

Putting the into Reproducible Research

Directors Cut

Dr Anna Krystalli

University of Sheffield RSE

2020/04/21

RSS Leeds / Bradford Local Group Seminar





Hello

me: **Dr Anna Krystalli**

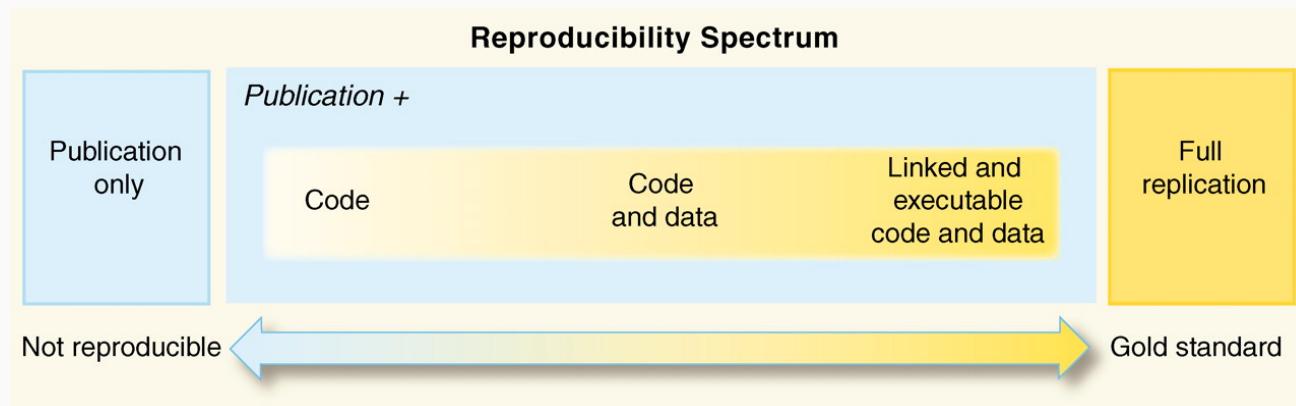
- **Research Software Engineer**, *University of Sheffield*
 - twitter **@annakrystalli**
 - github **@annakrystalli**
 - email **a.krystalli[at]sheffield.ac.uk**
- **Editor** **rOpenSci**
- **Co-organiser:** **Sheffield R Users group**

slides: **bit.ly/r-in-repro-research-dc-leeds**

Motivation

Calls for reproducibility

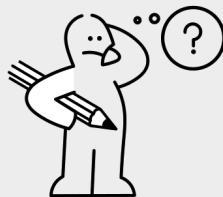
Reproducibility has the potential to serve as a **minimum standard for judging scientific claims** when full independent replication of a study is not possible.



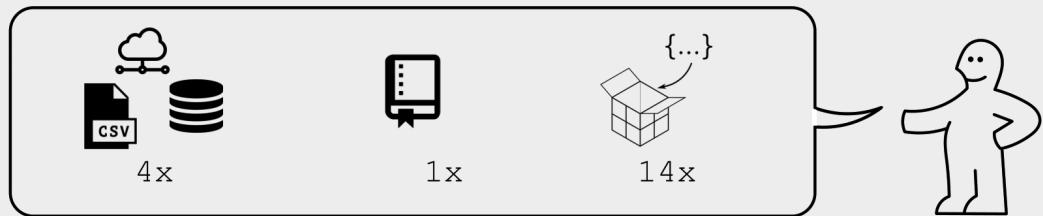
Reproducible Research in Computational Science ROGER D. PENG, SCIENCE 02 DEC 2011 : 1226-1227

Is code and data enough?

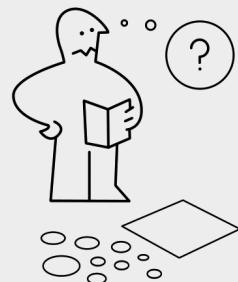
1.



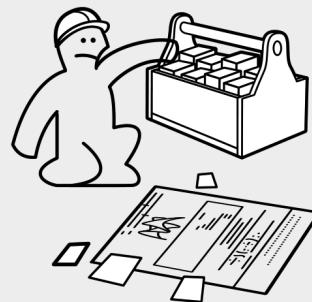
2.



3.



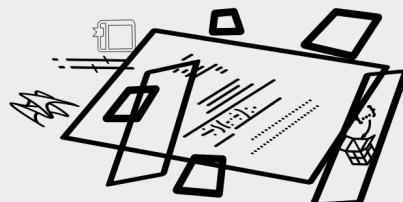
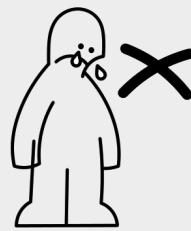
4.



5.



6.



slide: [Karthik Ram: rstudio::conf 2019 talk](#)

Calls for open science

Special Feature: 5th Anniversary of *Methods in Ecology and Evolution*

Shedding light on the ‘dark side’ of phylogenetic comparative methods

Natalie Cooper , Gavin H. Thomas, Richard G. Fitzjohn

First published: 13 June 2016 [Full publication history](#)

... highlight problems with users jumping straight into software implementations of methods (e.g. in r) that may lack documentation on biases and assumptions that are mentioned in the original papers.

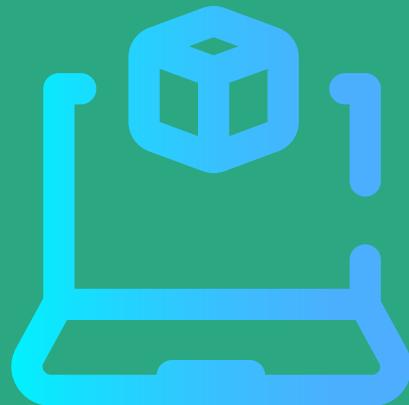
To help solve these problems, we make a number of suggestions including **providing blog posts** or **videos** to **explain new methods in less technical terms, encouraging reproducibility and code sharing**, making **wiki-style pages** summarising the literature on popular methods, more careful consideration and testing of whether a method is appropriate for a given question/data set, **increased collaboration**, and a shift from publishing purely novel methods to **publishing improvements to existing methods** and ways of detecting biases or testing model fit. Many of these points are applicable across methods in ecology and evolution, not just phylogenetic comparative methods.

