

# **Computational Reproducibility:** from theory to practice



Anna Krystalli  
@annakrystalli

useR!2020 EuropeanHub  
gone remote!





# me: Dr Anna Krystalli

- Research Software Engineer, University of Sheffield
  - twitter [@annakrystalli](https://twitter.com/annakrystalli)
  - github [@annakrystalli](https://github.com/annakrystalli)
  - email [a.krystalli\[at\]sheffield.ac.uk](mailto:a.krystalli@sheffield.ac.uk)
- Editor rOpenSci
- Co-organiser: Sheffield R Users group

# Background

# Marine Biology



Boats are fun!

# Marine Biology

So are maps!

# Quality Assurance

QA Auditor for a Contract Research Organisation subject to GLP regulation

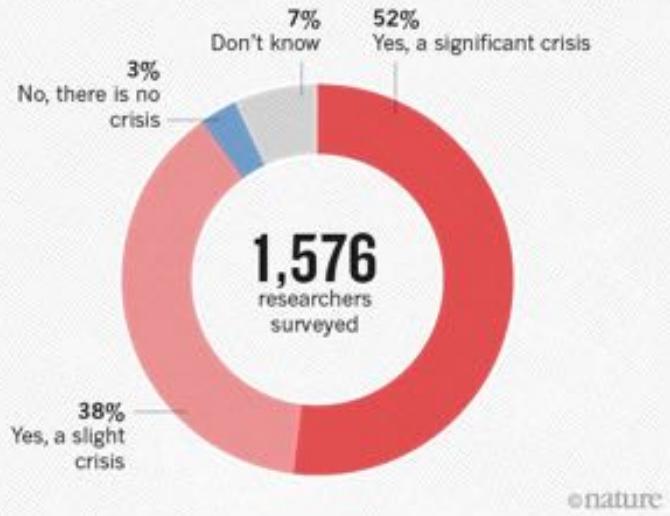
# Ultrasport

Brand coordinator for an extreme sports equipment distributor

# **Back to science:**

# Ooops, that's embarrassing!

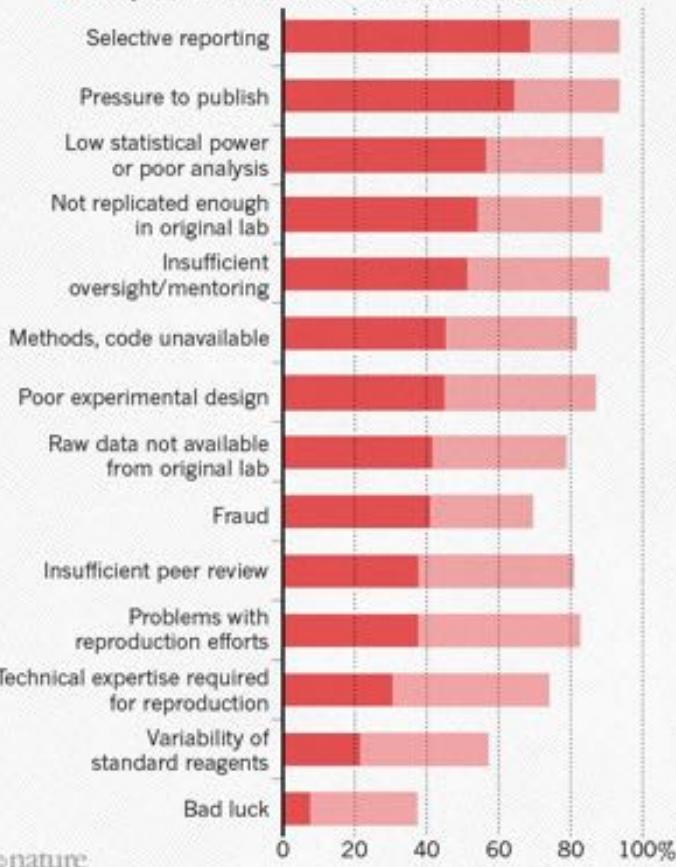
## IS THERE A REPRODUCIBILITY CRISIS?



## WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.

● Always/often contribute   ● Sometimes contribute



# The paper is the advertisement

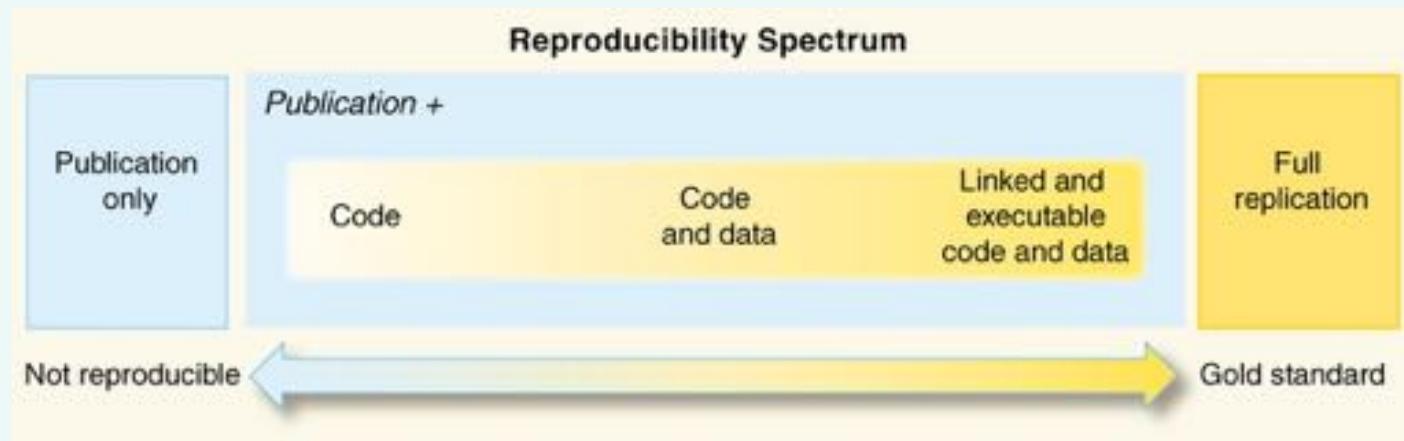
“an article about a computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.”

John Claerbout paraphrased in *Buckheit and Donoho (1995)*

**Why is our whole system geared towards reviewing, publishing, distributing, archiving the advertisement?**

# Progress: calls for reproducibility as minimum standard

Reproducibility has the potential to serve as a minimum standard for judging scientific claims when full independent replication of a study is not possible.



Reproducible Research in Computational Science ROGER D. PENG, SCIENCE 02 DEC 2011: 1226-1227

## **Benefit #1**

**transparency as a means of  
verification**

**There is a hidden superpower...**



# Forking github/training-kit

It should only take a few seconds.

