R: A Hitchhikers Guide to Reproducible Research

```
- Hello
```

Brendan Palmer,
Clinical Research Facility - Cork &
School of Public Health
@B_A_Palmer

CRF-C
HRB Clinical Research Facility Cork

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
School of Public Health

Daily overview

09.30 **Lecture**

- Introduction to the morning session and catch-up

10.30 **Coffee break**

11.00 **Practical**

- Mix of live demonstrations, walkthroughs and worksheets

13.00 **Lunch break**

14.00 **Lecture**

- Introduction to the afternoon session

14.45 **Practical**

- Mix of live demonstrations, walkthroughs and worksheets

16.00 **Lecture**

- Short presentation on the bigger picture

16.30 **Finish**

What does reproducible mean?

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Psychology is leading the way

SOCIETY FOR THE IMPROVEMENT OF PSYCHOLOGICAL SCIENCE

Program Travel And Hotels Social Events Contact

The need for greater research reproducibility

The Reinhart-Rogoff error – or how not to Excel at economics

April 22, 2013 9.40pm BST

Data and computer code should be made publicly available at an early stage – or else ... esarastudillo

Email
Twitter 88
Facebook 453
LinkedIn
Print

Last week we learned a famous 2010 academic paper, relied on by political big-hitters to bolster arguments for austerity cuts, contained significant errors; and that those errors came down to misuse of an Excel spreadsheet.

Sadly, these are not the first mistakes of this size and nature when handling data. So what on Earth went wrong, and can we fix it?

Harvard's Carmen Reinhart and Kenneth Rogoff are two of the most respected and influential academic economists active today.

241 shades of grey

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Full Length Articles

The secret lives of experiments: Methods reporting in the fMRI literature

Joshua Carp

University of Michigan, Department of Psychology, 530 Church Street, Ann Arbor, MI, 48109, USA

ARTICLE INFO

Article history:
Accepted 3 July 2012
Available online 10 July 2012

Keywords:
fMRI
Methods reporting
Reproducibility
Experimental design
Analysis methods
Statistical power

ABSTRACT

Replication of research findings is critical to the progress of scientific understanding. Accordingly, most scientific journals require authors to report experimental procedures in sufficient detail for independent researchers to replicate their work. To what extent do research reports in the functional neuroimaging literature live up to this standard? The present study evaluated methods reporting and methodological choices across 241 recent fMRI articles. Many studies did not report critical methodological details with regard to experimental design, data acquisition, and analysis. Further, many studies were underpowered to detect any but the largest statistical effects. Finally, data collection and analysis methods were highly flexible across studies, with nearly as many unique analysis pipelines as there were studies in the sample. Because the rate of false positive results is thought to increase with the flexibility of experimental designs, the field of functional neuroimaging may be particularly vulnerable to false positives. In sum, the present study documented significant gaps in methods reporting among fMRI studies. Improved methodological descriptions in research reports would yield significant benefits for the field.

Standing on the shoulder of giants?

Research Article

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
23(5) 524–532

DOI: 10.1177/0956797611430953
http://pss.sagepub.com
SAGE

Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling

Leslie K. John[1], George Loewenstein[2], and Drazen Prelec[3]
[1]Marketing Unit, Harvard Business School; [2]Department of Social & Decision Sciences, Carnegie Mellon University; and [3]Sloan School of Management and Departments of Economics and Brain & Cognitive Sciences, Massachusetts Institute of Technology

Abstract

Cases of clear scientific misconduct have received significant media attention recently, but less flagrantly questionable research practices may be more prevalent and, ultimately, more damaging to the academic enterprise. Using an anonymous elicitation format supplemented by incentives for honest reporting, we surveyed over 2,000 psychologists about their involvement in questionable research practices. The impact of truth-telling incentives on self-admissions of questionable research practices was positive, and this impact was greater for practices that respondents judged to be less defensible. Combining three different estimation methods, we found that the percentage of respondents who have engaged in questionable practices was surprisingly high. This finding suggests that some questionable practices may constitute the prevailing research norm.

COMMENTARY

Scientists behaving badly

To protect the integrity of science, we must look beyond falsification, fabrication and plagiarism, to a wider range of questionable research practices, argue **Brian C. Martinson, Melissa S. Anderson** and **Raymond de Vries**.

Table 1 | Percentage of scientists who say that they engaged in the behaviour listed within the previous three years ($n = 3,247$)

Top ten behaviours	All	Mid-career	Early-career
1. Falsifying or 'cooking' research data	0.3	0.2	0.5
2. Ignoring major aspects of human-subject requirements	0.3	0.3	0.4
3. Not properly disclosing involvement in firms whose products are based on one's own research	0.3	0.4	0.3
4. Relationships with students, research subjects or clients that may be interpreted as questionable	1.4	1.3	1.4
5. Using another's ideas without obtaining permission or giving due credit	1.4	1.7	1.0
6. Unauthorized use of confidential information in connection with one's own research	1.7	2.4	0.8 ***
7. Failing to present data that contradict one's own previous research	6.0	6.5	5.3
8. Circumventing certain minor aspects of human-subject requirements	7.6	9.0	6.0 **
9. Overlooking others' use of flawed data or questionable interpretation of data	12.5	12.2	12.8
10. Changing the design, methodology or results of a study in response to pressure from a funding source	15.5	20.6	9.5 ***
Other behaviours			
11. Publishing the same data or results in two or more publications	4.7	5.9	3.4 **
12. Inappropriately assigning authorship credit	10.0	12.3	7.4 ***
13. Withholding details of methodology or results in papers or proposals	10.8	12.4	8.9 **
14. Using inadequate or inappropriate research designs	13.5	14.6	12.2
15. Dropping observations or data points from analyses based on a gut feeling that they were inaccurate	15.3	14.3	16.5
16. Inadequate record keeping related to research projects	27.5	27.7	27.3

p-values should not define a study

nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 520 > Issue 7549 > Comment > Article

