](#)

[Study Design Standard](#)

[Study Participation Standard](#)

[Subject Data Standard](#)

FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g. in 21CFR part 11].

<http://www.fda.gov/drugs/development/clinical/regulatory-information/industry/default.htm>. "The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available." Sponsors are encouraged to consult with FDA review teams and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process.

Links:

[Using R in a regulatory environment – FDA Experiences](#)
[FDA Statistical Software Clarifying Statement](#) (2015)

 [Open source on GitHub](#) 

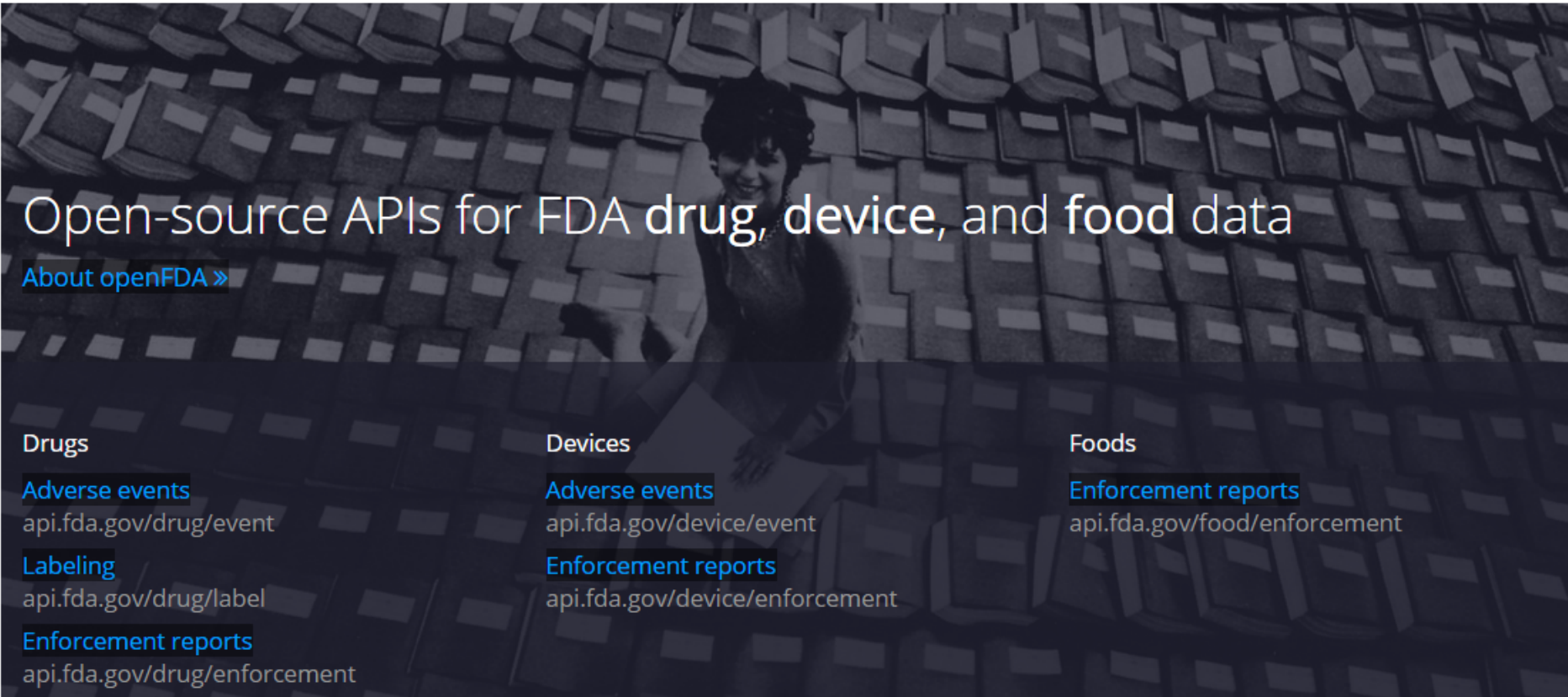
Log bugs, contribute, or fork openFDA

 [Q&A on StackExchange](#) 

Ask questions and get support

 [@openFDA on Twitter](#) 

Keep up to date on [#openFDA](#) 



Open-source APIs for FDA drug, device, and food data

[About openFDA](#) »

Drugs

[Adverse events](#)

api.fda.gov/drug/event

[Labeling](#)

api.fda.gov/drug/label

[Enforcement reports](#)

api.fda.gov/drug/enforcement

Devices

[Adverse events](#)

api.fda.gov/device/event

[Enforcement reports](#)