**Woah, It's  
evolution...**

⟳ Refresh



# Macroecological and macroevolutionary patterns emerge in the universe of GNU/Linux operating systems

doi:10.1111/ecog.03424



## Research

### Macroecological and macroevolutionary patterns emerge in the universe of GNU/Linux operating systems

Petr Keil<sup>1</sup>, A. A. M. MacDonald<sup>2</sup>, Kelly S. Ramirez<sup>3</sup>, Joanne M. Bennett<sup>4</sup>, Gabriel E. García-Peña<sup>5</sup>, Benjamin Yguel<sup>6</sup>, Bérenger Bourgeois<sup>7</sup> and Carsten Meyer<sup>8</sup>

<sup>1</sup> Petr Keil (keil@uni-leipzig.de) – 0000-0003-1952 (petrkeil@uni-leipzig.de); German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany; <sup>2</sup> MBI also as: Inst. of Biology, Marine Science Univ. Halle-Wittenberg, Halle (Saale), Germany; <sup>3</sup> A. A. M. MacDonald, G. E. García-Peña and B. Bourgeois, Centre de Synthèse et d'Analyse sur la Biodiversité – CESAR, Aix-en-Provence, France; <sup>4</sup> ANMIM also as: Univ. of British Columbia, Vancouver, BC, Canada; <sup>5</sup> GEG- also as: Centro de Ciencias de la Complejidad (CCS) y Facultad de Medicina Veterinaria y Zootecnia, Mexico; <sup>6</sup> Bébén Bourgeois, Institut National de la Recherche Agronomique (INRA), Unité Biogéoquímique Comté, Dijon, France; <sup>7</sup> S. & B. Bourgeois, Tiersantierklinik für Kleintiere, Berlin, Germany; <sup>8</sup> B. Iglesias, Unité MÉTACOM, Institut des Sciences de l'Evolution, UMR 7179 CNRS-MNHN, Banyuls, France; and <sup>9</sup> C. Meyer, Macroecology and Society, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany; Faculty of Bioresources, Pharmacy and Psychology, Univ. of Leipzig, Leipzig, Germany; and Dept of Ecology and Evolutionary Biology, Yale Univ., New Haven, CT, USA.

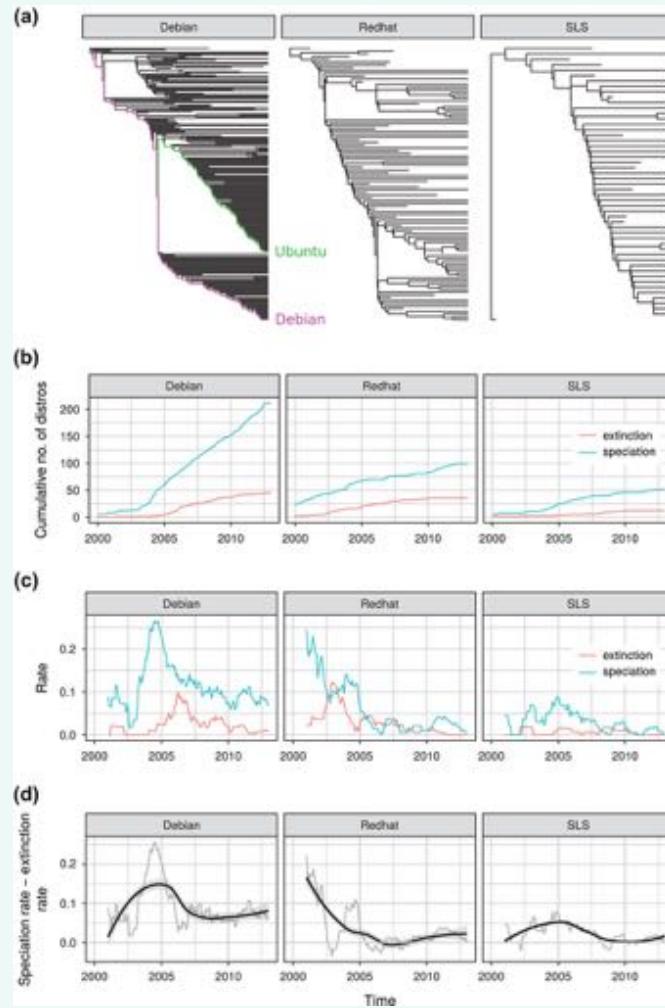
Ecography  
41: 1788–1800, 2018  
doi:10.1111/ecog.03424

Subject Editor: Thiago Rangel  
Editor-in-Chief: Miguel Araújo  
Accepted 2 February 2018



What leads to classically recognized patterns of biodiversity remains an open and contested question. It remains unknown if observed patterns are generated by biological or non-biological mechanisms, or if we should expect the patterns to emerge in non-biological systems. Here, we employ analogies between GNU/Linux operating systems (distros), a non-biological system, and biodiversity, and we look for a number of well-established ecological and evolutionary patterns in the Linux universe. We demonstrate that patterns in the Linux universe generally match macroecological patterns. Particularly, Linux distro composition and rarefaction curves follow a skewed distribution with a clear excess of rare species, as is observed in a particular scaling of phylogenetic fluctuations; however, there is only a weak relationship between niche breadth (number of software packages) and comunesity. The diversity in the Linux universe also follows general macroevolutionary patterns: the number of phylogenetic lineages increases linearly through time, with clear per-species diversification and extinction slowdowns, something that has been indirectly estimated, but not directly observed in biology. Moreover, the composition of functional traits (software packages) exhibits significant phylogenetic signal. The emergence of macroecological patterns across Linux suggests that the patterns are produced independently of system identity, which points to the possibility of non-biological drivers of fundamental biodiversity patterns. At the same time, our study provides a step towards using Linux as a model system for exploring macroecological and macroevolutionary patterns.

Keywords: complex systems, cultural evolution, Debian, speciation, species-abundance distribution, Taylor's power law



# Example: gapminder.org: 2006

liberating stories from data

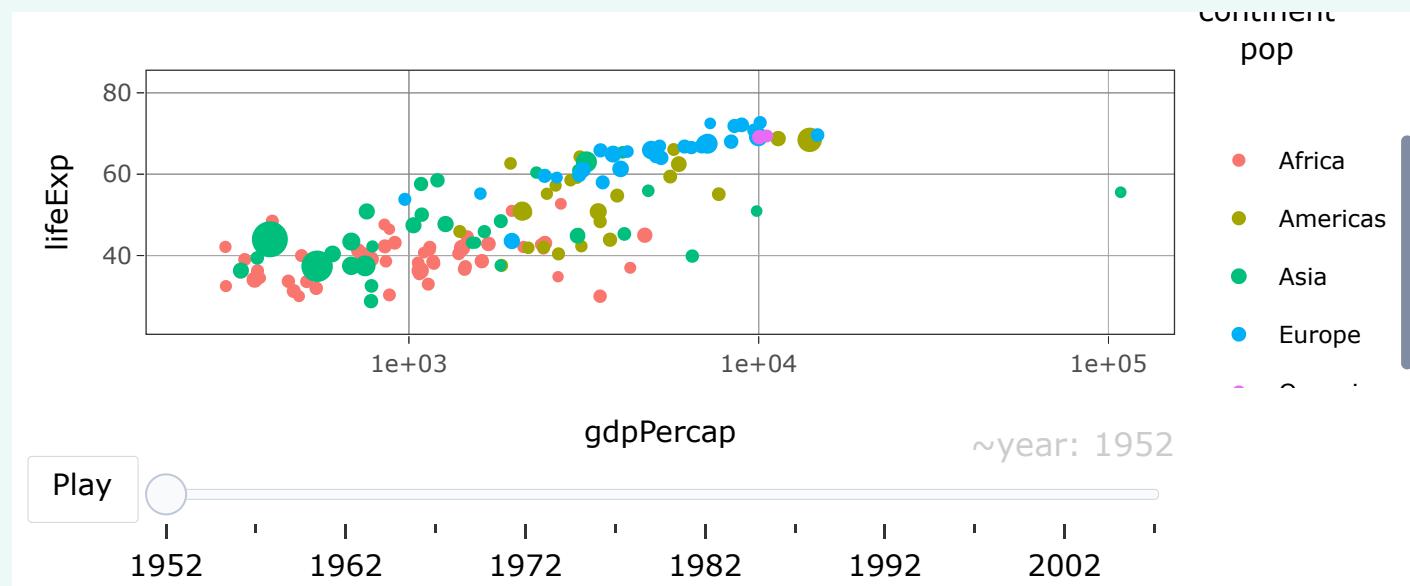


# Gapminder today

```
library(ggplot2)
```

```
p <- ggplot(gapminder::gapminder, aes(gdpPercap, lifeExp, size = pop, color = continent)) +  
  geom_point() + scale_x_log10() + theme_bw()
```