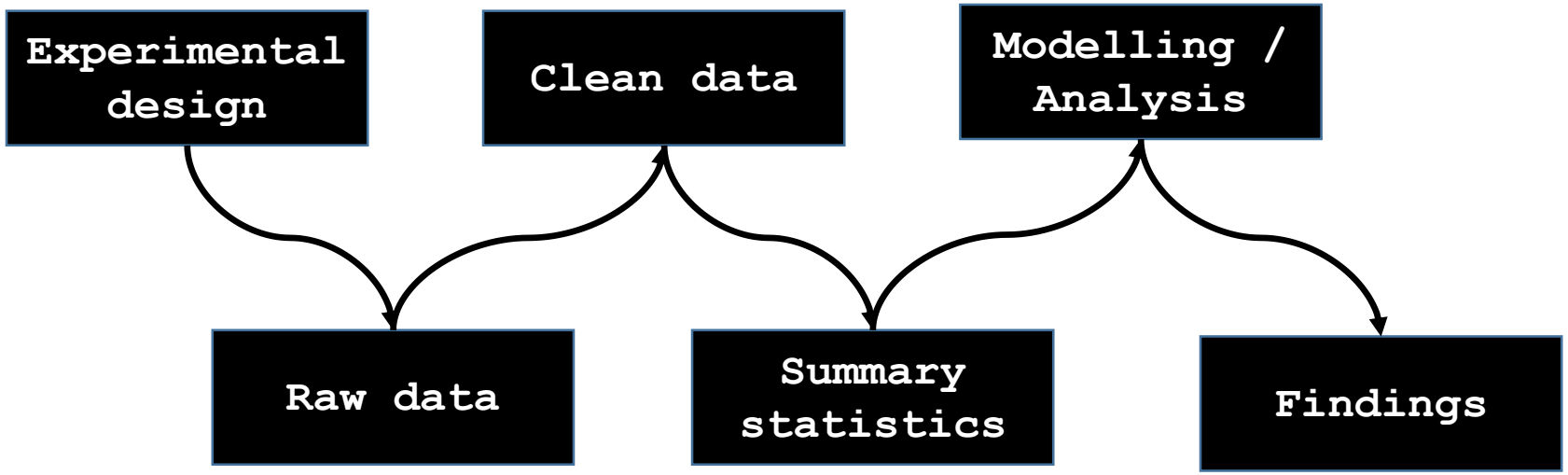
E-alert RSS Facebook Twitter

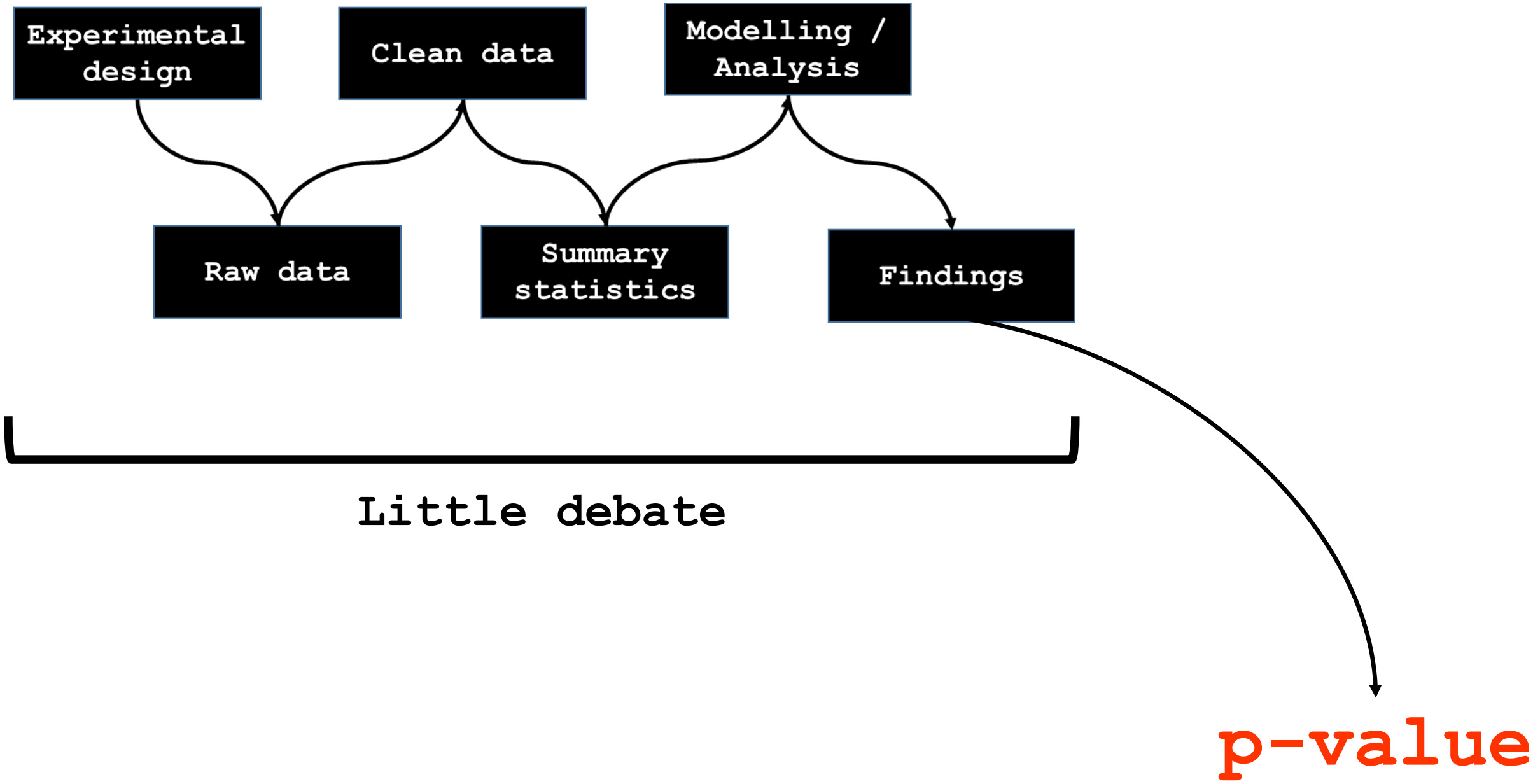
NATURE | COMMENT

Statistics: *P* values are just the tip of the iceberg

Jeffrey T. Leek & Roger D. Peng

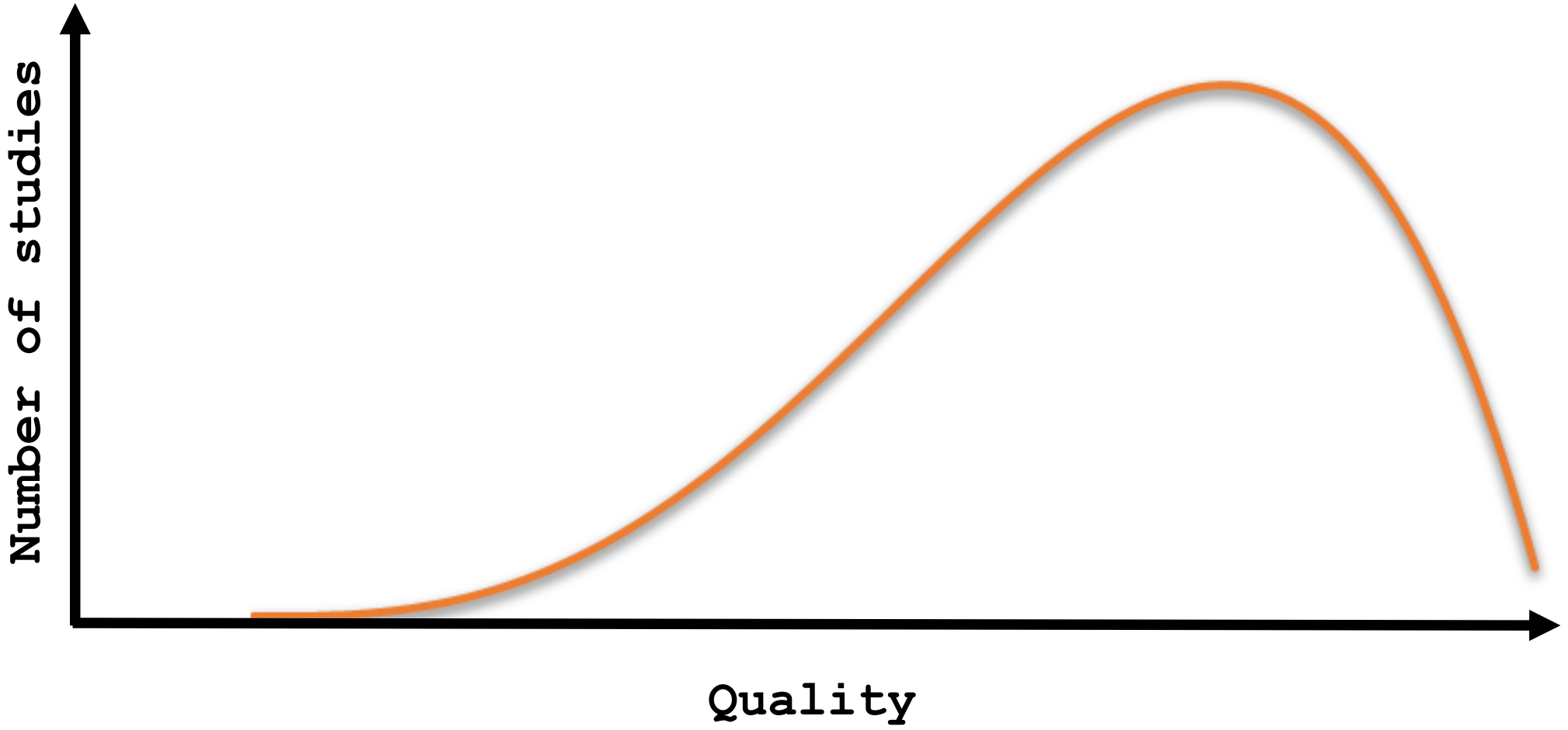
28 April 2015




Experimental design
Clean data
Modelling / Analysis
Raw data
Summary statistics
Findings
Little debate
p-value