api.fda.gov/device/enforcement

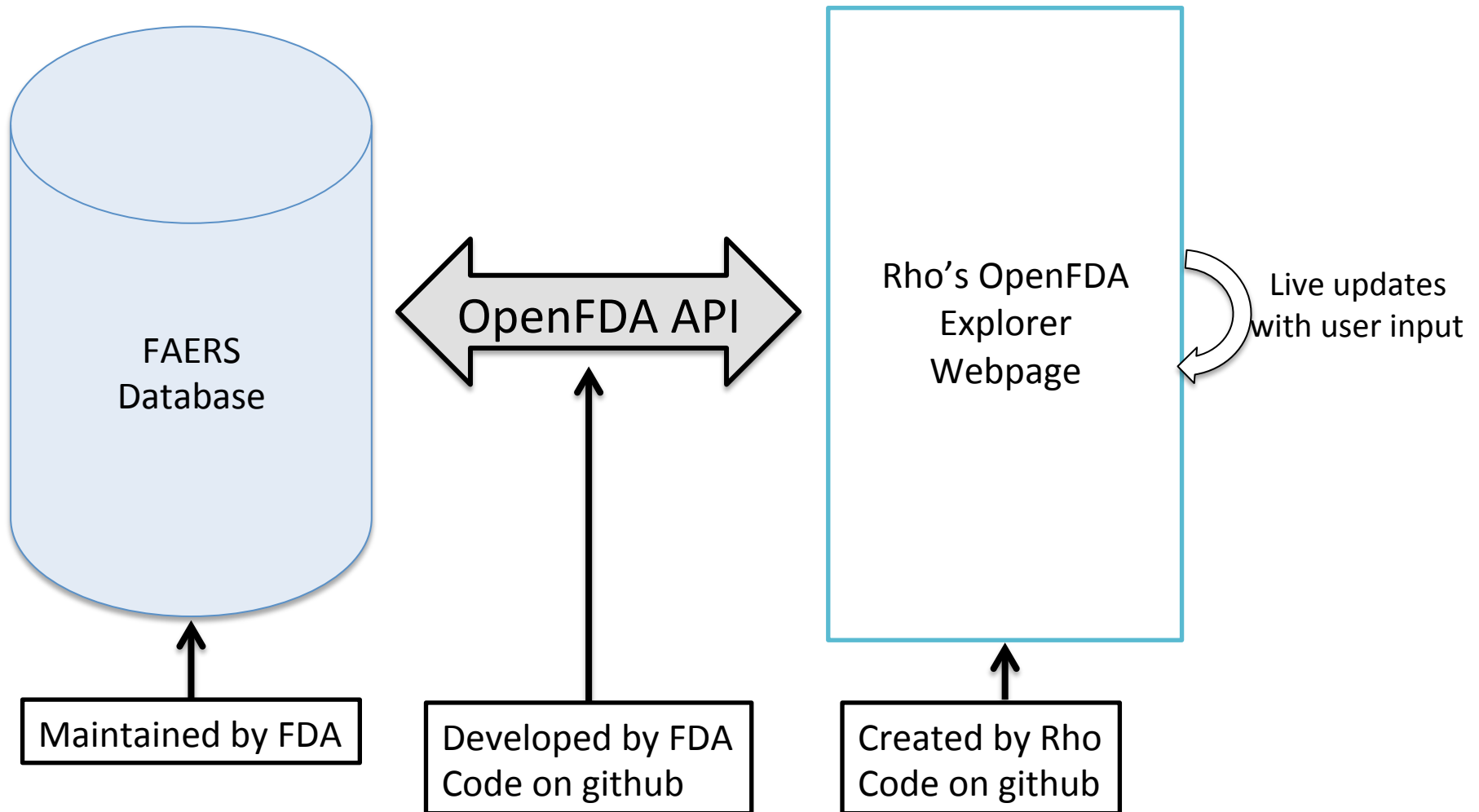
Foods

[Enforcement reports](#)

api.fda.gov/food/enforcement

Link: [OpenFDA homepage](#)

OpenFDA Explorer – System Design



Demo: OpenFDA AE Explorer

This page lets users explore more than 3.5 million adverse event reports collected between 1/1/2004 and 1/1/2015 from the [openFDA project](#).

Pick Comparison Variable

Drug Generic **Manufacturer** Preferred Term

Select Manufacturers to compare

Manufacturers with special characters (e.g. commas) may not appear.

Mylan Pharmaceuticals Inc. (1090580), Major Pharmaceuticals (774569), Qualitest Pharmaceuticals (742696) ▼

Show/Hide Results [Show All](#)

Gender Age **Adverse Events** Drug Generic Manufacturer Report Date

Manufacturer	Mylan Pharmaceuticals Inc.	Major Pharmaceuticals	Qualitest Pharmaceuticals
Count	1,090,580 reports	774,569 reports	742,696 reports
Gender	1 Female 2 Male 3 Unknown 45633 reports with no value (4%)	1 Female 2 Male 3 Unknown 31321 reports with no value (4%)	1 Female 2 Male 3 Unknown 28955 reports with no value (4%)
Age	1 0 To 18 2 18 To 35 3 35 To 45 4 45 To 60 5 60+ 337495 reports with no value (31%)	1 0 To 18 2 18 To 35 3 35 To 45 4 45 To 60 5 60+ 240523 reports with no value (31%)	1 0 To 18 2 18 To 35 3 35 To 45 4 45 To 60 5 60+ 224626 reports with no value (30%)
Adverse Events	1 Nausea 2 Drug Ineffective 3 Dyspnoea 4 Fatigue 5 Diarrhoea 6 Vomiting	1 Nausea 2 Dyspnoea 3 Drug Ineffective 4 Fatigue 5 Pain 6 Diarrhoea	1 Nausea 2 Dyspnoea 3 Drug Ineffective 4 Fatigue 5 Diarrhoea 6 Dizziness

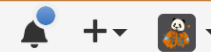
Top 10 Adverse Events

Add 5 Remove 5

Search for Preferred Term

search

Link: [Rho's OpenFDA AE Explorer](#)



Search

openfda

Search



Repositories

89



Code

3,954



Issues

324



Wikis

49



Users

1

We've found 89 repository results

Sort: Best match ▾

FDA/openfda

openFDA is a research project to provide open APIs, raw data downloads, documentation and examples, and a developer c...

Python ★ 254 🔗 60 Updated on Feb 9

rOpenHealth/openfda

Convenient access to the *OpenFDA* API

R ★ 34 🔗 10 Updated on Nov 2, 2015

esridc/openfda

Demonstration of Agile Process

JavaScript ★ 2 🔗 6 Updated on Jul 7, 2015

Languages

JavaScript 26

CSS 9

R 9

HTML 7

Java 7

Python 6

Go 2

Jupyter Notebook 2

Ruby 2

Link: [OpenFDA repositories on Github](#)

Trend 4: Reproducible Research

Availability of data, material and methods

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must **also** be disclosed in the submitted manuscript.