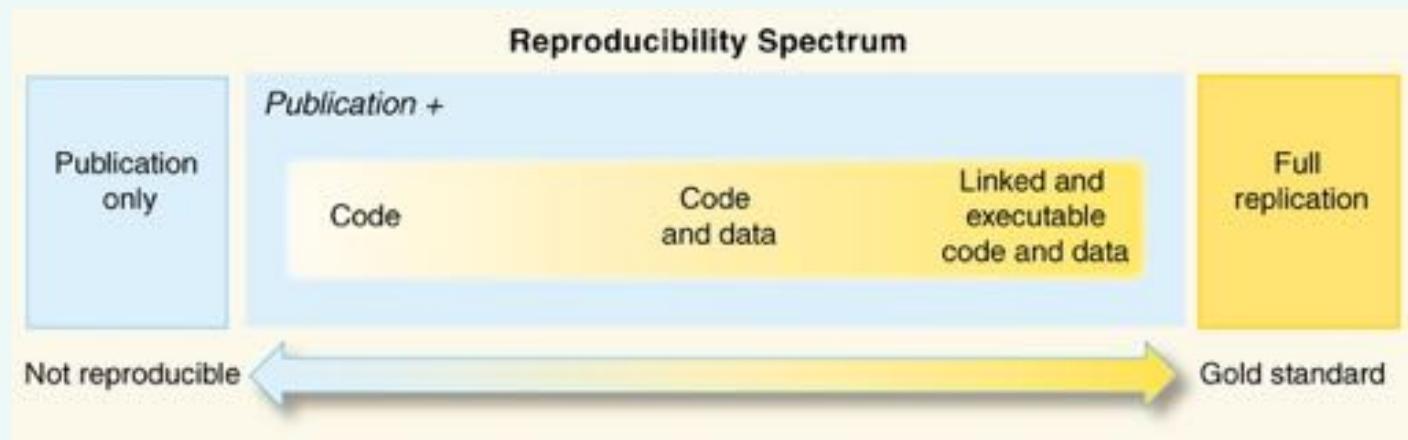
```
plotly::ggplotly(p)
```



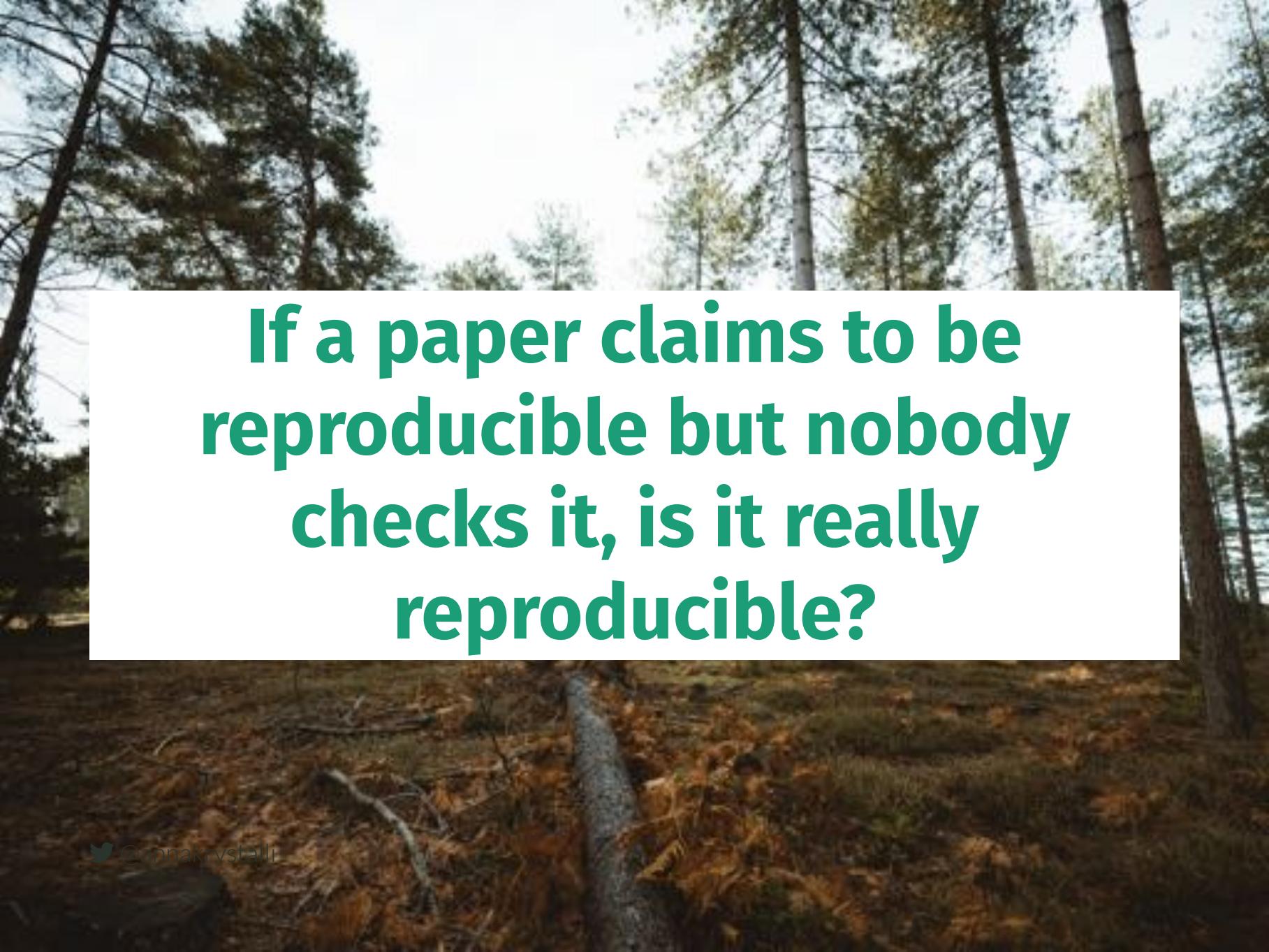
## **Benefit #2**

**transparency as a means of  
supercharging research cycle**

# So how are we doing?



Reproducible Research in Computational Science ROGER D. PENG, SCIENCE 02 DEC 2011 : 1226-1227



**If a paper claims to be  
reproducible but nobody  
checks it, is it really  
reproducible?**



# Practice

---

**The less you do, the more you s\*\*k**

# Reprohack

One day reproducibility hackathons

---

- **How reproducible are papers?**
- **How can we practice reproducibility?**

# ReproHack History

OpenCon Satellite: Berlin, 2016

OpenCon Satellite: London, 2017

Inspired by Owen Petchey's [Reproducible Research in Ecology, Evolution, Behaviour, and Environmental Studies](#) course,

- Reproduce published results from raw data
- Over a few months and a number of sessions

**ReproHack mission: Reproduce paper in a day  
from code and data**

# Software Sustainability Institute Fellowship 2019



🐦 @annakrystalli



**CoMcr19**  
CarpentryConnect Manchester 2019

# Reprohack Core Team formation



# ReproHackNL! - *Leiden*

 ReproHack 

This thing is really happening!