Extreme scrutiny

Today



Past failings

Retraction Watch

Tracking retractions as a window into the scientific process

PAGES

Help us: Here's some of what we're working on

How you can support Retraction Watch

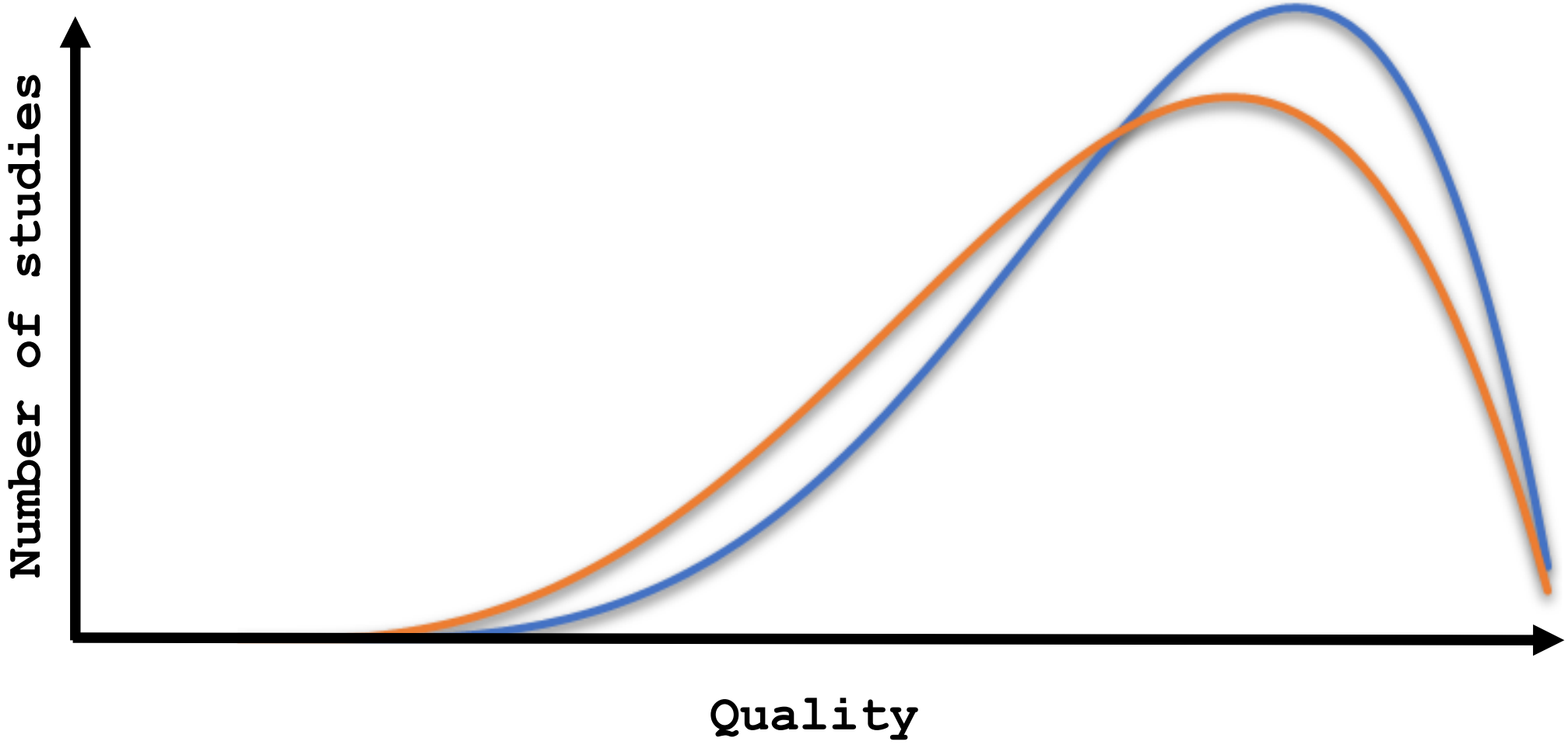
Meet the Retraction Watch staff

Diederik Stapel now has 58 retractions

Social psychologist Diederik Stapel has notched his 58th retraction, after admitting he fabricated data in yet another article.

He's holding onto his 4th place spot on our leaderboard.

Tomorrow



Who benefits most from reproducibility?