After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings.

See sections below for details on:

- [reporting requirements](#)
- [availability of data](#)
- [availability of materials](#)
- [availability of computer code](#)
- [experimental protocols](#)
- [clinical trials](#)
- [further reading](#)

Link: [Nature Policies for Publication](#)

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

FEBRUARY 26, 2015

VOL. 372 NO. 9

Randomized Trial of Peanut Consumption in Infants at Risk for Peanut Allergy

George Du Toit, M.B., B.Ch., Graham Roberts, D.M., Peter H. Sayre, M.D., Ph.D., Henry T. Bahnson, M.P.H., Suzana Radulovic, M.D., Alexandra F. Santos, M.D., Helen A. Brough, M.B., B.S., Deborah Phippard, Ph.D., Monica Basting, M.A., Mary Feeney, M.Sc., R.D., Victor Turcanu, M.D., Ph.D., Michelle L. Sever, M.S.P.H., Ph.D., Margarita Gomez Lorenzo, M.D., Marshall Plaut, M.D., and Gideon Lack, M.B., B.Ch., for the LEAP Study Team*

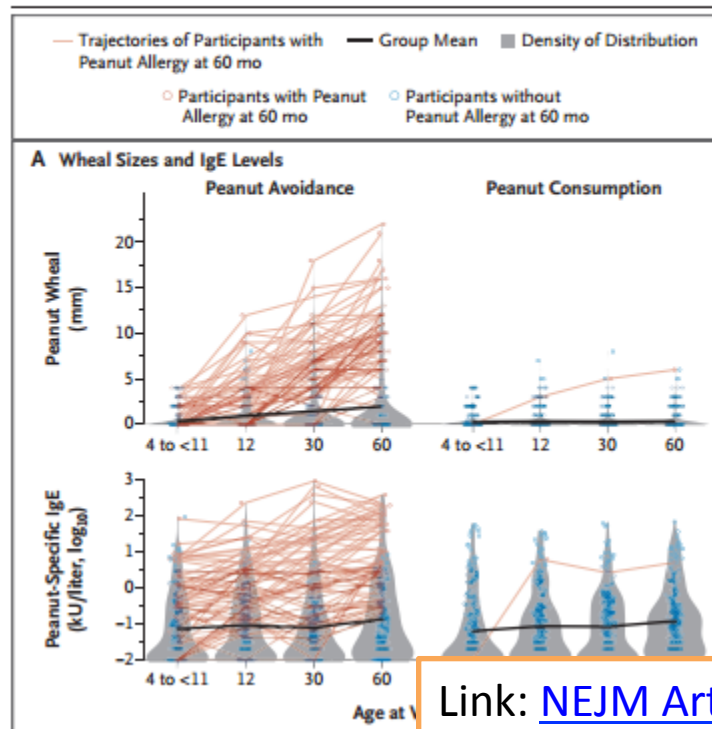
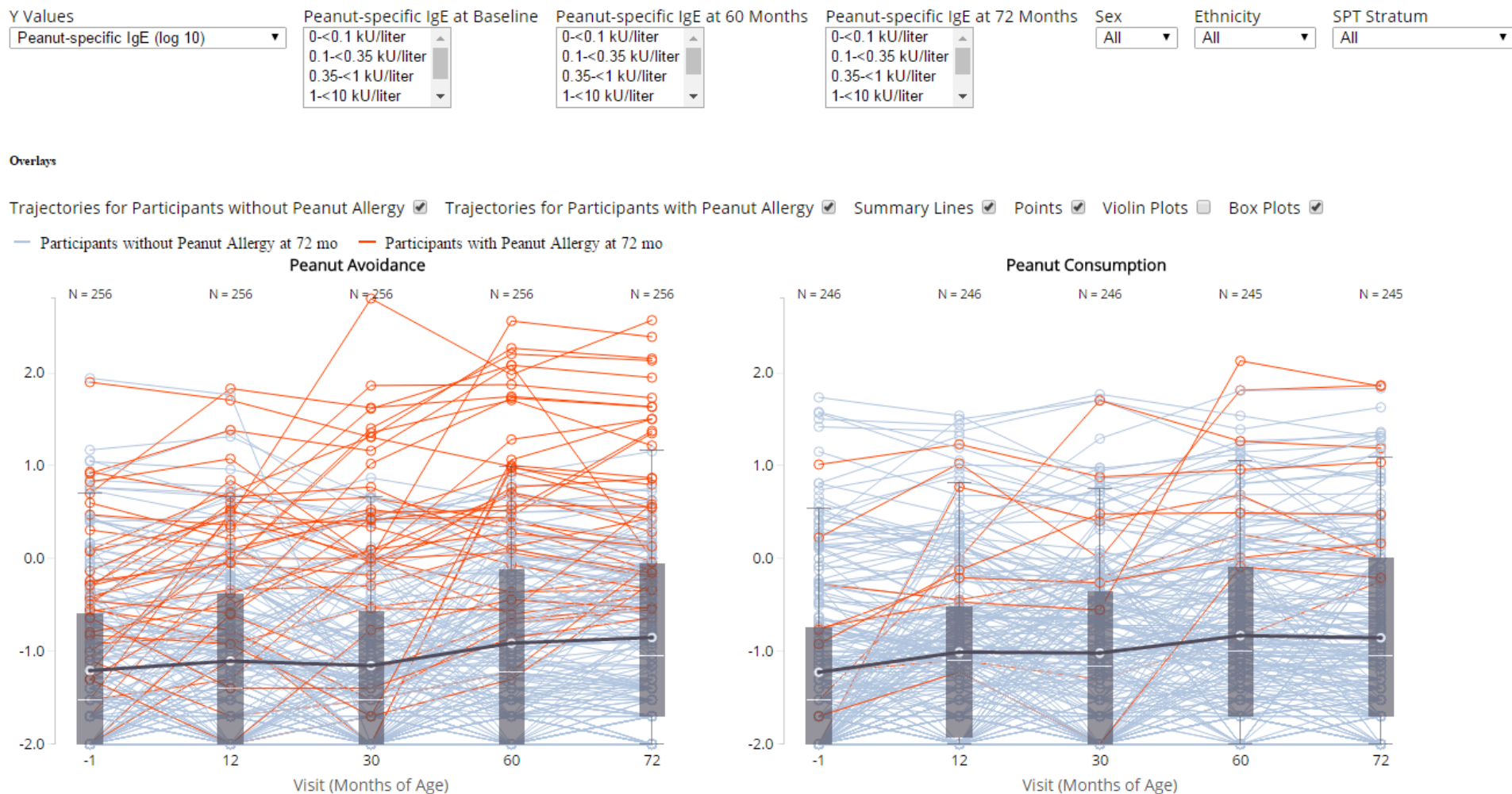


