Tools for data processing: R

Pieter Provoost

IOC-UNESCO

OBIS OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

United Nations Educational, Scientific and Cultural Organization

Intergovernmental Oceanographic Commission

Tools: R

- Open source environment for statistical computing and graphics
- Created in 1993 at the University of Auckland, inspired by S
- Rapidly gaining popularity for machine learning and business intelligence

FAST COMPANY

7 MINUTE READ | TECHNOLOGY

Why The R Programming Language Is Good For Business

InfoWorld

Home > Business Intelligence > Analytics

Microsoft Azure welcomes R language, with more to come

DataInformed
Big Data and Analytics in the Enterprise

The Rise of R Analytics Programming Language for Business

Tools: R

- Reproducible research!
  - https://rpubs.com/marschmi/105639

PERSPECTIVE

The Economics of Reproducibility in Preclinical Research

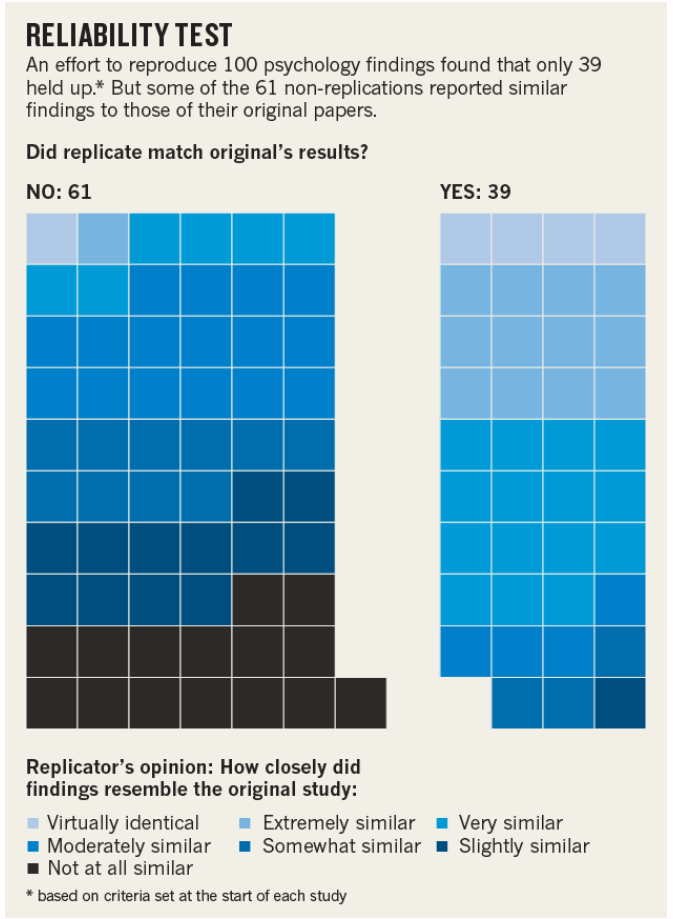
Leonard P. Freedman[1]*, Iain M. Cockburn[2], Timothy S. Simcoe[2,3]

1 Global Biological Standards Institute, Washington, D.C., United States of America, **2** Boston University School of Management, Boston, Massachusetts, United States of America, **3** Council of Economic Advisers, Washington, D.C., United States of America

* lfreedman@gbsi.org

Abstract

Low reproducibility rates within life science research undermine cumulative knowledge production and contribute to both delays and costs of therapeutic drug development. An analysis of past studies indicates that the cumulative (total) prevalence of irreproducible preclinical research exceeds 50%, resulting in approximately US$28,000,000,000 (US $28B)/year spent on preclinical research that is not reproducible—in the United States alone. We outline a framework for solutions and a plan for long-term improvements in reproducibility rates that will help to accelerate the discovery of life-saving therapies and cures.



Code share

Papers in Nature journals should make computer code accessible where possible.

A theme in *Nature*'s ongoing campaign for the replicability and reproducibility of our research papers is that key components of publications should be available to peers who wish to validate the techniques and results.