11:48 AM · Sep 20, 2019   
@annakrystalli

 16  See ReproHack 's other T...



# N8 CIR ReproHack Series!

<https://n8cir.org.uk/news/reprohacks/>



annakrystalli  
@annakrystalli



Manchester @N8CIR  
@ReproHack CANCELLED

Due to the changing situation with #COVID19 we've decided that despite the low risk, it would be irresponsible to expose participants unnecessarily to it.

We'll either reschedule or explore a remote option.

7:19 AM · Mar 12, 2020 from Sheffield, England



19 See the latest COVID-19 infor...

@annakrystalli



Much of the team made it!

People from afar were able to join:

including Japan, Argentina, Netherlands, Sweden and the USA!

# **How does it work?**

# Call for papers

annakrystalli  
@annakrystalli

Do you champion #reproducible #research?

Do you have a reproducible paper with open code and data?

The @SoftwareSaved #ReproHack series needs you!

Help others learn & engage with your work by submitting it to our 1-day Reproducibility hackathons!

sheffield-university.shinyapps.io/ReproHack\_CC Mc...

12:27 PM · Jun 12, 2019

@annakrystalli 86 69 people are Tweeting abou...

PROPOSE

Nominate a paper for Reproduction:

We invite nominations for papers that have both associated code and data publicly available. We also encourage analyses based on open source tools as we cannot guarantee participants will have access to specialized licensed software.

[Nominate Paper](#)

Proposed papers:

1. Spatial modelling of rice yield losses in Tanzania due to bacterial leaf blight and leaf blast in a changing climate  
Santos, A., H. Sparks, S. J. Zwart, Climate Change 135(3–4):2016, 369–383. Springer Nature doi: 10.1007/s10584-016-1380-2  
submitted by Adam Sparks

Why should we attempt to reproduce this paper?  
This was my first attempt at making a paper fully reproducible. To date I'd say the most reproducible that I have published. I'm interested to know what stumbling blocks exist that I'm not aware of (aside from needing software like ArcGIS to fully render the complex analysis).

Paper URL: <https://link.springer.com/article/10.1007/s10584-016-1380-2>  
Data URL: <https://zenodo.org/record/340801>  
Code URL: <https://github.com/AdamSparks/MCCRCODE>  
Used programming skills: R, Python, ArcGIS

2. Climate change may have limited effect on global risk of potato late blight.  
Sparks, A., H. Sparks, G. A. Higgins, R. J. A. Gareth & K. A. (2014). Climate change may have limited effect on global risk of potato late blight. Global Change Biology doi:10.1111/gcb.12596  
submitted by Adam Sparks

Why should we attempt to reproduce this paper?  
This is a clear one. The repository contains code for companion papers, the model development and the model implementation and analysis. As the repository notes, some data are not freely available so I've made an effort to allow the paper to be replicated as best possible with what's available.

Paper URL: <https://onlinelibrary.wiley.com/doi/10.1111/gcb.12596>  
Data URL: <https://zenodo.org/record/340802>  
Code URL: [https://github.com/AdamSparks/Global\\_Late\\_Blight\\_Modeling](https://github.com/AdamSparks/Global_Late_Blight_Modeling)  
Used programming skills: R

3. Sea level regulated tetraploid diversity dynamics through the Jurassic/Cretaceous interval  
Tessier, L., P. Mansell, P. D., & Uchupi, P. (2016). Sea level regulated tetraploid diversity dynamics through the Jurassic/Cretaceous Interval. *Nature Communications*, 7, 12731.  
submitted by Jon Tessier

Why should we attempt to reproduce this paper?  
Because it's a fun paper involving dinosaurs! But one which I myself have also attempted to reproduce in the past, and struggled with. There are a few additional twists that might throw some people off too.

Paper URL: <https://www.nature.com/articles/ncomms12731>  
Data URL: <https://zenodo.org/record/340803>  
Code URL: <https://zenodo.org/record/340803#supplementary-information>  
Used programming skills: R, Perl

4. Genotyping Polyploids from Messy Sequencing Data  
David Geman, Luis Felipe Valencia, Antonio Augusto Freitas Gama, and Christine Stephens. GENETICS November 2010 vol. 185 no. 3 1789-1807. <https://doi.org/10.1534/genetics.110.310486>  
submitted by David Geman

Why should we attempt to reproduce this paper?  
Reproducing this paper will give us exposure to organizing reproducible results with a workflow. I'm excited to see what changes I should make to make my future work more reproducible.

# On the day

- **Select paper and form groups**
- **Work with materials and reproduce**
- **Discuss**
- **Feed back to authors**

# Tips for Reproducing & Reviewing



# Selecting Papers

- Information submitted by authors:
  - Languages / tools used
  - Why you should attempt the paper.
- No. attempts ↗ No. times reproduction has been attempted
- Mean Repro Score ✓ Mean reproducibility score (out of 10)
  - lower == harder!



# Review as an auditor



# Access

- How **easy** was it to **gain** access to the materials?
- Did you manage to download all the files you needed?

# Installation

- How **easy / automated** was **installation**?
- Did you have any problems?
- How did you solve them?

# Data

- Were data clearly separated from code and other items?
- Were large data files deposited in a trustworthy data repository and referred to using a persistent identifier?
- Were data documented ...somehow...

# Documentation

Was there adequate documentation describing:

- how to install necessary software including non-standard dependencies?
- how to use materials to reproduce the paper?
- how to cite the materials, ideally in a form that can be copy and pasted?

# Analysis

- Were you able to fully reproduce the paper? 
- How automated was the process of reproducing the paper?
- How easy was it to link analysis code to:
  - the plots it generates
  - sections in the manuscript in which it is described and results reported

## If the analysis was not fully reproducible

- Were there missing dependencies?
- Was the computational environment not adequately described / captured?
- Was there bugs in the code?
- Did code run but results (e.g. model outputs, tables, figures) differ to those published? By how much?

# Review as a user



New User



Invested User



# Feedback as a community member

Acknowledge author effort

Give feedback in good faith

Focus on community benefits and system level solutions

*Help build convention on what form a Reproducible paper should take and how we should be able to use it*