R for Open Reproducible Research

A whistle-stop tour of tools, practices and conventions in R for more:

- **Reproducible**
- **Robust**
- **Transparent**
- **Reusable**
- **Shareable** research materials

Project management



Icon by [Freepik](#) from [flaticon.com](#)

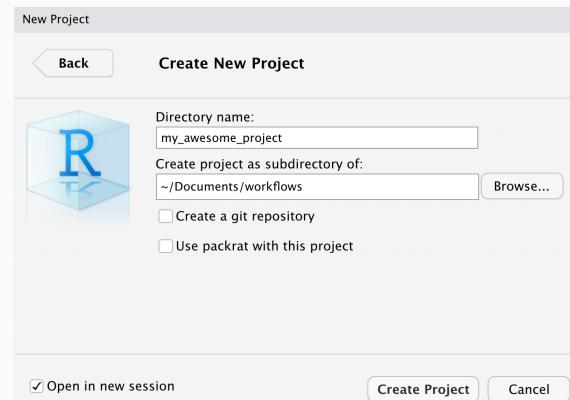
Rstudio Projects

Use Rstudio projects to keep materials associated with a particular analysis together

- **Self contained** and **portable**
- **Working directory set to root** of project on launch
- **Fresh session** everytime the project is launched

See Jenny Bryan's post on **project oriented workflows** for more details

File > New Project > New Directory



~/Documents/workflows/my_awesome_project - RStudio

Addins

my_awesome_project

Console Terminal Jobs

~/Documents/workflows/my_awesome_project/ ↵

```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Hey Anna, welcome back, time to Rrrrrrrrock!
> |
```

Environment History Connections

Import Dataset List C

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Documents > workflows > my_awesome_project Rproj ...

Name	Size
..	
my_awesome_project.Rproj	205 B



here

Use 📦 here to create robust relative paths

- **Robust paths relative to project root**

- portable
- independent of:
 - working directory
 - source code location



```
here :: here()
```

```
## [1] "/Users/Anna/Documents/workflows/talks"
```

```
here :: here("data", "summaries.csv")
```

```
## [1] "/Users/Anna/Documents/workflows/talks/data/summaries.csv"
```

Icon by **Freepik** from [flaticon.com](https://www.flaticon.com)

Dependency management

Minimal approach

include an `install.R` script

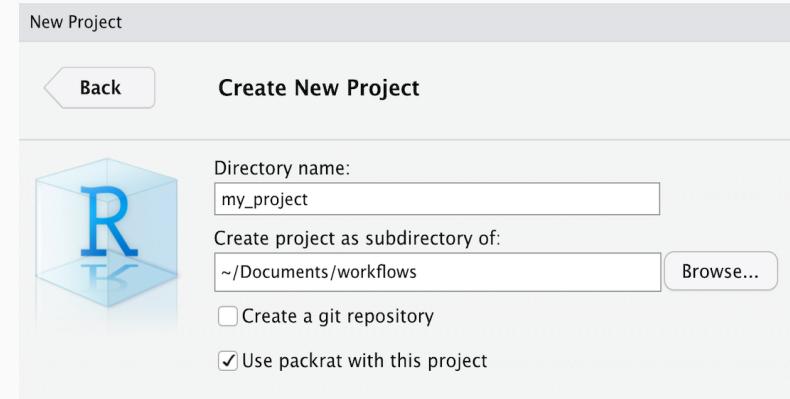
```
install.packages("dplyr")
install.packages("purrr")
```

Most robust approach

use  `renv` (previously `packrat`)

- Create and manage a per project library of packages
- initialise during project set up

will revisit later on





drake

Use 📦 drake to orchestrate your workflows



Scale the work
you need.



Skip the work
you don't.



See evidence
of reproducibility.



make plan

```
plan ← drake::drake_plan(  
  raw_data = readr::read_csv(here::here("data", "iris.csv")),  
  data = raw_data %>%  
    dplyr::mutate(Species = forcats::fct_inorder(Species)),  
  fit = lm(Sepal.Width ~ Petal.Width + Species, data))
```

Plan

view plan

```
plan
```

```
## # A tibble: 3 x 2
##   target    command
##   <chr>     <expr_lst>
## 1 raw_data  readr::read_csv(here::here("data", "iris.csv"))
## 2 data       raw_data %>% dplyr::mutate(Species = forcats::fct_inorder(Species))
## 3 fit        lm(Sepal.Width ~ Petal.Width + Species, data)
```

re-execute plan

```
drake::make(plan)
```

```
## All targets are already up to date.
```