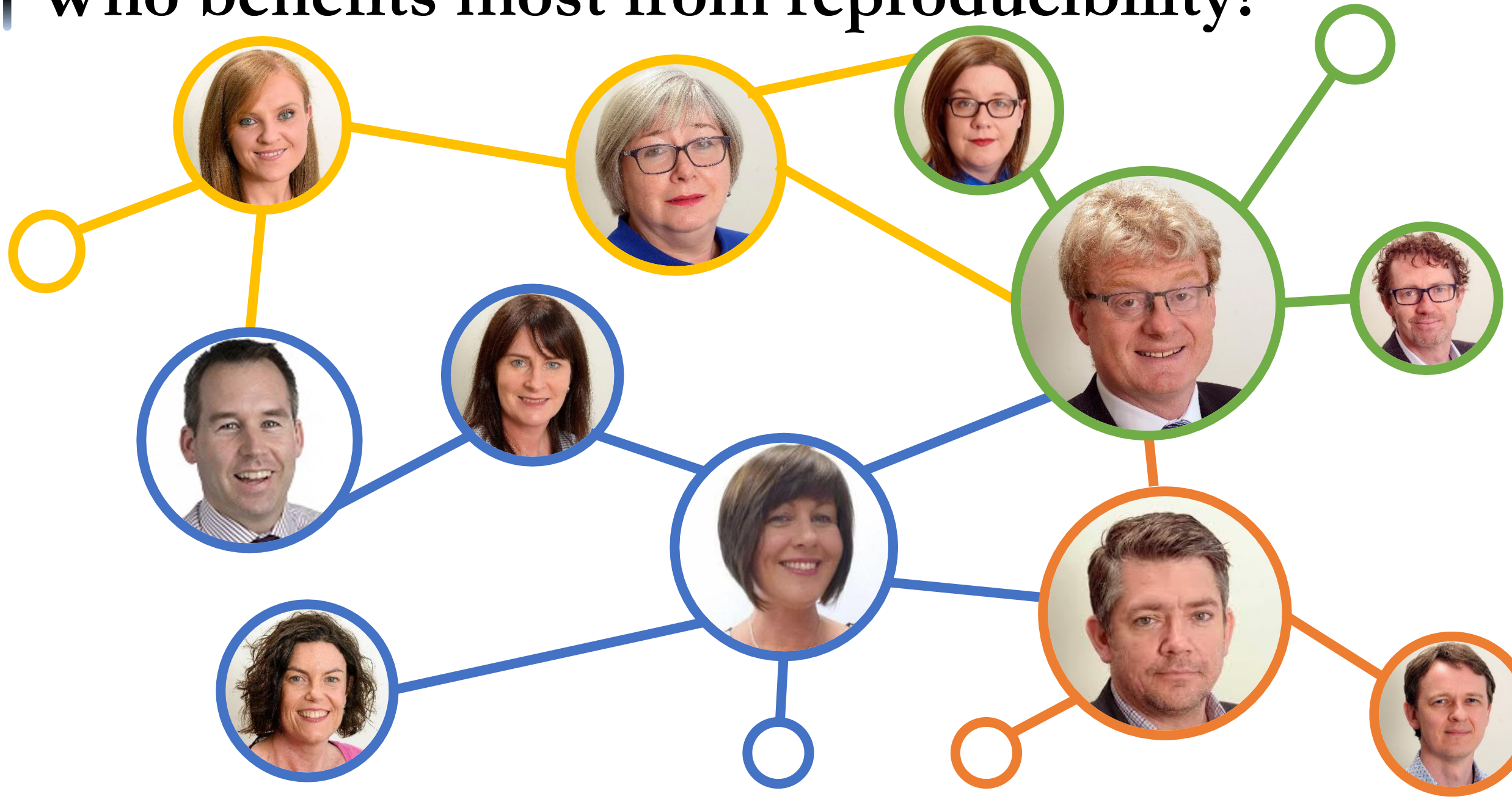
Casey Greene
@GreeneScientist

Follow

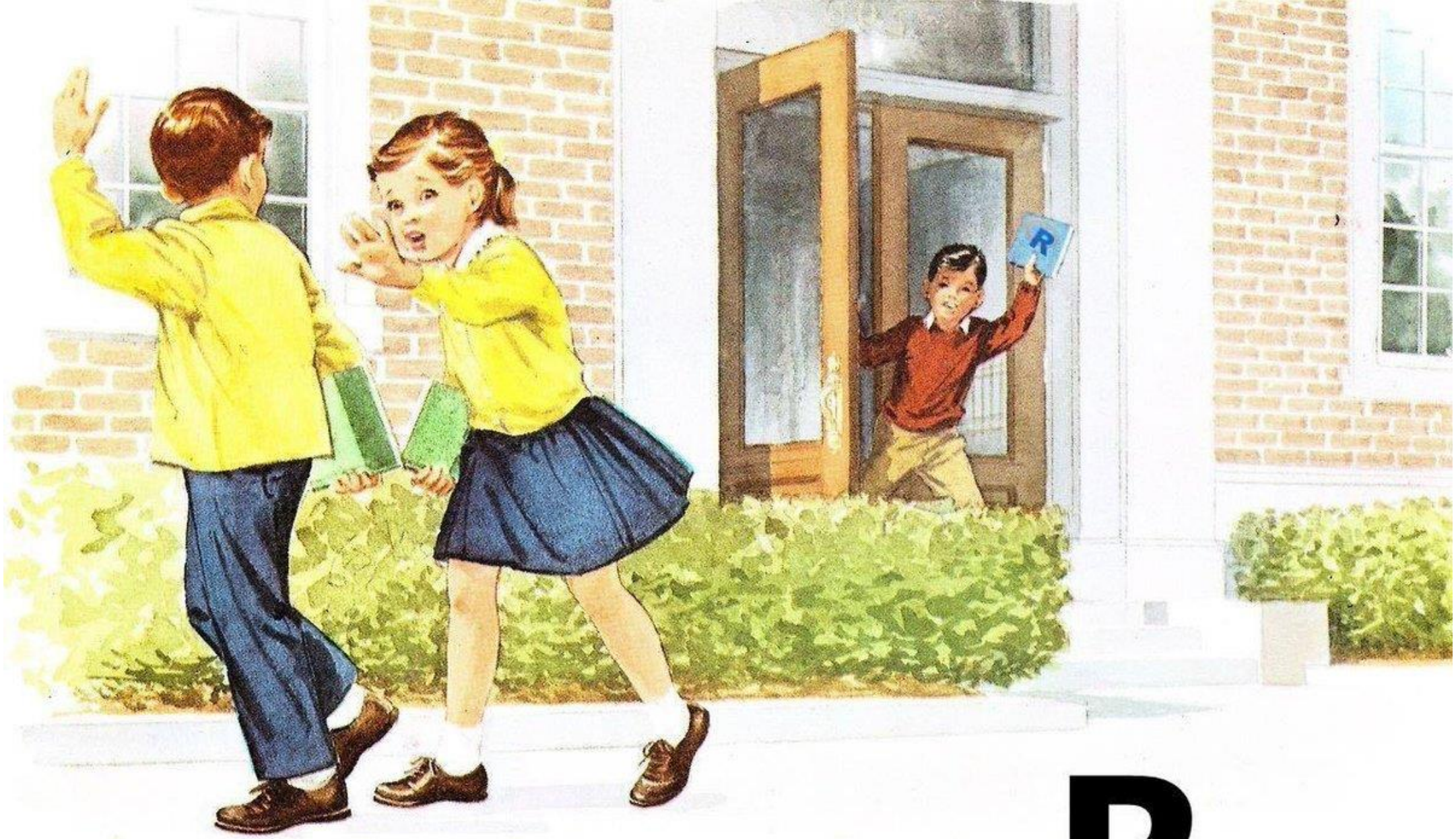
Reproducibility is important because the you of 3 months ago is terrible at answering email! - @tracykteal at #2016dssummit

1:17 PM - 26 Oct 2016 from Manhattan, NY

Who benefits most from reproducibility?



Where to begin…



Run, or he's going to tell us about **R** again!

R is for Resources

Vignettes


Package 'ggplot2'

December 30, 2016

Version 2.2.1

Title Create Elegant Data Visualisations Using the Grammar of Graphics

Description A system for 'declaratively' creating graphics, based on "The Grammar of Graphics". You provide the data, tell 'ggplot2' how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

Depends R (>= 3.1)

Imports digest, grid, gtable (>= 0.1.1), MASS, plyr (>= 1.7.1), reshape2, scales (>= 0.4.1), stats, tibble, lazyeval

Suggests covr, ggplot2movies, hexbin, Hmisc, lattice, mapproj, maps, maptools, mgcv, multcomp, nlme, testthat (>= 0.11.0), quantreg, knitr, rpart, rmarkdown, svglite

Enhances sp

License GPL-2 | file LICENSE

URL http://ggplot2.tidyverse.org, https://github.com/tidyverse/ggplot2

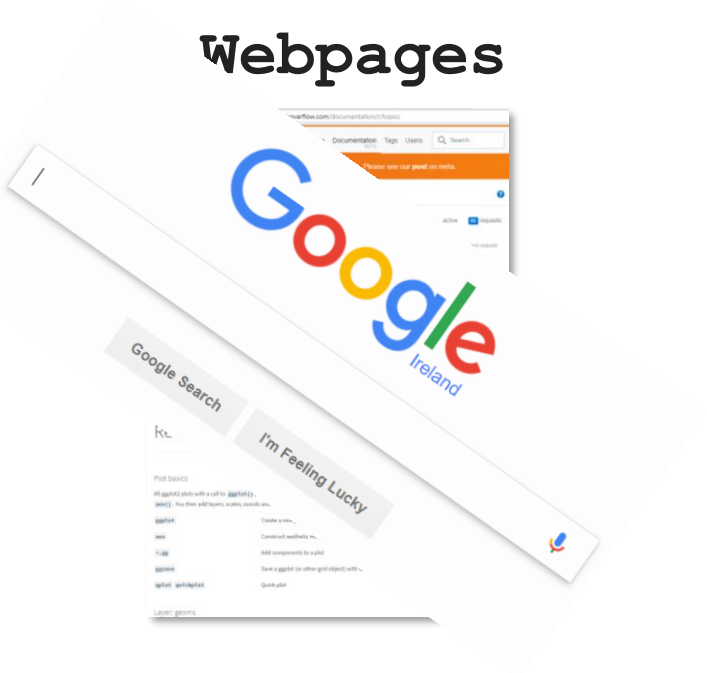
BugReports https://github.com/tidyverse/ggplot2/issues

LazyData true

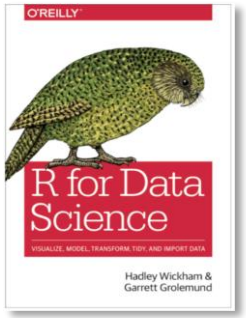
Collate 'ggproto.r' 'aaa-.r' 'aes-calculated.r' 'aes-colour-fill-alpha.r' 'aes-group-order.r' 'aes-linetype-size-shape.r' 'aes-position.r' 'utilities.r' 'aes.r' 'legend-draw.r' 'geom-.r' 'annotation-custom.r' 'annotation-logticks.r' 'geom-polygon.r' 'geom-map.r' 'annotation-map.r' 'geom-raster.r' 'annotation-raster.r' 'annotation.r' 'autolayer.r' 'axis-secondary.R' 'bench.r' 'bin.R' 'coord-.r' 'coord-cartesian-.r' 'coord-fixed.r' 'coord-flip.r' 'coord-map.r' 'coord-munch.r' 'coord-polar.r' 'coord-quickmap.R' 'coord-transform.r' 'data.R' 'facet-.r' 'facet-grid-.r' 'facet-null.r' 'facet-wrap.r' 'fortify-lm.r' 'fortify-map.r' 'fortify-multcomp.r' 'fortify-spatial.r' 'fortify.r' 'stat-.r' 'geom-abline.r' 'geom-rect.r' 'geom-bar.r' 'geom-bin2d.r' 'geom-blank.r' 'geom-boxplot.r' 'geom-col.r' 'geom-path.r' 'geom-contour.r' 'geom-count.r'