Figure 3. Immunologic Outcomes for the Peanut-Avoidance and Peanut-Consumption Groups at Baseline (4 to <11 Months of Age) and at 12, 30, and 60 Months of Age.

Panel A shows wheal sizes after the peanut-specific skin-prick test and the levels of peanut-specific IgE in participants in the avoidance and consumption groups who met the per-protocol criteria. The solid black lines show the group mean over the course of the study period; the mean wheal size after the peanut-specific skin-prick test differed significantly between the randomized groups at all time points after baseline ($P=0.002$ at 12 months and $P<0.001$ at 30 months and 60 months). The thin red lines represent the trajectory of the development of allergic responses among participants who were allergic at 60 months of age. Panel B shows the levels of peanut-specific IgG and IgG4 and the peanut-specific IgG4:IgE ratio over the course of the study period. The means of each of these measures differed significantly between the two study groups at all postbaseline time points ($P<0.001$). The log₁₀ of the ratio of peanut-specific IgG4:IgE was calculated after the peanut-specific IgG4 levels were converted from micrograms per liter to nanograms per milliliter and the peanut-specific IgE level from kilo unit per liter to nanograms per milliliter using the use of the formula $\text{IgG4} \div (\text{IgE} \times 2.4)$.

Link: [NEJM Article \(pdf\)](#)



Links:

[TrialShare Repository](#) (sign up required)
[Interactive Figure](#)