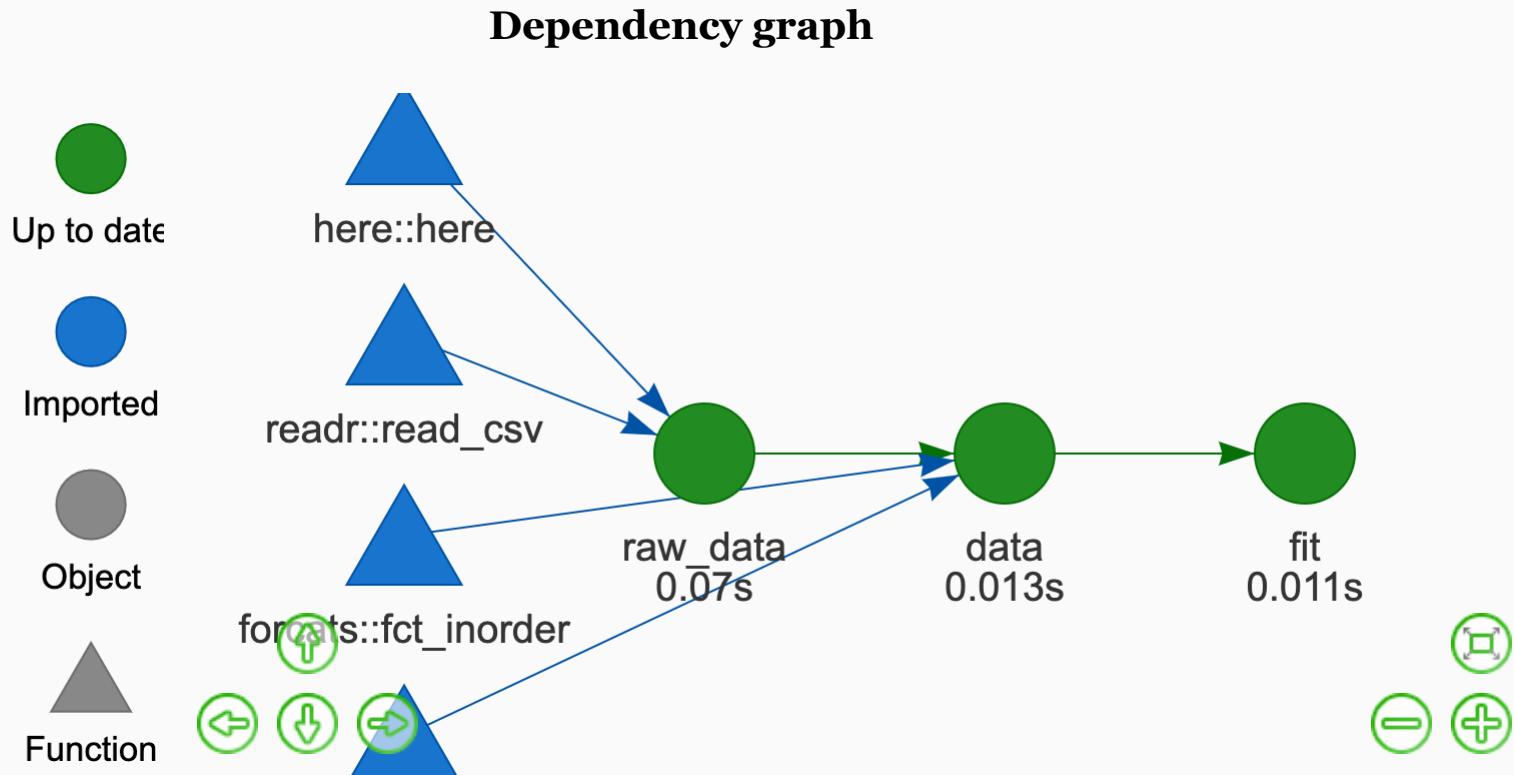
inspect targets

```
drake :: readd(fit)

##  
## Call:  
## lm(formula = Sepal.Width ~ Petal.Width + Species, data = data)  
##  
## Coefficients:  
## (Intercept)      Petal.Width  Speciesversicolor  Speciesvirginica  
##           3.236          0.781          -1.501          -1.844
```

visualise workflow

```
drake :: vis_drake_graph(drake :: drake_config(plan))
```



Version Control

Version Control

What is it? 🤔

The **management of changes** to documents, computer programs, large web sites, and other collections of information.

Git 🚀 git

Open source (free to use) **Version control software**.

GitHub 🐫

A **website** (<https://github.com/>) that allows you to **store your Git repositories online** and makes it easy to collaborate with others.

Why use them in research?

Exhibit A

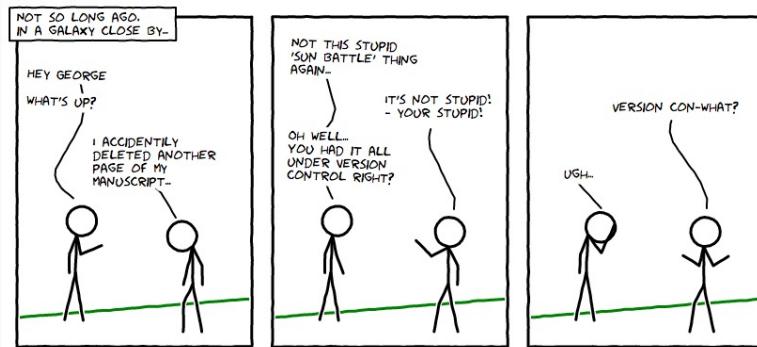


Image: xkcd CC BY-NC 2.5

Exhibit B

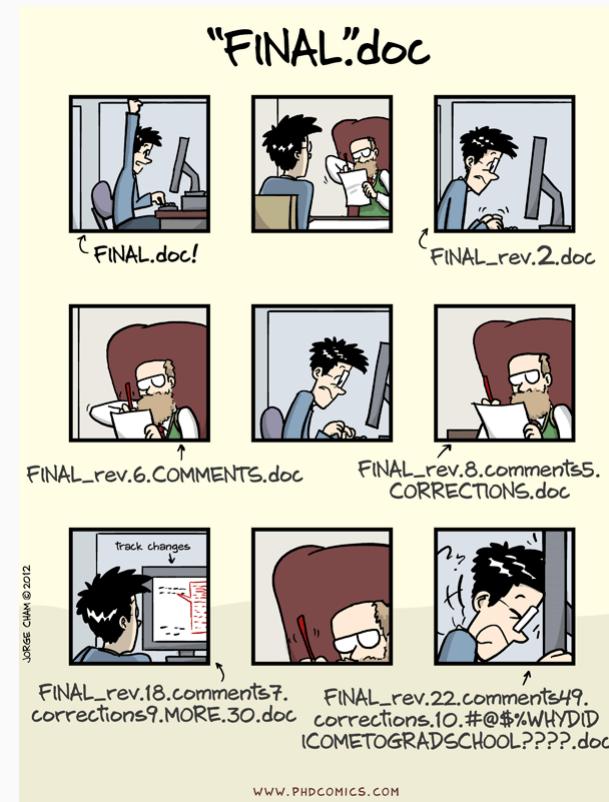


Image: Jorge Cham www.phdcomics.com

Git, Github & Rstudio

Before: git only through the terminal 😢

Now: Rstudio + usethis 📦 == ❤️ Git & GitHub 😊



Configure git & GitHub

Configure git

Check your configuration

```
usethis::git_sitrep()
```

Set your configuration

Use your github username and and the email you used to sign-up on GitHub

```
usethis::use_git_config(  
  user.name = "Jane",  
  user.email = "jane@example.org")
```