A core element of many papers is the computer code used by authors in models, simulations and data analysis. In an ideal world, this code would always be transportable and easily used by others. In such a world, our editorial policy would be to insist on sharing to allow free use, as we already do (as far as is practicable) with data and research materials. Unfortunately, such an ideal is not easy to attain owing to the amount of extra funding and effort it would require to render some major pieces of code shareable. Nevertheless, we at *Nature* and the Nature research journals want to encourage as much sharing as possible.

Climate modellers have made some strides in this regard. The journal *Geoscientific Model Development* has a good example of such a policy (see go.nature.com/jv8g1w), and an article in *Nature Geoscience* discusses some of the opportunities presented by code sharing, as well as

Tools: R

- The R ecosystem
  - R Core Team, http://r-project.org
  - Packages
    - CRAN (The Comprehensive R Archive Network): 8600 packages
    - Bioconductor
    - rOpenSci
    - Omegahat
    - GitHub
  - R User Groups
  - Enterprise
    - Revolution Analytics (Microsoft R Server)
    - RStudio: RStudio IDE, Shiny (web applications)

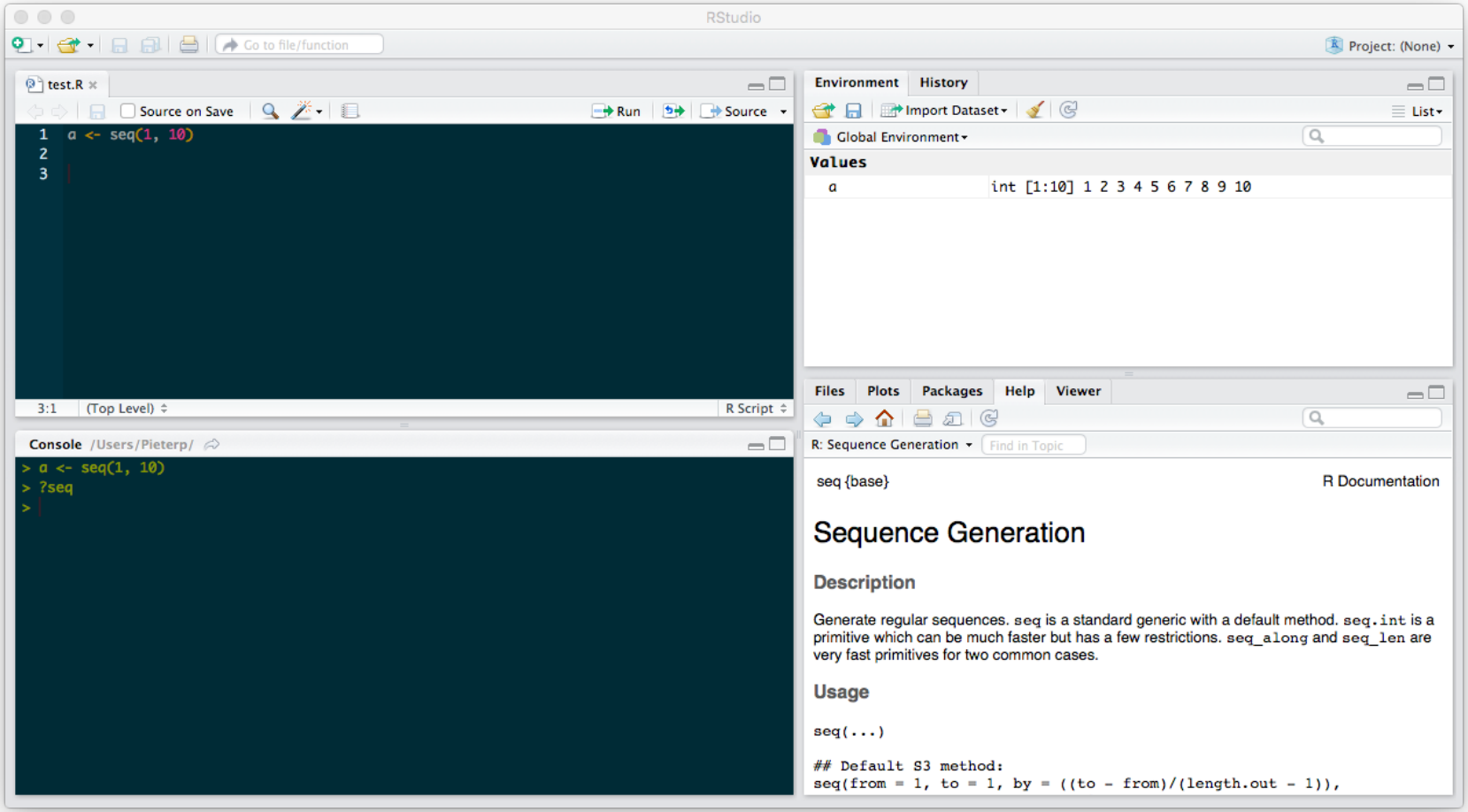
Tools: R

- Introduction to R
  - http://beta.iobis.org/manual/intror/