Interlude: SAS vs. R





Trend 5: R can no longer be ignored



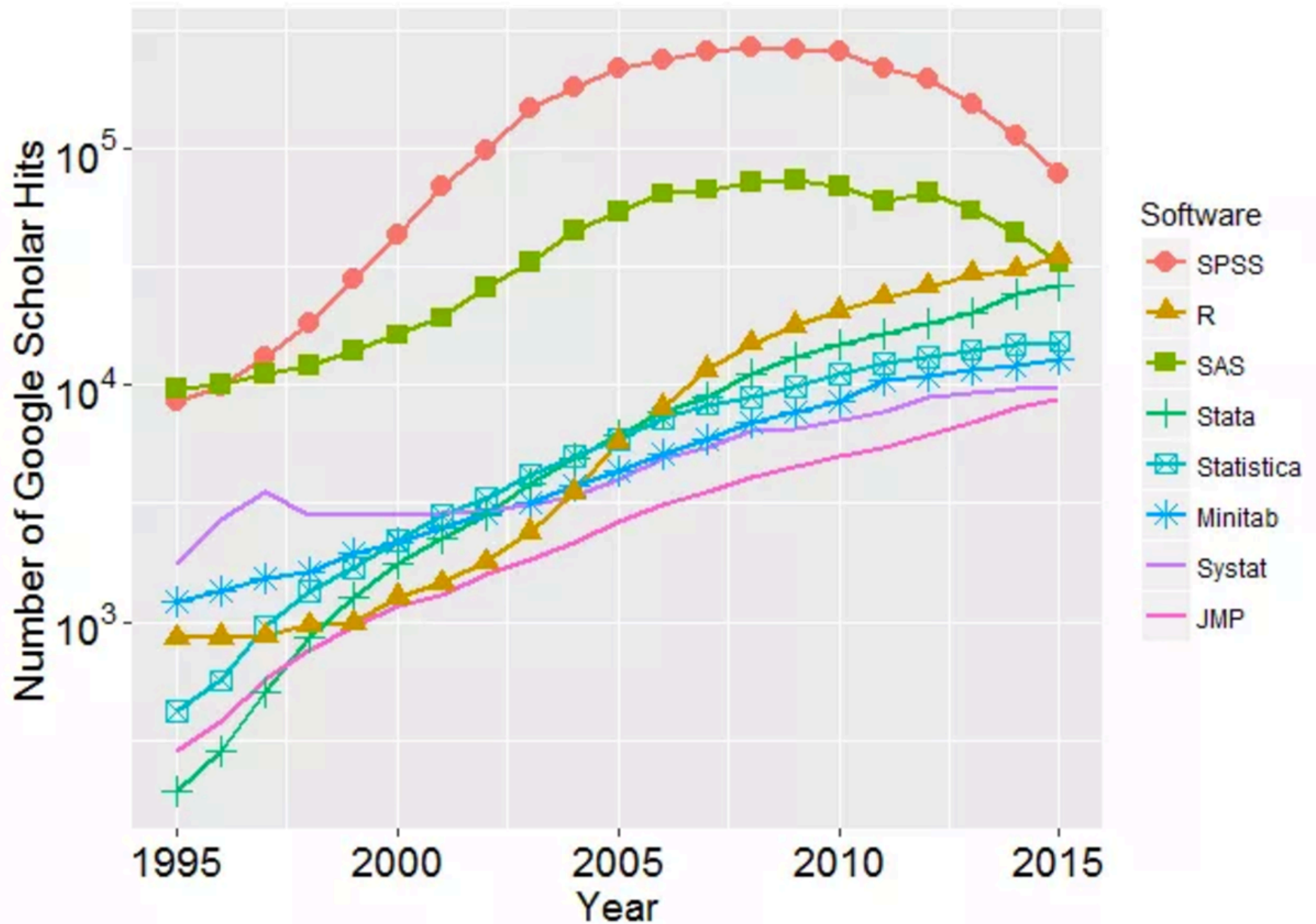
Job Trends from Indeed.com

Orange line: "R D" "A R" "H R" "R N" toys kids " R Walgreen" walmart "HVAC R" "R Bard" and (

Blue line: SAS "system administrator" "school age" sata firmware scsi raid samsung scandina

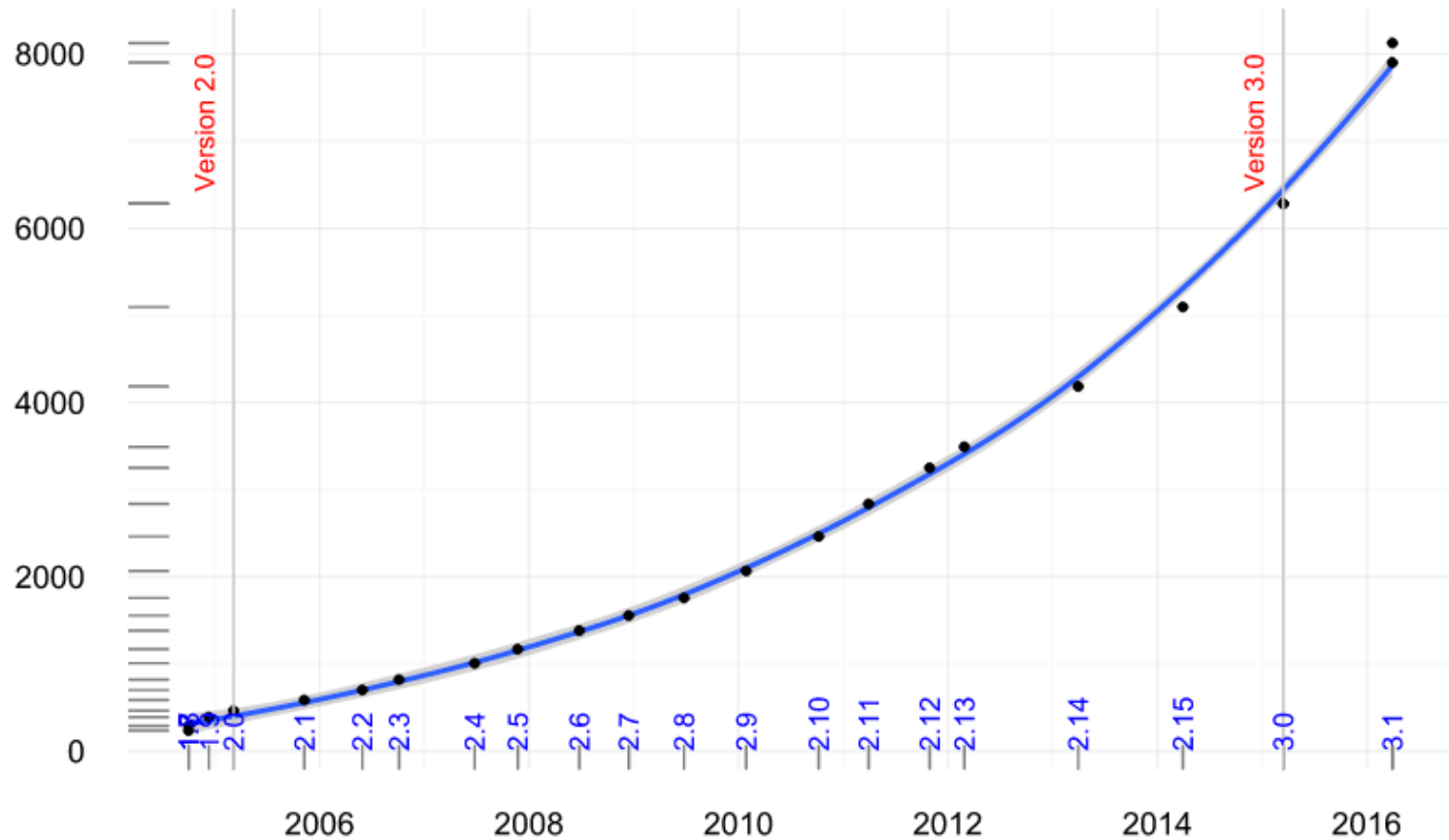


Link: [The Popularity of Data Analysis Software](#)



Link: [The Popularity of Data Analysis Software](#)

Number of CRAN packages per R version



Link: [On the Growth of CRAN Packages](#)

Select Inputs

Drug Variable

patient.drug.openfda.brand_name

Time Variable

receiptdate

Select Drug and Event...