Configure GitHub authentication

Get GITHUB Personal Authorisation Token

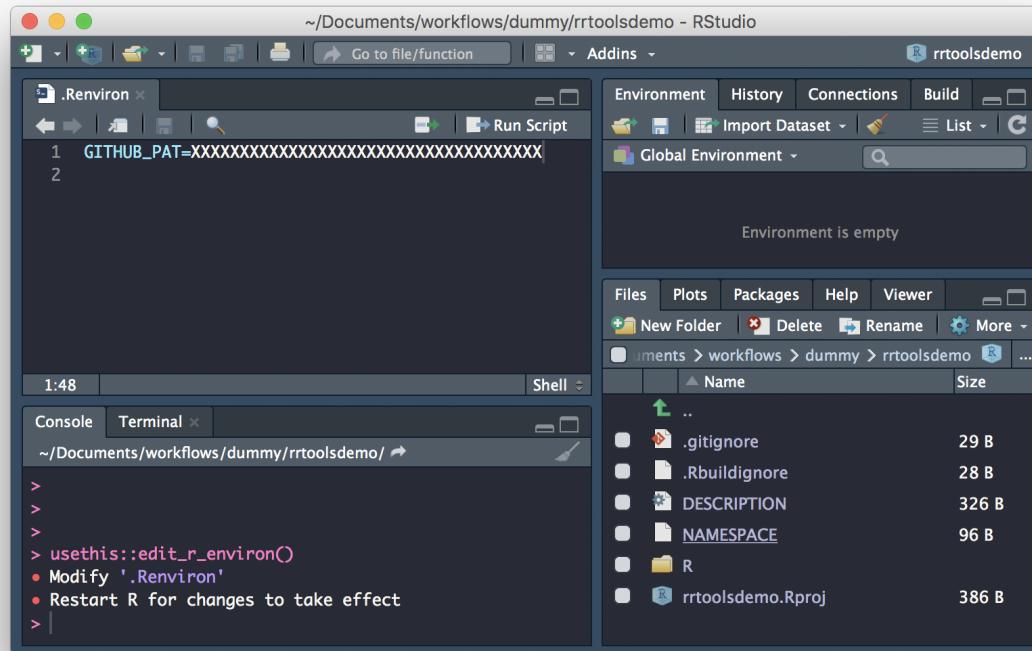
```
usethis :: browse_github_pat()
```

The screenshot shows the GitHub 'Personal access tokens' creation interface. On the left, a sidebar menu includes 'OAuth Apps', 'GitHub Apps', and 'Personal access tokens', with 'Personal access tokens' being the active tab. The main area is titled 'New personal access token'. It contains a 'Token description' input field containing 'R:GITHUB_PAT', a 'What's this token for?' section, and a 'Select scopes' section. The 'Select scopes' section lists several options, all of which are checked:

<input checked="" type="checkbox"/> repo	Full control of private repositories
<input checked="" type="checkbox"/> repo:status	Access commit status
<input checked="" type="checkbox"/> repo_deployment	Access deployment status
<input checked="" type="checkbox"/> public_repo	Access public repositories
<input checked="" type="checkbox"/> repo:invite	Access repository invitations

Store in `.Renvironment` file

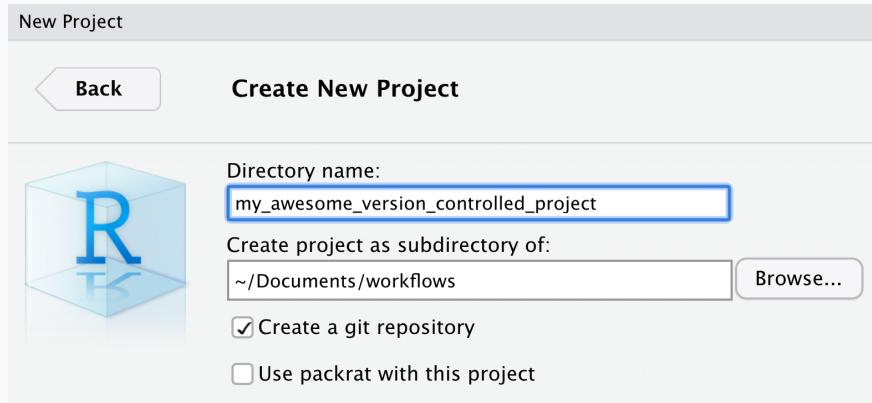
```
usethis::edit_r_environ()
```



Initialise git

Initialise Rstudio project with Git by just checking a box!