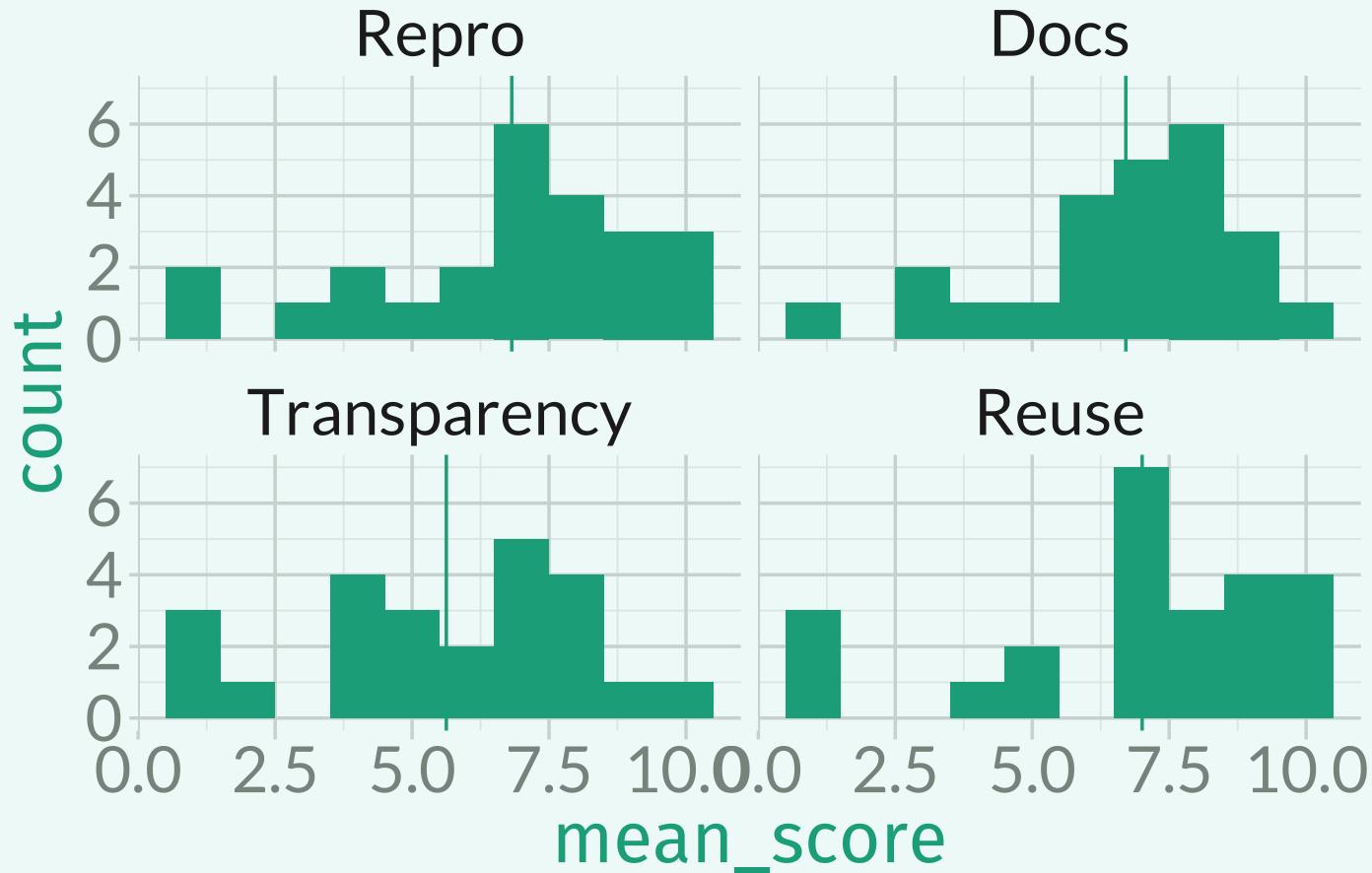


# **What did we learn?**

# N8 CIR ReproHack Series Stats

- **38 papers submitted so far**
- **Total of ~ 70 participants**
- **39 completed reviews over 27 papers**

# Review Scores

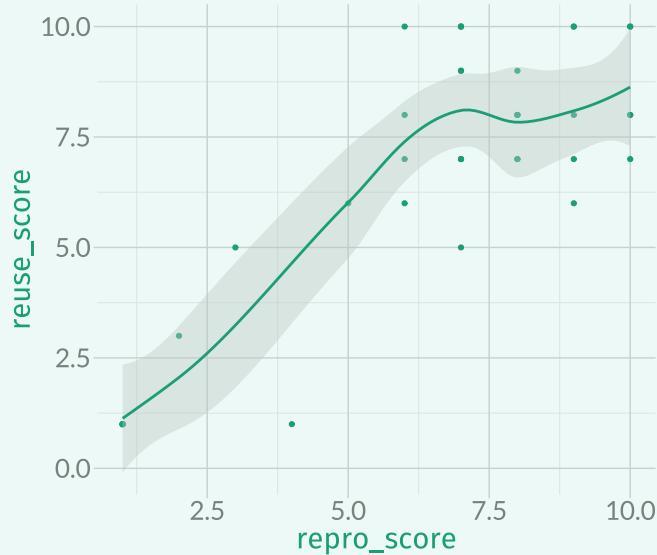
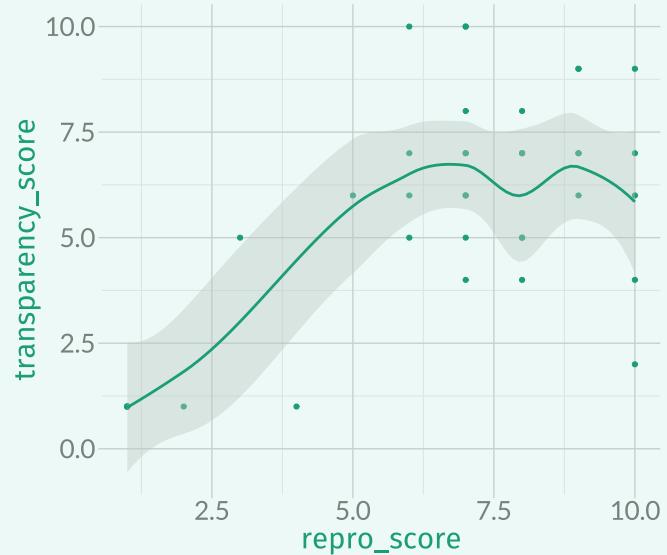


# Positives vs challenges

Difference in relative frequency of the top 40 terms used when talking about **positive** and **challenging** aspects of approaches to reproducibility.



# Trade-offs



# ReproHacks are fun



# Opportunity for peer skill sharing

- CCMcr: Contributing to open source
- Leiden: Syncing GitHub repositories with Zenodo
- Remote Reprohack: Docker school

# Fit for purpose



Durham University Advanced Re...  
@ARC\_DU



Replying to @annakrystalli @N8CIR and 3 others

On the way home, [@df3n5](#) said quite rightly, if all [code-producing/data-analysing] researchers would take part in at least one [@ReproHack](#), the code reproducibility and quality of documentation would generally soar!

8:19 AM · Jan 22, 2020



6



See Durham University Advan...

# The Way Forward

**Define**

**Practice**

**Create**

**Review**

# **Define: Research Compendium**

# The concept of a Research Compendium

“ ...We introduce the **concept of a compendium** as both a **container for the different elements** that make up the document and its computations (i.e. text, code, data, ...), and as a **means for distributing, managing and updating the collection.**”

*Gentleman and Temple Lang, 2004*

## Principles

- Stick with peer conventions
- Keep data, methods and outputs separate
- Specify the computations environment as clearly as possible

# Research compendia in R

Ben Marwick, Carl Boettiger & Lincoln Mullen (2018) *Packaging Data Analytical Work Reproducibly Using R (and Friends)*

| R package structure is an excellent way for sharing research compendia.