Drug Name: *None*

Match drug name:

- ☐ Exactly
☒ Any Term

Event Term: *None*

Match event name:

- ☐ Exactly
☒ Any Term

Plot PRR between

1989-06-30

to

2016-11-30

Dynamic PRR

PRR Over Time

Report Counts and PRR

Counts For Drugs In Selected Reports

Counts For Events In Selected Reports

Meta Data and Queries

Other Apps

Data Reference

About

Query: https://api.fda.gov/drug/event.json?search=&limit=999&count=patient.drug.openfda.brand_name.exact&skip=0

Most Common Drugs In Selected Reports

Tables

Word Cloud

Show 25 entries

Search:

D	L	Drug Name	Count	Cumulative Sum
D	L	REGULAR STRENGTH PAIN RELIEF	264,137	264137
D	L	ECOTRIN	229,754	493891
D	L	BUFFERIN LOW DOSE BUFFERED ASPIRIN	227,128	721019
D	L	BAYER LOW DOSE	226,778	947797
D	L	BAYER GENUINE ASPIRIN	226,449	1174246
D	L	ASPRIN	226,434	1400680
D	L	LOW DOSE ASPIRIN ENTERIC SAFETY COATED	226,210	1626890

Link: [OpenFDA Analytic and Research Tools](#)

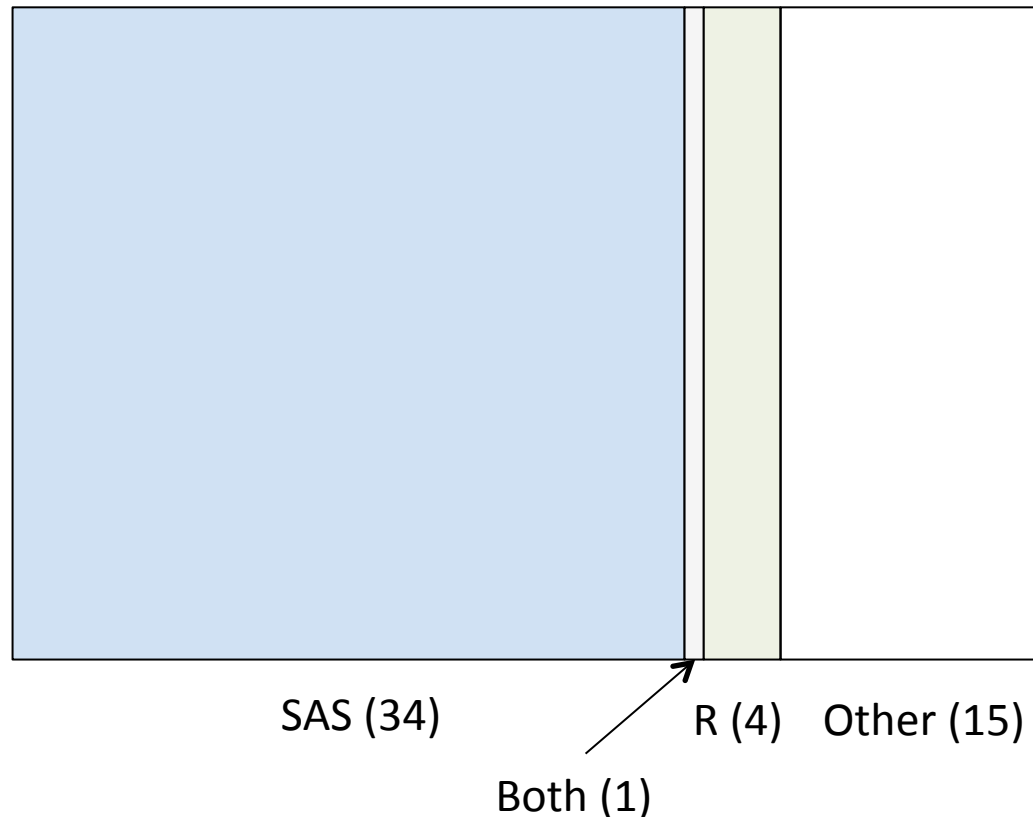
Trend 6: Slow (but accelerating) Industry Adoption

Distribution of industries among GitHub Enterprise customers



Link: [State of the Octoverse](#) (aka github)

> 60% of Articles from PhUSE 2016 focused on SAS



I reviewed 54 abstracts from PHUSE 2016 under “Applications and Software development”, “Coder’s Corner”, “Data Visualization”, “Coding Solutions”, “Industry Starters” and “Trends and Technology”

Link: [PhUSE 2016 papers](#)

Paper DV06

Interactive Visualization of Linked Data

Tim Williams, UCB Biosciences Inc., Raleigh, USA

About: [A Placebo-controlled Study of Levetiracetam In Children \(1mo to 4yrs of Age\) With Partial Onset Seizures.](#)

An Entry of Type : http://bio2rdf.org/clinicaltrials_vocabulary:Resource, within Data Space : od.openlinksw.com associated with source [document\(s\)](#)

Type: [clinicaltrials resource \[clinicaltrials_vocabulary:Resource\]](#) | [New Facet based on Instances of this Class](#)

Attributes	Values
type	Clinical_Study [clinicaltrials_vocabulary:Clinical_Study] clinicaltrials_resource [clinicaltrials_vocabulary:Resource]
label or name	A Placebo-controlled Study of Levetiracetam In Children (1mo to 4yrs of Age) With Partial Onset Seizures. [clinicaltrials:NCT00175890]
Title	A Placebo-controlled Study of Levetiracetam In Children (1mo to 4yrs of Age) With Partial Onset Seizures.
Identifier	clinicaltrials:NCT00175890
in_dataset	http://bio2rdf.org/clinicaltrials_resource:bio2rdf:dataset:clinicaltrials_R3
identifiers.org URI	http://identifiers.org/clinicaltrials/NCT00175890
Bio2RDF uri	http://bio2rdf.org/clinicaltrials:NCT00175890
Bio2RDF identifier	NCT00175890
Bio2RDF namespace	clinicaltrials
intervention brows...ervention-browse]	Piracetam [clinicaltrials_resource:3352605dc14d7f0e09469973fc77bc61] Etiracetam [clinicaltrials_resource:61e5650c36f3b7bbecd755cb46761545]

Figure 7 Virtuoso faceted browser view of clinical trial information from ClinicalTrials.Gov, hosted on OpenLinkSW.