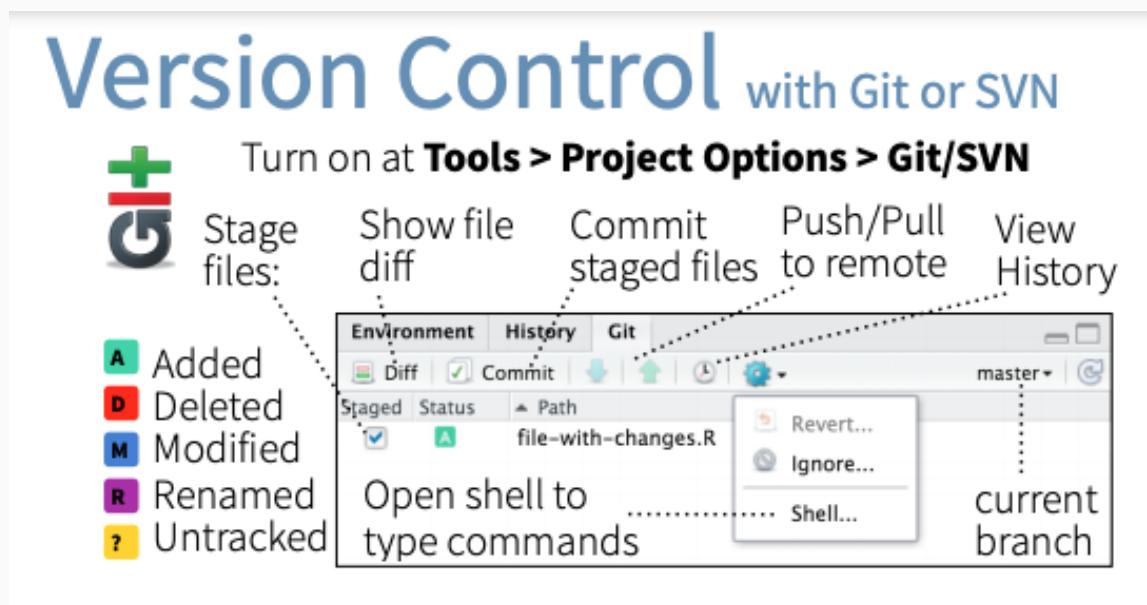
It's now a **repository**



Forgot to? use `usethis::use_git()`

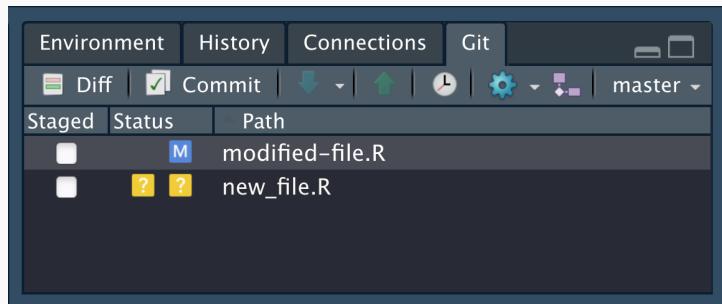
Git panel

Integrated graphical user interface

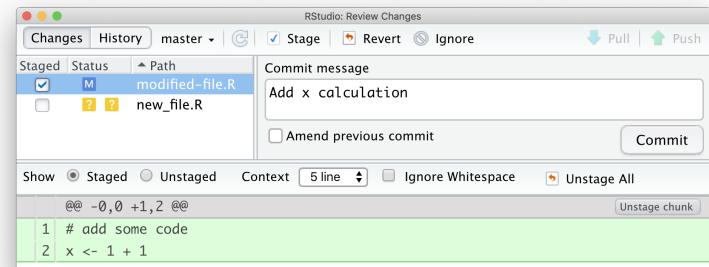


Git Rstudio workflow

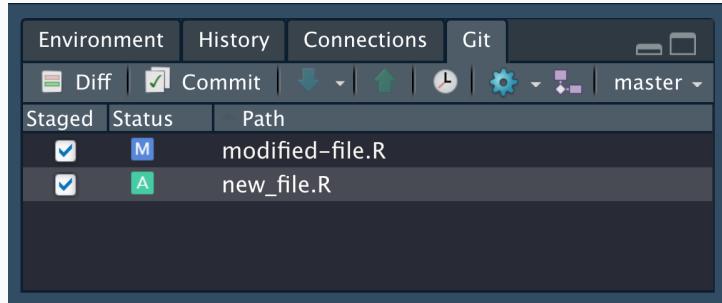
view file status



commit changes



stage files



Share on GitHub

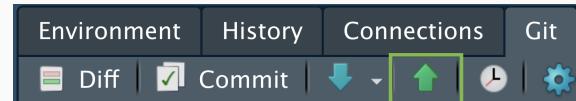
Create repo

```
useThis::use_github(protocol = "https")
```

The screenshot shows a GitHub repository page for 'annakrystalli / my_awesome_project'. The repository has 4 commits, 1 branch, 0 releases, and 1 contributor. The latest commit was made 18 hours ago. The commit history includes:

- Initial commit (18 hours ago)
- Add x calculation (18 hours ago)
- initial commit (18 hours ago)
- rename file (18 hours ago)

Push further changes



Anatomy of a GitHub Repo

- **README**. Explain what your project is, and how to use it.
 - `usethis::use_readme_md()`
 - `usethis::use_readme_rmd()`
- **LICENSE**. Without a licence, the contents of the repository are technically closed.
 - Examples licence **MIT**: `usethis::use_mit_license(name = "Anna Krystalli")`
 - `?licenses`: details of functions available to generate licenses
 - <https://choosealicense.com/> help on choosing a licence.
- **CONTRIBUTING.md** - guidelines for contributors.
 - `usethis::use_tidy_contributing()` provides a relatively strict but instructive template
- **CODE_OF_CONDUCT.md** set the tone for discourse between contributors.
 - `use_code_of_conduct()`

GitHub issues

use GitHub issues to plan, record and discuss tasks.

issues

A screenshot of the GitHub Issues page for the repository 'annakrystalli/gaitr'. The page shows 15 open issues and 14 closed issues. The issues are listed in chronological order, with the most recent at the top. Each issue card includes a title, a brief description, and labels. The interface includes filters for 'Issues 15', 'Labels 17', and 'Milestones 2'.