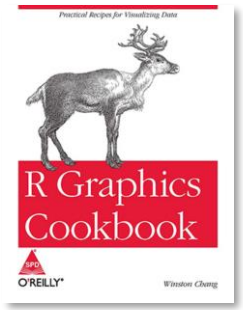
1

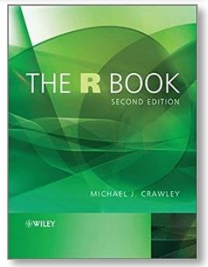
Webpages

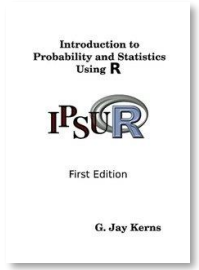


eBooks



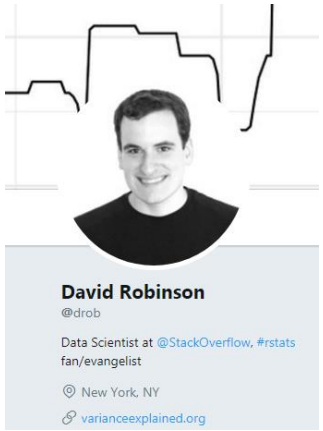




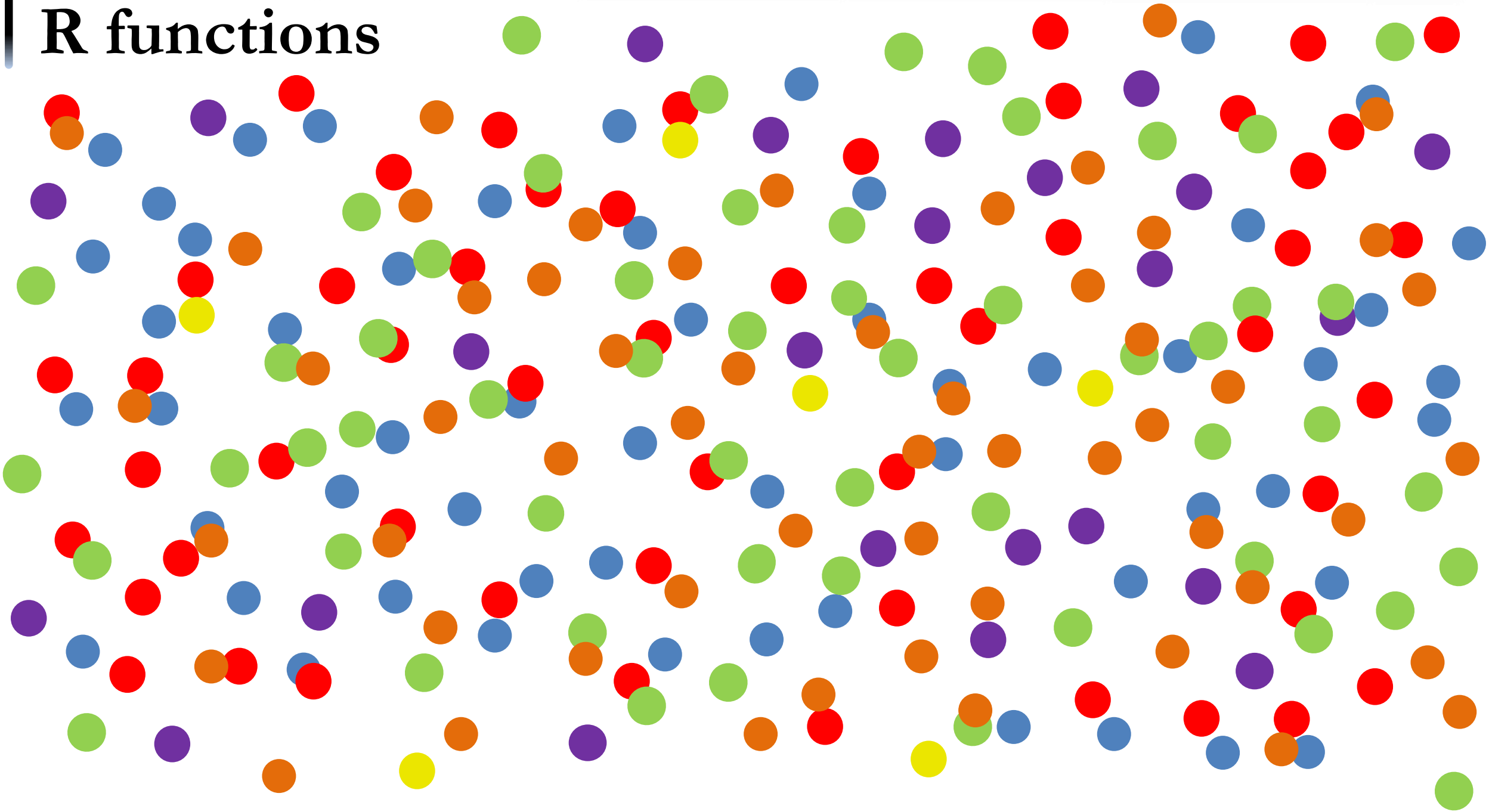


Cheatsheets

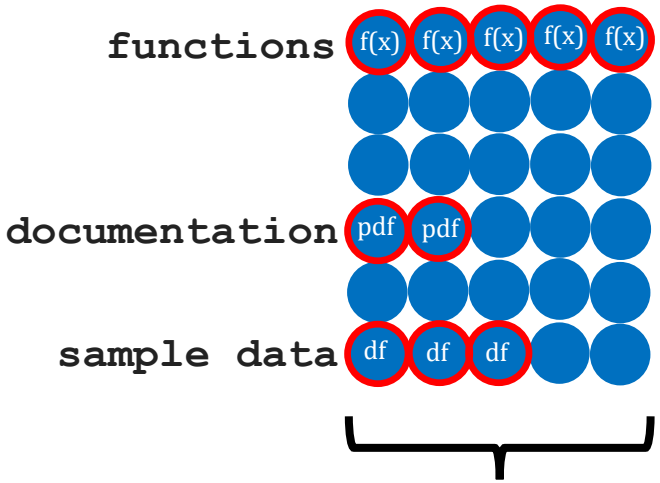
Twitter



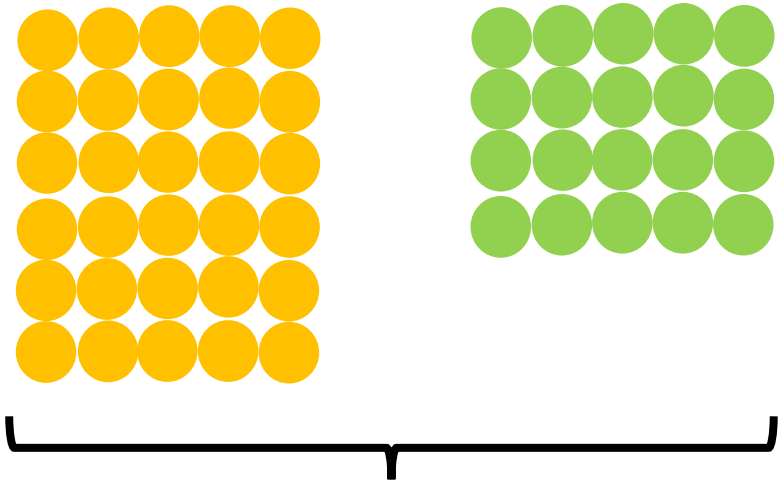
R functions



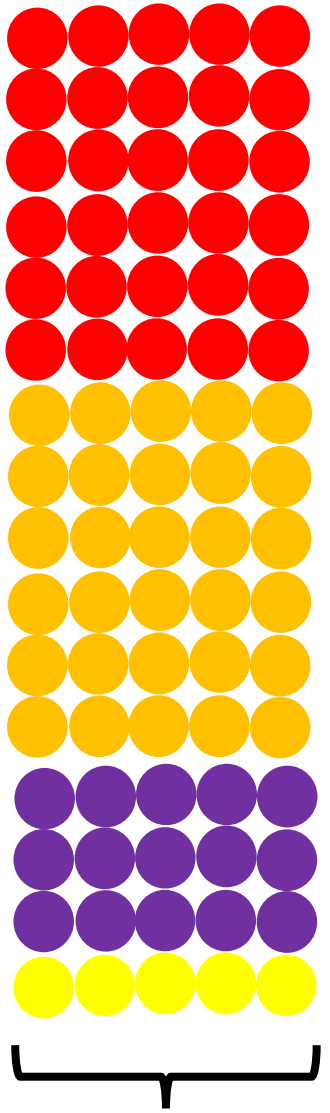
R packages



R comes pre-loaded with ~30 other packages (e.g. base, stats, graphics etc.)



Other packages:
Install once
Update regularly
Load each session



tidyverse

So many packages, so little time

https://cran.r-project.org/web/views/

CRAN Task Views

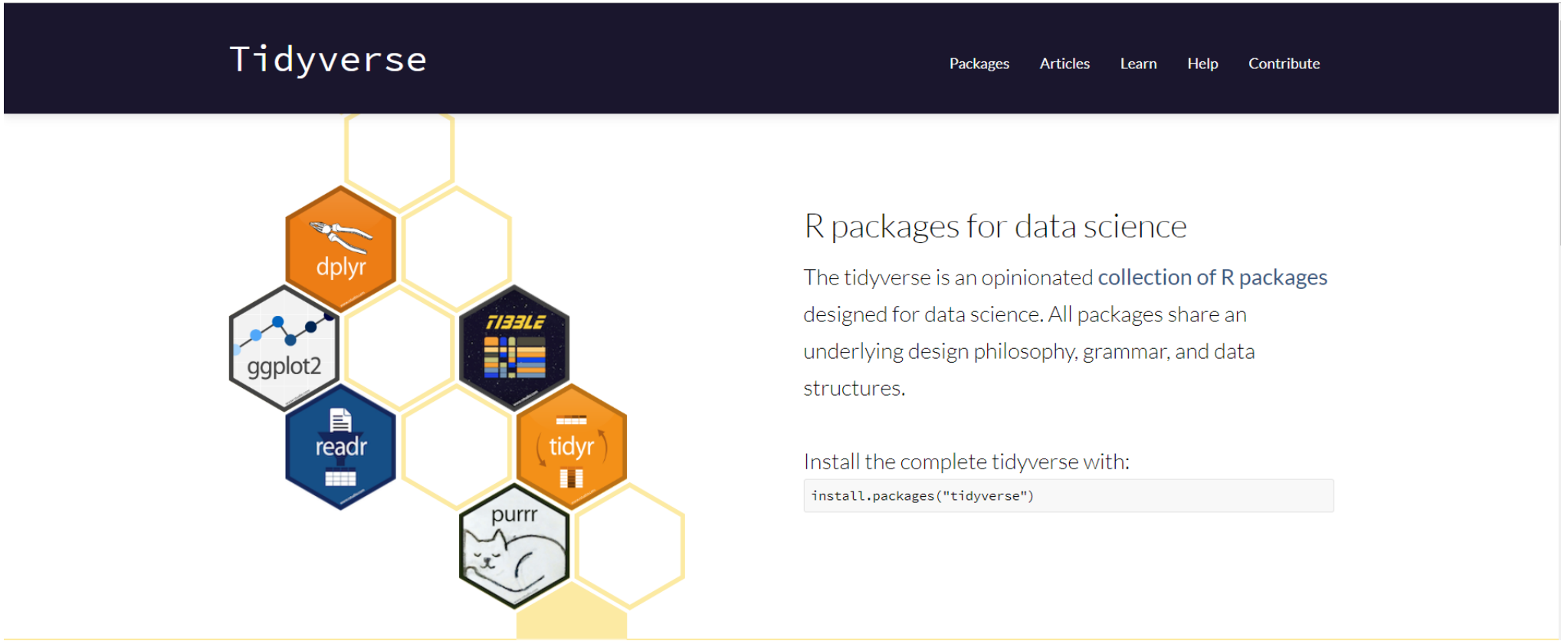
CRAN task views aim to provide some guidance which packages on CRAN are relevant for tasks related to a certain topic. They give a brief overview of the included packages and can be automatically installed using the ctv package. The views are intended to have a sharp focus so that it is sufficiently clear which packages should be included (or excluded) - and they are *not* meant to endorse the "best" packages for a given task.

- To automatically install the views, the ctv package needs to be installed, e.g., via
  `install.packages("ctv")`
  and then the views can be installed via `install.views` or `update.views` (where the latter only installs those packages are not installed and up-to-date), e.g.,