The network graph in **Figure 8** displays information about studies that contributed to two data pools used for submissions. Study phase and contribution are easily identified in the display, with drill-down to study information available by clicking on the study nodes. The legend is also interactive. A click on "Phase 1" opens a faceted browser view listing all Phase 1 studies. From there the user can follow the hyperlinked data to obtain details about each of Phase 1 study.

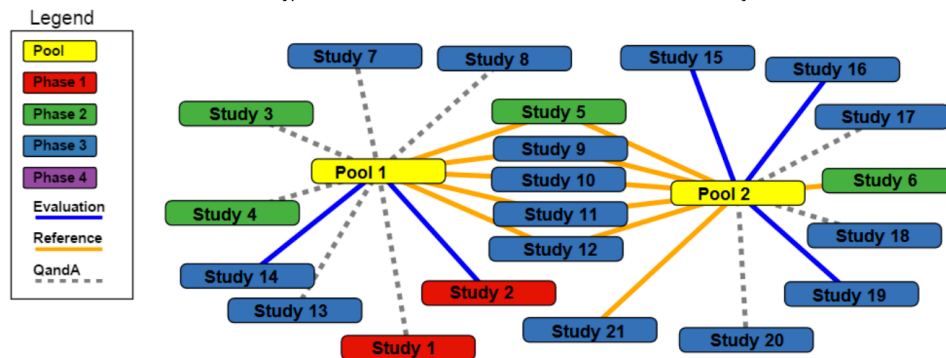


Figure 8 Force network graph showing studies that contribute to data pools.

Link: [PhUSE Paper](#)

Trend 7: Open Source and SAS



This organization

Search

Pull requests

Issues

Gist



PhUSE



<http://www.phuse.eu/>

Repositories

People 1

Search repositories...

Type: All

Language: All

phuse-scripts

Delivery standard industry analyses, built upon CDISC standards for analysis data

SAS ★ 12 14 Updated 4 hours ago



Reg2RDF

The Reg2RDF project focuses on representing 21CFR as RDF data.

Web Ontology Language ★ 1 Updated on Jul 20



Top languages

Web Ontology Language SAS
HTML

People

1 >



glow-mdsol
Geoff Low

Links: [PhUSE github](#) & [Paper](#)



Rho Inc.

The contract research organization with unsurpassed expertise, integrity, and passion for clinical research.

Chapel Hill, NC

<http://graphics.rhoworld.com>

graphics@rhoworld.com

Repositories

People 27

Teams 11

Projects 0

Settings

Type: All

Language: SAS

Customize pinned repositories

New

11 results for repositories written in SAS

Clear filter

sas-violinPlot

Violin Plots in SAS

SAS 1 Updated 6 days ago

sas-snapshotCompare Private

Compare dataset-level, variable-level, and ID-level metadata with SAS

SAS Updated 8 days ago

sas-consort Private

A SAS-based tool to aid in the production of CONSORT diagrams.

SAS Updated 9 days ago

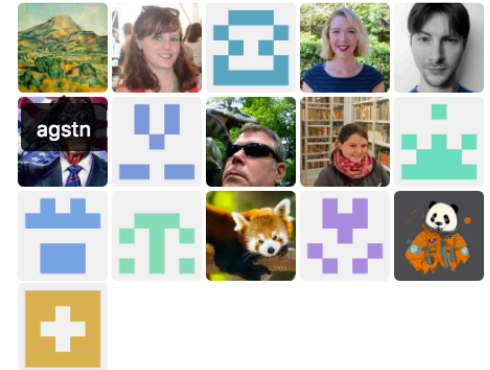
Link: [Github](#)