Issue #	Title	Labels	Opened By	Opened On
#15	y axis radar plot	data-munging, stats, summaries	lore86	26 Mar
#14	P_adj value wilcox test	data-munging, stats, summaries	lore86	26 Mar
#13	Add documentation of get_ & set_group & vars functions	documentation, enhancement	annakrystalli	5 Feb
#12	Publication strategy	publication	annakrystalli	5 Feb
#11	Anonymise variable names in examples (and tests?)	publication, testing	annakrystalli	5 Feb
#10	Add significance asterisks to plots	enhancement, plot	lore86	29 Jan
#9	Collate outputs into one file	data-munging, stats, summaries	lore86	29 Jan
#8	15 Open	✓ 14 Closed		

projects

A screenshot of the GitHub Projects page for the repository 'annakrystalli/gaitr'. The page displays a workflow titled 'Complete group analysis code workflow' with three columns: 'To do', 'In progress', and 'Done'. Each column contains several cards representing tasks. The 'To do' column has 2 items, 'In progress' has 3, and 'Done' has 4. Each card includes a title, a brief description, and labels. The interface includes a search bar for 'Filter cards'.

Column	Card 1	Card 2	Card 3	Card 4	Card 5	Card 6	Card 7	Card 8
To do	Collate outputs into one file	Parametric version of zscore, standard deviation in summary	Ability to choose between parametric/non-parametric testing	Bonferroni correction	Pvalue no bonferroni correction	Determine which arguments of each statistics test should be available through map function	Add ability to choose metadata column with variable labels	
In progress	data-munging, stats, summaries	enhancement, summaries	enhancement, stats, summaries	duplicate, stats	stats	question	nnnnull	
Done	enhancement, summaries	enhancement, stats, summaries	enhancement, stats	duplicate, stats	stats	question	nnnnull	

Literate programming with Rmarkdown

Literate programming

Programming paradigm first introduced by **Donald E. Knuth**.

Treat program as literature to be understandable to human beings

- focus on the logic and flow of human thought and understanding
- single document to integrate data analysis (executable code) with textual documentation, **linking data, code, and text**

Literate programming in R

Rmarkdown (.Rmd) integrates:

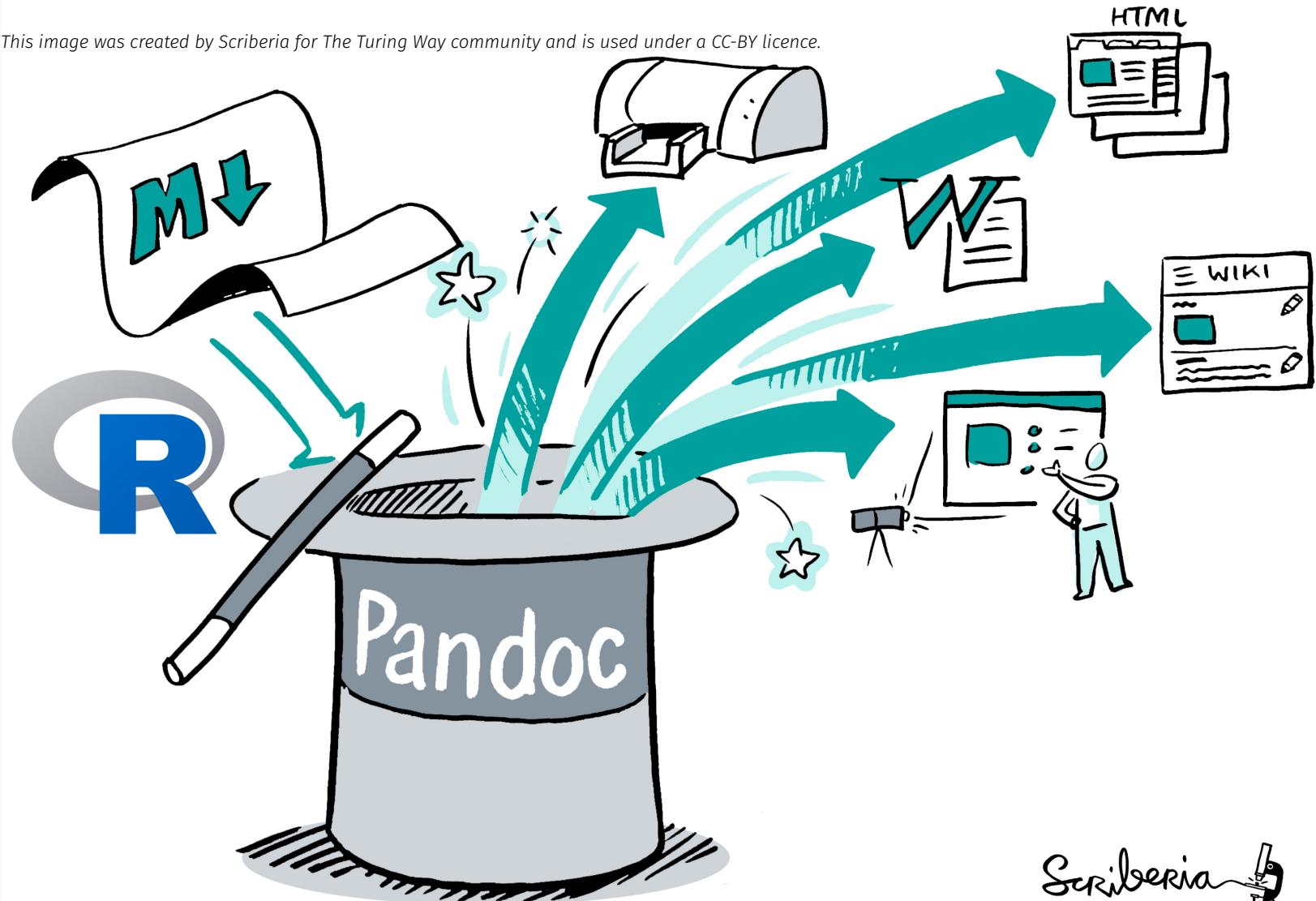
- a **documentantion** language (.md)
- a **programming** language (R)
- functionality to "**knit**" **them together** through  knitr

features

-  provides a framework for writing narratives around code and data
-  Code re-run in a clean environment every time the document is "knit"