  `ctv::install.views("Econometrics")`
  `ctv::update.views("Econometrics")`
- The task views are maintained by volunteers. You can help them by suggesting packages that should be included in their task views. The contact e-mail addresses are listed on the individual task view pages.
- For general concerns regarding task views contact the ctv package maintainer.

Topics

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
Databases	Databases with R
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data

What is the tidyverse?



- Joined up collection of packages for data analysis
    - Consistent functions
    - Uses (tidy) data
    - Supports end-to-end workflows

What is the tidyverse?

```
> install.packages(c("broom", "cli2", "crayon",
"dbplyr", "dplyr", "forcats", "ggplot2", "haven",
"hms", "httr", "jsonlite", "lubridate",
"magrittr", "modelr", "pillar", "purrr", "readr",
"readxl", "reprex", "rlang", "rstudioapi",
"rvest", "stringr", "tibble", "tidyr", "xml2")
```

```
> install.packages("tidyverse")
```

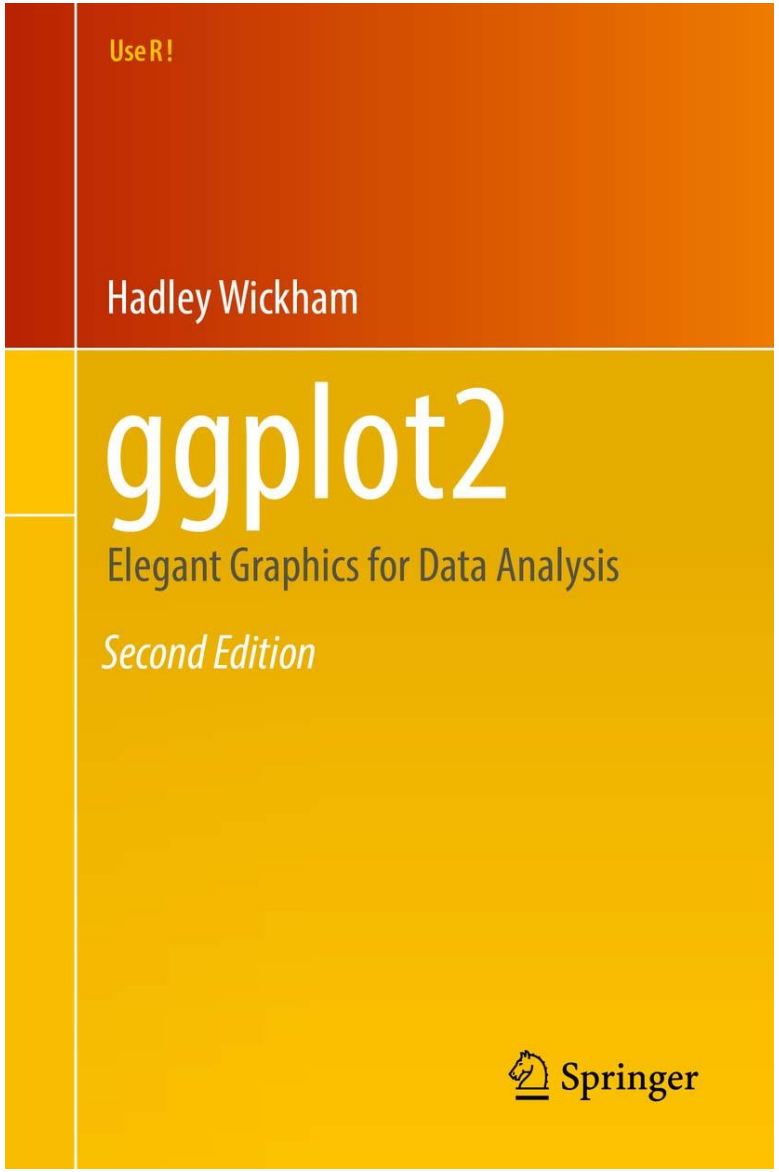
The tidyverse Oct 2017

```
> library(tidyverse)
Loading tidyverse: ggplot2
Loading tidyverse: tibble
Loading tidyverse: tidyr
Loading tidyverse: readr
Loading tidyverse: purrr
Loading tidyverse: dplyr
```