Top languages

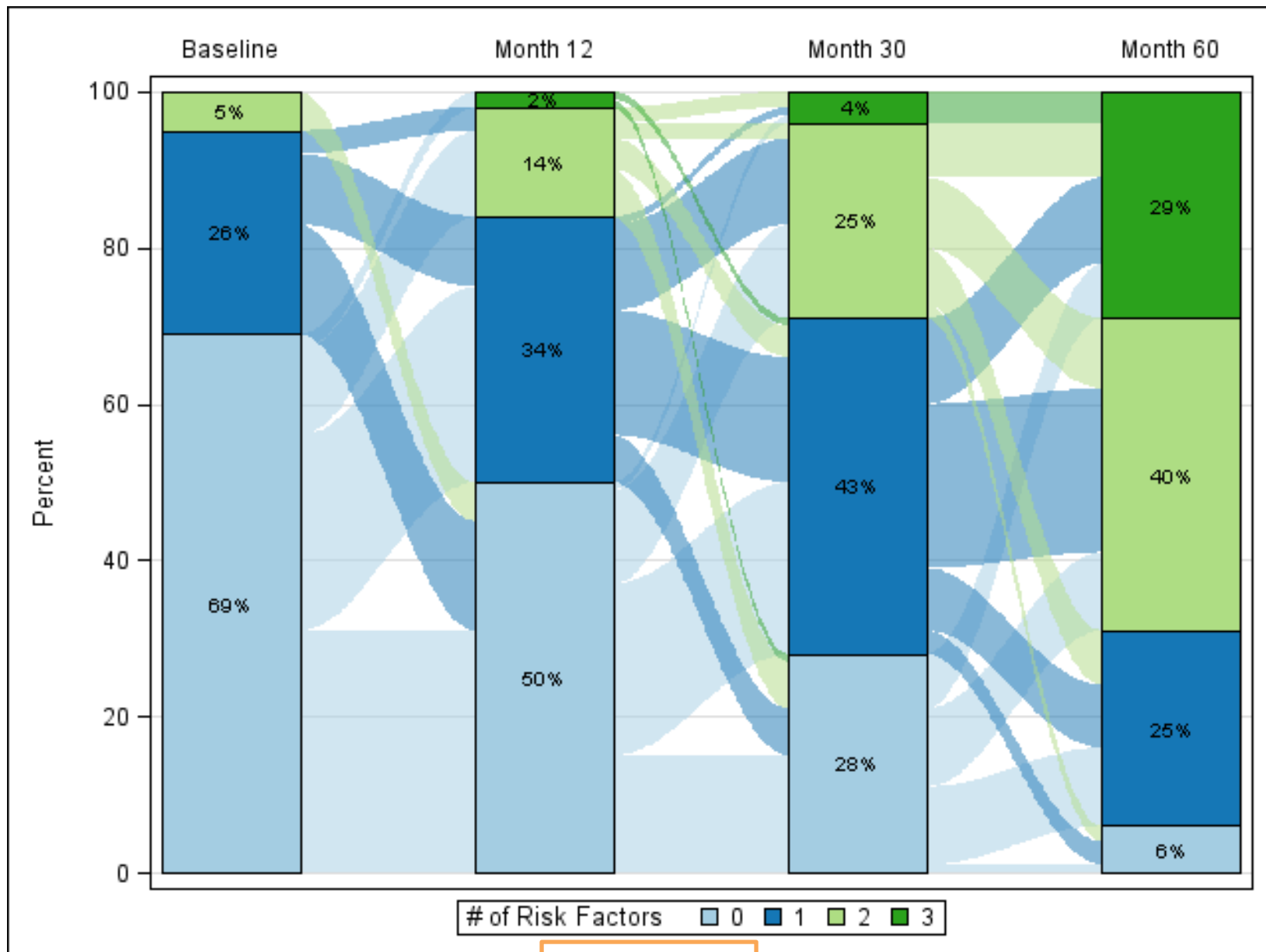
JavaScript SAS R HTML

People

27



Invite someone



7 Trends Towards Open Source

1. Mature Open Source Ecosystem
2. Data Science
3. Regulatory Guidance Updates
4. Reproducible Research
5. Slow (but accelerating) Industry Adoption
6. R can't be ignored
7. Open Source and SAS

Open Source at Rho

Open source development reflects [Rho's Core values](#). A *team culture, critical and creative thinking, and innovation* are at the heart of the open source philosophy. We demonstrate our *integrity and quality* by releasing the details of our process and allowing others to examine and enhance our work. The open source process encourages *agility and adaptability, profitability, and stability* by encouraging thorough documentation, reducing rework, and increasing the visibility of our work.

Link: [Rho's Open Source Handbook](#)

Guiding Principle

When possible, code should be made public.



Rho Inc.

The contract research organization with unsurpassed expertise, integrity, and passion for clinical research.

📍 Chapel Hill, NC

🌐 [http://graphics.rhow...](http://graphics.rhoworld.com)

✉ [graphics@rhoworld.c...](mailto:graphics@rhoworld.com)

📁 Repositories

👤 People 1

Type: All ▾

Language: All ▾

sas-sankeybarchart

A set of SAS macros for creating longitudinal bar charts with Sankey-style overlays.

● SAS ★ 1 Updated 7 hours ago



Top languages

● JavaScript ● SAS ● R

sas-violinPlot

Violin Plots in SAS

● SAS 🍴 1 Updated 6 days ago



People

1 >



qspencer
Quentin Spencer

Webcharts

Reusable, flexible, interactive charts with JavaScript

● JavaScript ★ 9 🍴 1 Updated 7 days ago



open-source-handbook

Open Source Guidelines for Rho Inc.

🍴 1 Updated 7 days ago



Link: [Rho Github](#)



Software Metapapers

Webcharts – A Web-based Charting Library for Custom Interactive Data Visualization

Authors: [Nathan Bryant](#), [Jeremy Wildfire](#) [✉](#)

Abstract

Webcharts is a JavaScript library built on top of D3.js that creates reusable, flexible, interactive charts that are highly customizable. Webcharts provides a method for creating commonly-used charts, including bar charts, scatterplots, and timelines, through a simple configuration scheme. Charts created with Webcharts allow users to dynamically manipulate chart data, appearance, and behavior both through callback functions and input elements that are tied to chart objects. This approach allows users to create reusable charts that range from simple static graphics to complex interactive data exploration tools with custom user interfaces, all using the same library.

Keywords: [Data Visualization](#), [Javascript](#), [d3.js](#)

JUMP TO

0 COMMENTS

[Abstract](#)[\(1\) Overview](#)[\(2\) Availability](#)[\(3\) Reuse potential](#)[Acknowledgements](#)[Competing Interests](#)[References](#)

Link: [Journal of Open Research Software Manuscript](#)

Speed Bumps

- New skills to master
- Integrating open source and commercial systems is complex
- Invites scrutiny ... which can lead to work
- Rethinking benefits of proprietary work
- Regulatory environment is a work in progress