Rmarkdown outputs

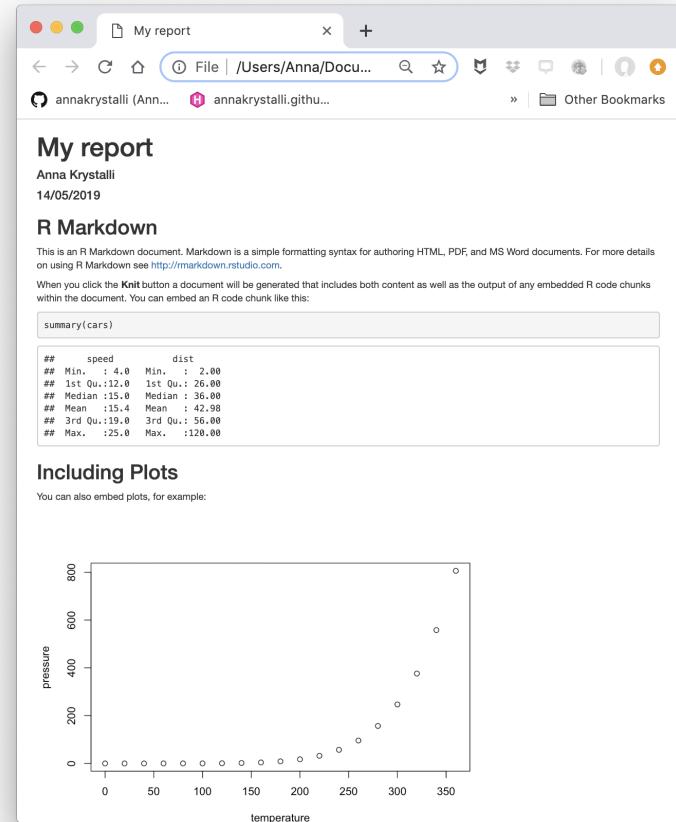
This image was created by Scriberia for The Turing Way community and is used under a CC-BY licence.



Rmarkdown to html

File > New File > RMarkdown... > Document

```
my-report.Rmd
1 ---  
2 title: "My report"  
3 author: "Anna Krystalli"  
4 date: "14/05/2019"  
5 output: html_document  
6 ---  
7  
8 ```{r setup, include=FALSE}  
9 knitr::opts_chunk$set(echo = TRUE)  
10```  
11  
12 ## R Markdown  
13  
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see  
http://rmarkdown.rstudio.com.  
15  
16 When you click the **Knit** button a document will be generated that includes both content as  
well as the output of any embedded R code chunks within the document. You can embed an R code  
chunk like this:  
17  
18 ```{r cars}  
19 summary(cars)  
20```  
21  
22 ## Including Plots  
23  
24 You can also embed plots, for example:  
25  
26 ```{r pressure, echo=FALSE}  
27 plot(pressure)  
28```  
29  
30 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the  
R code that generated the plot.  
31
```



Applications in research

Rmd documents can be useful for a number of research related **long form documentation** materials:

- Documentation of code & data (eg  DataMaid)
- Electronic Notebooks
- Supplementary materials
- Reports
- Papers

DOCUMENT



memegenerator.net

Rmd Vs Word



Spell check in Rstudio!

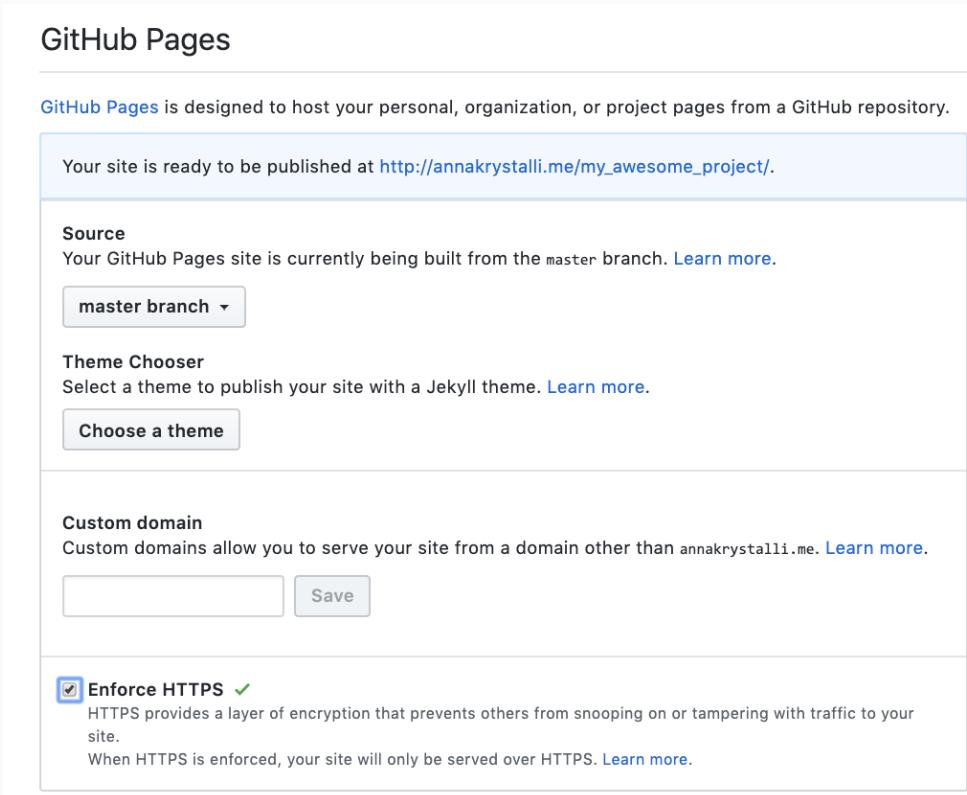


redoc HOT OFF THE PRESS

Enables a two-way R Markdown-Microsoft Word workflow. It generates Word documents that can be de-rendered back into R Markdown, **retaining edits on the Word document**, including tracked changes.