The tidyverse May 2019

```
> library(tidyverse)
-- Attaching packages ------------------------------------ tidyverse 1.2.1 --
v ggplot2 3.1.1       v purrr   0.3.2
v tibble  2.1.1       v dplyr   0.8.0.1
v tidyr   0.8.3       v stringr 1.4.0
v readr   1.3.1       v forcats 0.4.0
-- Conflicts --------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

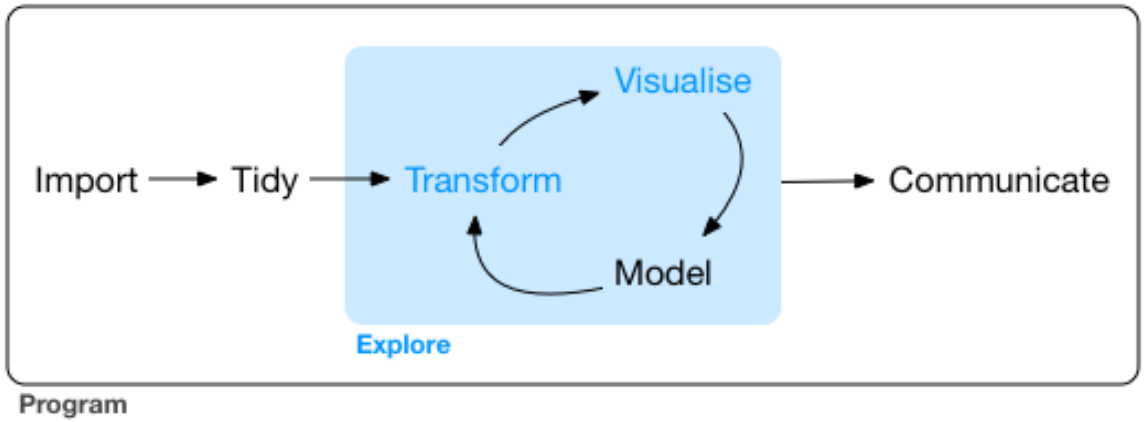
You could write a book on that!!



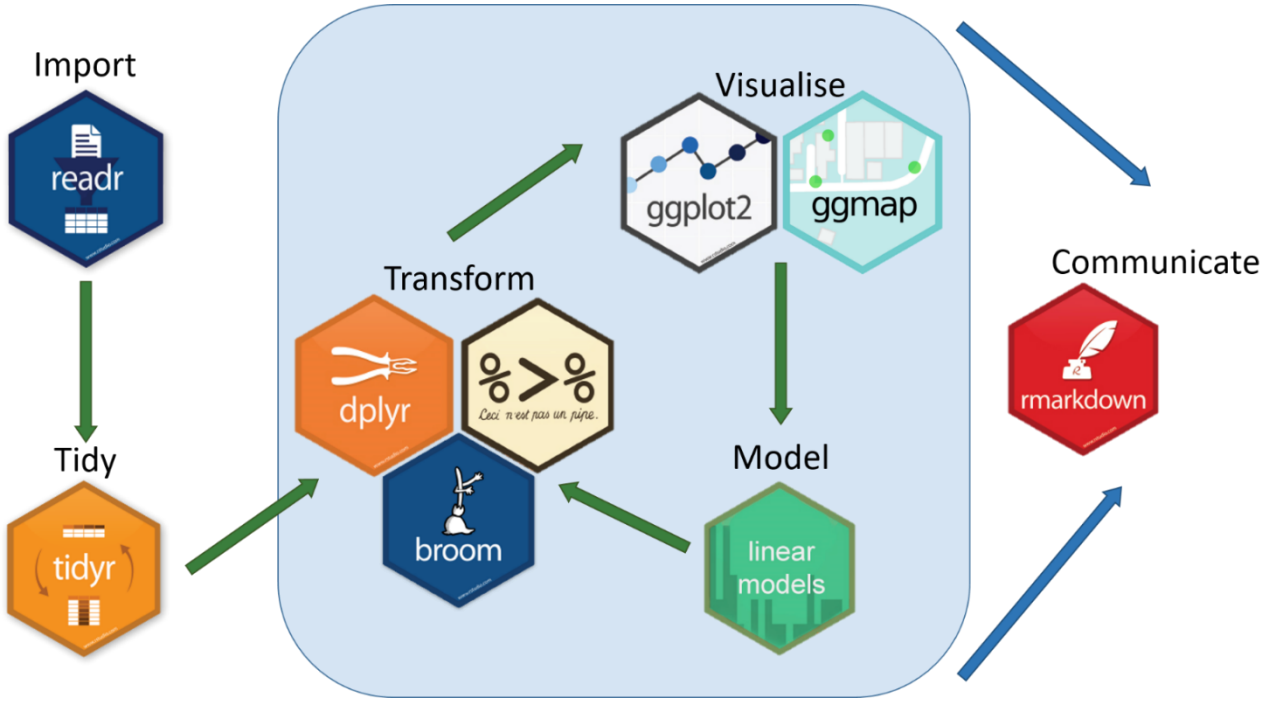
Data analysis in a nutshell

"The good news about computers is that they do what you tell them to do. The bad news is that they do what you tell them to do." - Ted Nelson

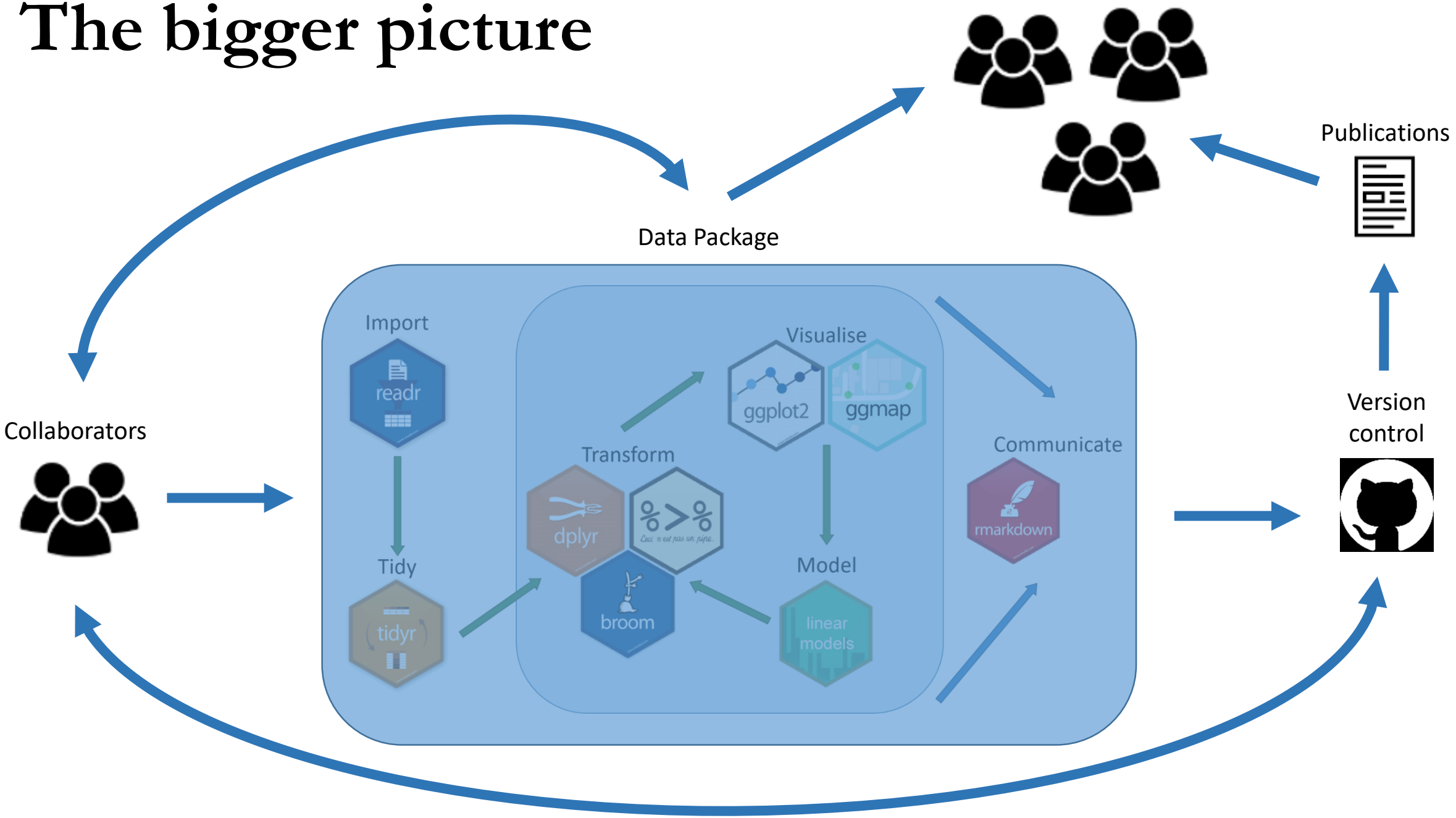
The tidyverse makes R code more human readable
- it is easier to write, run and read



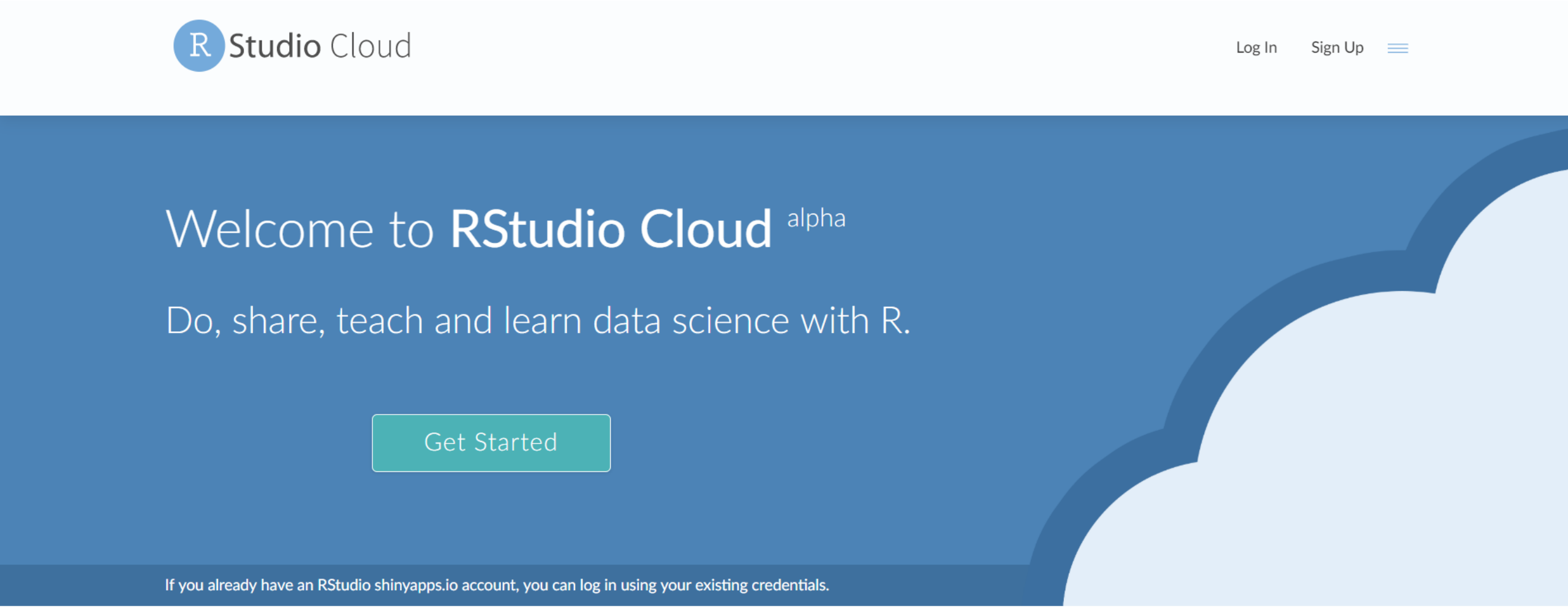
Putting the pieces together via the tidyverse



The bigger picture



RStudio Cloud



GitHub

Search or jump to... / Pull requests Issues Marketplace Explore

Set status

Brendan Palmer

bapalmer

Edit profile

Twitter: @B_A_Palmer

https://crfcsdau.github.io/about/

Organizations

Overview | Repositories 14 | Projects 0 | Stars 0 | Followers 10 | Following 10

Find a repository... Type: All Language: All New

R-A_Hitchhikers_Guide_to_Reproducible_Research ★ Star

A 3-day R course given in University College Cork that encompasses RStudio projects, an introduction to the R tidyverse language, project workflows and reporting using R Markdown.

MIT License Updated 6 days ago

lunchtime_sessions ★ Star

Short 1 hour introductions to R-related topics such as creating R projects, using GitHub through RStudio and more

HTML ★ 1 MIT License Updated 12 days ago