  - http://rforcats.net/ (Scott Chamberlain)
  - http://swirlstats.com/

- Installation
  - Install R from https://cloud.r-project.org/
  - Install RStudio from https://www.rstudio.com/products/rstudio/download/

Tools: R

- RStudio



Tools: R

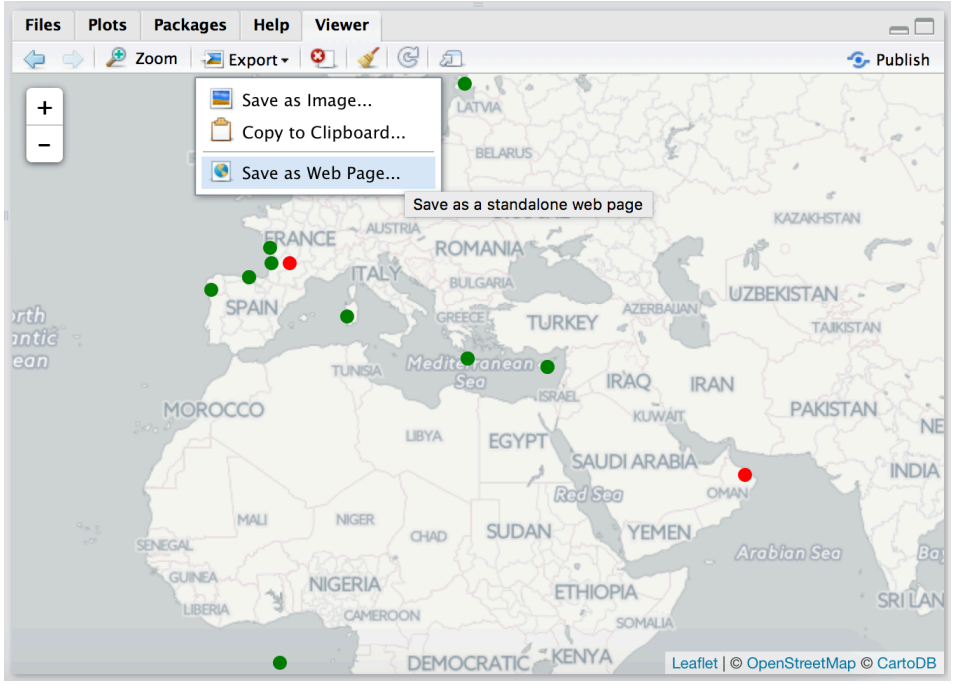
- Demo 1: introduction
  - https://github.com/iobis/training/tree/master/fixo3
  - OBIS R package
    - https://github.com/iobis/robis
    - Installation

```
install.packages("devtools")
devtools::install_github("iobis/robis")
```

  - ggplot
    - http://www.r-graph-gallery.com/all-graphs/
    - http://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html
    - https://github.com/jrnold/ggthemes

Tools: R

- Demo 2: cleaning data
  - leaflet
    - https://rstudio.github.io/leaflet/
    - export map



Tools: R

- Demo 3: rOpenSci packages
  - https://ropensci.org
  - packages
    - **rfishbase**: interface to fishbase.org
    - **rgbif**: interface to GBIF
    - **mapr**: species occurrence maps
    - **spocc**: species occurrence toolkit
    - **rredlist**: interface to IUCN Red List
    - **taxize**: search taxonomic data sources
    - **finch**: read DwC-A files

Tools: R

- Demo 4: data analysis
  - reshape
    - http://www.statmethods.net/management/reshape.html
  - vegan
    - http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pd