## Convention →

## Automation, templates and checklists ✨

# Create: rrtools

# rrtools: Creating Compendia in R

**"The goal of *rrtools* is to provide instructions, templates, and functions for making a basic compendium suitable for writing reproducible research with R."**

Install *rrtools* from GitHub

```
# install.packages("devtools")
devtools::install_github("benmarwick/rrtools")
```

# Create compendium

```
rrtools::create_compendium("~/Documents/workflows/rrcompendium")
```

- ✓ Setting active project to '/Users/Anna/Documents/workflows/rrcompendium'
- ✓ Creating 'R/'
- ✓ Creating 'man/'
- ✓ Writing 'DESCRIPTION'
- ✓ Writing 'NAMESPACE'
- ✓ Writing 'rrcompendium.Rproj'
- ✓ Adding '.Rproj.user' to '.gitignore'
- ✓ Adding '^rrcompendium\\.Rproj\$', '^\\.Rproj\\\\.user\$' to '.Rbuildignore'
- ✓ Opening new project 'rrcompendium' in RStudio
- ✓ The package rrcompendium has been created
- ✓ Opening the new compendium in a new RStudio session...

Next, you need to: ↓ ↓ ↓

- Edit the DESCRIPTION file
- Use other 'rrtools' functions to add components to the compendium

# DESCRIPTION file

## Package metadata

```
Package: rrcompendiumDTB
Title: Partial Reproduction of Boettiger Ecology Letters 2018;21:1255-1267
Version: 0.0.0.9000
Authors@R:
  person(given = "Anna",
          family = "Krystalli",
          role = c("aut", "cre"),
          email = "annakrystalli@googlemail.com")
Description: This repository contains the research compendium of the partial reproduction of Boettiger Ecology Letters 2018;21:1255-1267. The core contains all data, code, and text associated with this sub-section of analysis.
```

# Prepare for sharing

```
rrtools::use_readme_rmd()
```

- ✓ Creating 'README.Rmd' from template.
- ✓ Adding 'README.Rmd' to `Rbuildignore`
- Modify 'README.Rmd'
- ✓ Rendering README.Rmd to README.md for GitHub
- ✓ Adding code of conduct.
- ✓ Creating 'CONDUCT.md' from template.
- ✓ Adding 'CONDUCT.md' to `Rbuildignore`
- ✓ Adding instructions to contributors.
- ✓ Creating 'CONTRIBUTING.md' from template.
- ✓ Adding 'CONTRIBUTING.md' to `Rbuildignore`

## rcompendium

This repository contains the data and code for our reproduction paper:

Krystalli, A. (2018). *Partial Reproduction of Boettiger Ecology Letters 2018;21:1255–1267 with rrtools*.  
<https://doi.org/DOI-to-paper>

Our reproduction pre-print is online here:

Krystalli, A. (2018). *Partial Reproduction of Boettiger Ecology Letters 2018;21:1255–1267 with rrtools*, Accessed 27 Nov 2018. Online at <https://doi.org/DOI-to-compendium>

### How to cite

Please cite this compendium as:

Krystalli, A. (2018). Compendium of R code and data for 'Partial Reproduction of Boettiger Ecology Letters 2018;21:1255–1267 with rrtools'. Accessed 27 Nov 2018. Online at <https://doi.org/DOI-to-compendium>

### How to download or install

You can download the compendium as a zip from this URL:  
<https://github.com/annakrystalli/rcompendium/archive/master.zip>

Or you can install this compendium as an R package, rcompendium, from GitHub with:

### Licenses

Text and figures : CC-BY-4.0 Copyright (c) 2018 Carl Boettiger.

Code : See the [DESCRIPTION](#) file

Data : CC-BY-4.0, Copyright (c) 2018 Carl Boettiger.

### Contributions

We welcome contributions from everyone. Before you get started, please see our [contributor guidelines](#). Please note that this project is released with a [Contributor Code of Conduct](#). By participating in this project you agree to abide by its terms.

# Create analysis folder

```
rrtools::use_analysis()
```

- ✓ Adding bookdown to Imports
- ✓ Creating 'analysis' directory and contents
- ✓ Creating 'analysis'
- ✓ Creating 'analysis/paper'
- ✓ Creating 'analysis/figures'
- ✓ Creating 'analysis/templates'
- ✓ Creating 'analysis/data'
- ✓ Creating 'analysis/data/raw\_data'
- ✓ Creating 'analysis/data/derived\_data'
- ✓ Creating 'references.bib' from template.
- ✓ Creating 'paper.Rmd' from template.

Next, you need to: ↓ ↓ ↓ ↓

- Write your article/report/thesis, start at the paper.Rmd file
- Add the citation style library file (csl) to replace the default provided here
- Add bibliographic details of cited items to the 'references.bib' file
- For adding captions & cross-referencing in an Rmd, see <https://bookdown.org/>
- For adding citations & reference lists in an Rmd, see <http://rmarkdown.rstudio.com>

# Capturing dependencies

```
rrtools::add_dependencies_to_description()
```

Imports:

```
bookdown,  
ggplot2 (>= 3.0.0),  
ggthemes (>= 3.5.0),  
here (>= 0.1),  
knitr (>= 1.20),  
rticles (>= 0.6)
```

Paper using **renv** / **packrat** & docker: <https://github.com/joelnitta/pleurosoriopsis>

# Further Helpers

## articles

Contains a [suite of custom R Markdown templates for popular journals](#), simplifying the creation of documents that conform to research paper submission standards.



## RStudio Add-in to Insert Markdown Citations

A screenshot of the RStudio interface demonstrating the citr add-in. The main window shows an R Markdown file with code and text. A modal dialog titled "Insert citation" is open, displaying a citation entry:

```
Bartlett (1960). Stochastic population models in ecology and epidemiology.  
[@Bartlett1960]
```

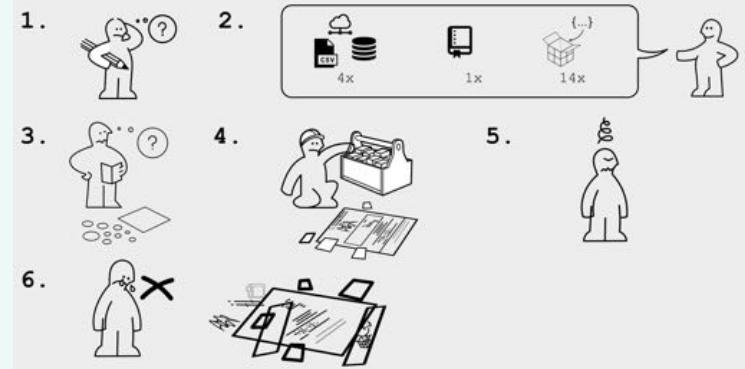
The "In parentheses" checkbox is checked. Below the dialog, a message indicates bibliography files found in YAML front matter: `refs.bib`. The RStudio sidebar shows a file tree for the project "rtools-wkhp-materials". The console at the bottom shows the command `citr:::insert_citation()` being run and the server listening on port 7464.