RCR ★ Star

Section of the UCC Reproducible Conduct of Research digital badge dedicated to exposing researchers to reproducible research practices.

HTML MIT License Updated 23 days ago

Twitter

UK Reproducibility Network
@ukrepro

UK Reproducibility Network: a peer-led consortium to investigate factors which contribute to robust research, provide training, and disseminate best practice.

Retraction Watch
@RetractionWatch

Tracking retractions as a window into the scientific process. Sign up for our daily newsletter: eepurl.com/bNRlUn Tips? team@retractionwatch.com

James Heathers
@jamesheathers

Research scientist. Biosignals, meta-science, error detection. Yelling has vitamins. Cohost of @hertzpodcast.

The warm embrace of Uncle Sam

jamesheathers.com

Brian Nosek
@BrianNosek

Executive Director @ Center for Open Science, Professor @ University of Virginia, and co-Founder of Project Implicit

Open Science MOOC
@OpenScienceMOOC Follows you

A community designed for students and researchers to help make 'Open' the default setting for the future of research. Slack: osmooc.herokuapp.com

Everywhere

Malcolm Macleod #FBPE
@Maclomaclee Follows you

clinical neurologist, stroke trialist, and interested in improving the quality of laboratory research

Kate Button
@ButtonKate Follows you

Academic. Psychologist. Cognitive mechanisms of depression & anxiety. Meta-science & scientific rigour. Sporadic Twitterer.

Jon Tennant
@Protohedgehog

Rogue Palaeontologist currently at @criparis. I do @paleorXiv, @OpenScienceMOOC, @EGU_GC & @PLOSpaleo. Tweets snarky, irresponsible and my own. Do epic shit.

Jurassic Park

Workshop content