# **Is sharing this enough?**

# Case Study: Sharing a Geospatial Analysis in R

On a computer without System Library **GDAL** ✗

```
package 'rgdal' successfully unpa  
and MD5 sums checked  
  
configure: gdal-config: gdal-conf  
checking gdal-config usability...  
line 1353: gdal-config: command n  
no  
Error: gdal-config not found  
...  
ERROR: configuration failed for  
package 'rgdal'
```

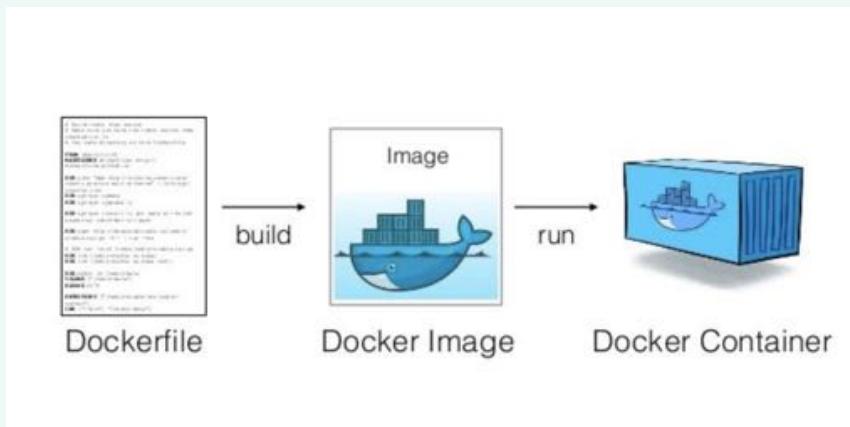


slide: Karthik Ram: rstudio::conf 2019 talk

# What are Docker containers?

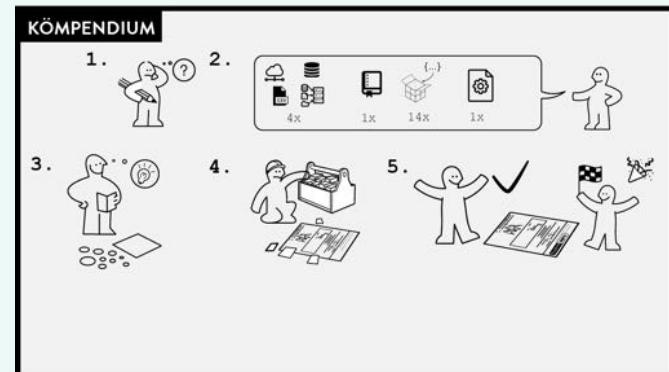
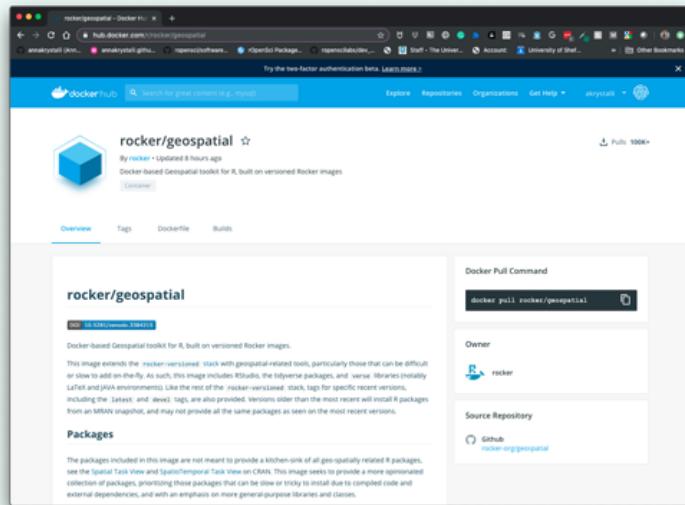
standardized units of software that **package up everything needed to run an application**: *code, runtime, system tools, system libraries and settings* in a lightweight, standalone, executable package

- **Dockerfile**: Text file containing recipe for setting up computation environment.
- **Docker Image**: Executable **built** from the **Dockerfile** with all required dependencies installed. Can have many images from the same **Dockerfile**.
- **Docker Container**: Docker Images become containers at **runtime**



# Rocker on DockerHub

using the [rocker/geospatial](#) Docker Image 



slide: Karthik Ram: rstudio::conf 2019 talk

# Create Dockerfile w/ rrtools

```
rrtools::use_dockerfile()
```

- ✓ Creating '**Dockerfile**' from template.
- ✓ Adding '**Dockerfile**' to `'.Rbuildignore`.
- Modify

Next:

- \* Edit the dockerfile with your name & email
- \* Edit the dockerfile to include system dependencies, such as linux lib
- \* Check the last line of the dockerfile to specify which Rmd should be

# Dockerfile

```
# get the base image, the rocker/verse has R, RStudio and pandoc
FROM rocker/verse:3.6.0

# required
MAINTAINER Anna Krystalli <annakrystallil@googlemail.com>

COPY . /rrcompendiumDTB

# go into the repo directory
RUN . /etc/environment \
    # Install linux dependencies here
    # e.g. need this for ggforce::geom_sina
    && sudo apt-get update \
    && sudo apt-get install libudunits2-dev -y \
    # build this compendium package
    && R -e "devtools::install('/rrcompendiumDTB', dep=TRUE)" \
    # render the manuscript into a docx, you'll need to edit this if you've
    # customised the location and name of your main Rmd file
    && R -e "rmarkdown::render('/rrcompendiumDTB/analysis/paper/paper.Rmd"
```

# Docker + Travis

## Create .travis.yml

```
rrtools::use_travis()
```

- ✓ Creating '.travis.yml' from template.
- ✓ Adding '.travis.yml' to `Rbuildignore`.

Next:

- \* Add a travis shield to your README.Rmd:

[![Travis-CI Build Status](<https://travis-ci.org/annakrystalli/rrcompenc>)

- \* Turn on travis for your repo at <https://travis-ci.org/annakrystalli/i>

\* To connect Docker, go to <https://travis-ci.org/>, and add your environment variables: DOCKER\_EMAIL, DOCKER\_USER, DOCKER\_PASS to enable pushing to Docker Hub

# .travis.yml

```
env:  
  global:  
    - REPO=$DOCKER_USER/rrcompendiumdtb  
  
sudo: required  
  
warnings_are_errors: false  
  
language: generic  
  
services:  
  - docker  
  
before_install:  
  - docker build -t $REPO .
```

Create & build image using dockerfile, i.e. compile pkg and render Rmd to Word doc

# .travis.yml

Push our custom docker image to docker hub, env vars stored on travis-ci.org

```
after_success:  
  - docker login -u $DOCKER_USER -p $DOCKER_PASS  
  - export REPO=$DOCKER_USER/rrcompendiumdtb  
  - export TAG=`if [ "$TRAVIS_BRANCH" == "master" ]; then echo "latest"; else echo $TRAVIS_BRANCH; fi`  
  - docker build -f Dockerfile -t $REPO:$TAG .  
  - docker tag $REPO:$TAG $REPO:$TAG  
  - docker tag $REPO:$TAG $REPO:travis-$TRAVIS_BUILD_NUMBER  
  - docker push $REPO
```

Travis repository settings

The screenshot shows the 'Environment Variables' section of a Travis CI repository settings page. It contains two entries:

- DOCKER\_PASS**: A variable set to a masked password, available to all branches.
- DOCKER\_USER**: A variable set to a masked user name, available to all branches.

@annakrystalli

# Travis build passes!

annakrystalli / rrcompendiumDTB Build passing

Current Branches Build History Pull Requests More options ⚙

✓ master correct compendium name on README

- Commit 90157b3 ↗
- ↳ Compare 3ea62f9..90157b3 ↗
- ▷ Branchmaster ↗

Anna Krystalli

↳ ↗ no language set

AMD64

build passing

🐦 @annakrystalli

# Image on Dockerhub

The screenshot shows the Dockerhub profile for user **akrystalli**. The profile includes a blue fingerprint icon, the username **akrystalli**, a link to [Edit profile](#), a status of **Community User**, and a joining date of **November 7, 2015**. Below the profile, there are tabs for **Repositories** (which is selected), **Starred**, and **Contributed**. The message "Displaying 3 of 3 repositories" is shown above a list. The first repository listed is **akrystalli/rrcompendiumdtb**, created by **akrystalli** and updated a few seconds ago. It is categorized as a **Container** and has 1 download. A small gray cube icon is next to the repository name.

Docker Image: <https://hub.docker.com/repository/docker/akrystalli/rrcompendiumdtb>

Compendium Repository: <https://github.com/annakrystalli/rrcompendiumDTB>

 @annakrystalli

# **On reproducible lab culture**

# Documentation

**The heart of communities of practice**

## Turing Way

<https://the-turing-way.netlify.app/>

- Great source of general best practice.

**Needs to be translated to on-the-ground lab practice guidelines.**

# **Templates, checklists and automation in the lab**

**Define and document lab level procedures & conventions.**

- Clear and complete on-boarding.
- Guidance on creating and managing digital research outputs
- Clear off-boarding procedures including archiving of generated materials.

**Basics can be templated and provided in  
customisable formats**

# Checklib



<https://github.com/checklib/checklib>

#ColabW19 Project: a library of markdown checklists that can be incorporated as GitHub issue templates into your own projects.



## Set up GitHub Authentication

Set up GitHub authentication through a Personal Access Token

[Get started](#)



## Set up GitHub Authentication in R

Set up GitHub authentication in R through a Personal Access Token

[Get started](#)

<https://github.com/annakrystalli/IJCNextR/issues/1>

## Set up GitHub Authentication in R

[Create](#) annakrystalli opened this Issue 4 days ago · 0 comments

annakrystalli commented 4 days ago · edited

Owner

Assignees  
No one—assign yourself

Labels  
None yet

Projects  
None yet

Milestone  
No milestone

Notifications  
[Unsubscribe](#)  
You're receiving notifications because you authored the thread.

1 participant

[Lock conversation](#)

[Pin issue](#)

[Transfer issue](#)  
Beta

[Delete issue](#)

**Set up GitHub Access in R/RStudio**

Set up GitHub authentication through a Personal Access Token in R

- Create New personal access token (PAT)
- Store PAT as Environment Variable

**Guidance**

**Create Personal Access Tokens**

Personal access tokens function like ordinary OAuth access tokens. They can be used instead of a password for GitHub over HTTPS, or can be used to authenticate to the API over Basic Authentication.

You can set PAT authentication through R using package `useRtlib`, and is easier through RStudio. To install the package run:

```
install.packages("useRtlib")
```

To create a new PAT use:

```
useRtlib::github_pat()
```

This launches a browser and navigates to the GitHub URL for creating new Personal Access Token. The URL incorporates the PAT specification with:

- A description for the token: `RISETHUR_PAT`
- The scope of the authorization. This defines the access for personal tokens and defaults to full access to repos and gists only. You can amend this if required. Read more about OAuth scopes

[Click Generate token](#).

Then copy the token to your clipboard. For security

# **On the future of Reviewing**

REVIEWER

AUTHOR

I FOUND AND CORRECTED SOME  
TYPOS AND A MINOR ERROR IN THE  
CODE RESULTS. I AM PLEASED.

😊 THANKS!

I REANALYSED  
PARAMETER X.  
SIGNIFICANT  
RESULTS

WOW! THAT'S AN INTERESTING FINDING. IT  
MAY ALSO BE USEFUL TO REOPEN A SIGNIFICANT  
MANUSCRIPT. WOULD YOU LIKE TO  
COLLABORATE ON AN UPDATED  
MANUSCRIPT?

I THINK YOU SHOULD PERFORM  
COMPLICATED ANALYSIS Y

WE FEEL THAT'S BEYOND THE SCOPE OF THIS  
PAPER. THE CODE AND DATA ARE FREELY  
AVAILABLE. LET US KNOW IF YOU WISH TO  
COLLABORATE ON A NEW PAPER.

# On the scope of reproducibility

- Reproducibility *ad infinitum*
  -  UNREALISTIC

# On the scope of reusability

## Openness can help:

- surface useful parts of code.
- facilitate user feedback and contribution

**MAINTENANCE?!**

# On the scope of reproducibility

- Reproducibility *ad infinitum*
  - **✗ UNREALISTIC**
- Reproducibility for 2-3 years post-publication
  - **✓ MORE REALISTIC**
  - Checked as part of publication process, e.g. CODE CHECK  
<https://codecheck.org.uk/>

**CODE WORKS ✓**

...in the meantime

**take any opportunity to practice!**

# ReproHack

Multiple ways to run a ReproHack



Are the participants geographically located in the same place?

YES

## Event ReproHack



Conference



University

## Research Group ReproHack



Team

- Team with people of different backgrounds.
- Decide which paper reproduce from a variety of options.
- Networking.

- Let your team reproduce you article before is submitted.
- Reproduce papers related to your research topic
- Improve the capabilities of your team in scientific reproducibility.

NO

## Remote ReproHack



- Participants can join as a research group or work together in a particular paper selecting different breaking rooms.
- It allows the presence of scientists around the world.



## Ways to participate

### Propose a paper

You've put a lot of effort into making your work reproducible. Now let people learn from and engage with it!

#### Benefits to authors:

- Feedback on the reproducibility of your work.
- Appreciation for your efforts in making your work reproducible.
- Opportunity to engage others with your research.

[Submit paper!](#)

### Reproduce

Join a ReproHack and get working with other people's material!

#### Benefits to participants:

- Practical experience in reproducibility with real published materials
- Opportunity to explore different tools and strategies.
- Opportunity to for meaningful contribution.
- Inspiration to work more openly.

[Join an event!](#)

### Organise an event

Help create a practical learning space

#### Benefits to community:

- Help build capacity in reproducibility throughout the research community.
- Highlight community value of reproducibility beyond validation of results.
- Help community evaluate how successful current practices are and for what purpose.
- Help identify what works and where the most pressing weaknesses in our approaches are<sup>1</sup>.

[Submit an event!](#)

# Interested in ReproHacking?

## **reprohack/rephack-hq GH repository**

Chat to us:

slack [join us](#)

### **Host your own event!**

### **Submit your own papers!**

- Challenges remain to moving from theory to practice
- We need to clearly define our expectations of a research compendium
- This will allow to develop tools and templates
- ReproHacks provide great opportunities to practice

 **Thanks for** 

?

# Resources

- [The Turing Way](#): a lightly opinionated guide to reproducible data science.
- [Statistical Analyses and Reproducible Research](#): Gentleman and Temple Lang's introduction of the concept of Research Compendia
- [Packaging data analytical work reproducibly using R \(and friends\)](#): how researchers can improve the reproducibility of their work using research compendia based on R packages and related tools
- [How to Read a Research Compendium](#): Introduction to existing conventions for research compendia and suggestions on how to utilise their shared properties in a structured reading process.
- [Reproducible Research in R with rrtools](#): Workshop: Create a research compendium around materials associated with a published paper (text, data and code) using [`rrtools`](#).
  - [Example Compendium](#): Demo Research compendium.

# Acknowledgements

Images throughout the slides watermark with **Scriberia** were created by **Scriberia** for The Turing Way community and is used under a CC-BY licence

- *The Turing Way Community, & Scriberia.* (2019, July 11). Illustrations from the Turing Way book dashes. Zenodo. <http://doi.org/10.5281/zenodo.3332808>

Photo on slide #24 by [Annie Spratt](#) on [Unsplash](#)

Photo on slide #25 [Sharon McCutcheon](#) on [Unsplash](#)