Publish to the web for free!

- **RPubs**: Publish rendered rmarkdown documents on the web with the click of a button
<http://rpubs.com/>
- **GitHub**: Host your site through **gh-pages** on GitHub. Enable in GitHub repo  **Settings**



The screenshot shows the GitHub Pages settings interface. At the top, it says "GitHub Pages" and describes its purpose: "GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository." Below this, a message says "Your site is ready to be published at http://annakrystalli.me/my_awesome_project/".
Source: Your GitHub Pages site is currently being built from the `master` branch. [Learn more.](#) A dropdown menu shows "master branch".
Theme Chooser: Select a theme to publish your site with a Jekyll theme. [Learn more.](#) A button says "Choose a theme".
Custom domain: Custom domains allow you to serve your site from a domain other than annakrystalli.me. [Learn more.](#) There is a text input field and a "Save" button.
Enforce HTTPS : HTTPS provides a layer of encryption that prevents others from snooping on or tampering with traffic to your site. When HTTPS is enforced, your site will only be served over HTTPS. [Learn more.](#)

Rmarkdown extensions

Many great packages and applications build on rmarkdown.

All this makes it **incredibly versatile**.

bookdown

Create and maintain online books

Authoring with R Markdown. Offers:

- cross-references,
- citations,
- HTML widgets and Shiny apps,
- tables of content and section numbering

The publication can be exported to HTML, PDF, and e-books (e.g. EPUB)

Examples

- rOpenSci Software Review policies
- Geocomputation in R

Thesisdown

An updated R Markdown thesis template using the bookdown package

pkgdown

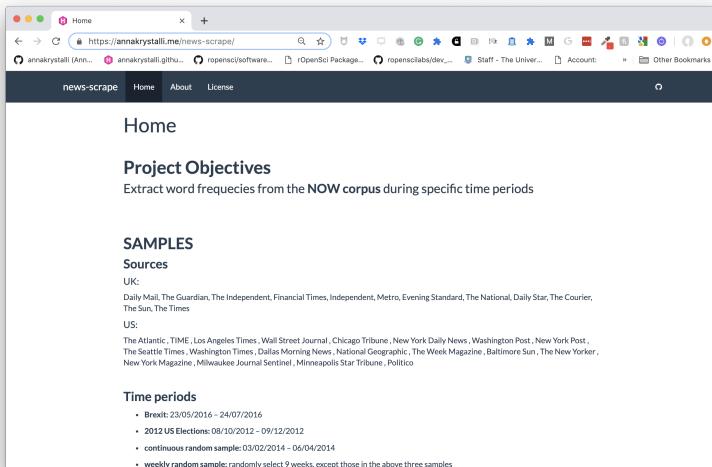
For building package documentation

Produce **function references** from `.Rd` files and **demonstrate function use** through long form documentation (vignettes).

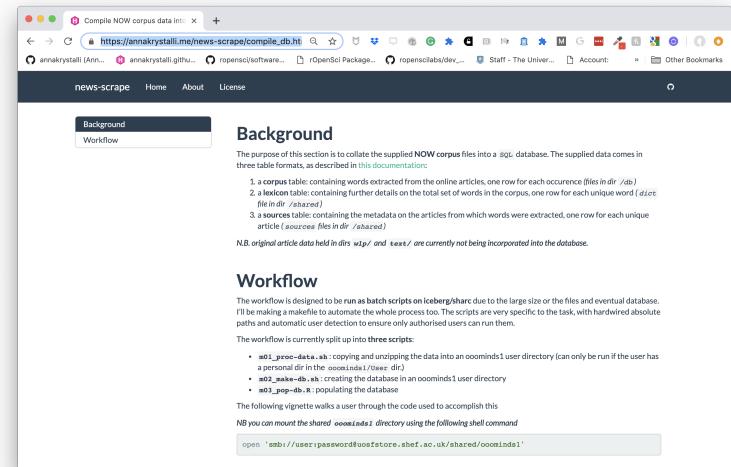
workflowr pkg

Build analyses websites and organise your project

Makes it easier for researchers to organize projects and share results. Includes **checks to ensure rendered versions correspond to up to date versions of code.**



The screenshot shows a web browser window with the URL <https://annakrystalli.me/news-scrape/>. The page has a dark header with the project name "news-scrape". Below the header, there are links for "Home", "About", and "License". The main content area is titled "Project Objectives" and contains the text: "Extract word frequencies from the NOW corpus during specific time periods". Under "SAMPLES", it lists "Sources" and "UK:" with a list of UK newspapers. Under "US:", it lists various US publications. At the bottom, there's a section for "Time periods" with a bulleted list of sample types: "Brexit: 23/05/2016 - 24/07/2016", "2012 US Elections: 08/10/2012 - 09/12/2012", "continuous random sample: 03/02/2014 - 06/04/2014", and "weekly random sample: randomly select 9 weeks, except those in the above three samples".



The screenshot shows a web browser window with the URL https://annakrystalli.me/news-scrape/compile_db.html. The page has a dark header with the project name "news-scrape". Below the header, there are links for "Background", "Workflow", "Home", "About", and "License". The main content area is titled "Background" and describes the purpose of the section: "The purpose of this section is to collate the supplied NOW corpus files into a sql database. The supplied data comes in three table formats, as described in [this documentation](#)". It lists three tables: "1. a corpus table: containing words extracted from the online articles, one row for each occurrence (files in dir /dir1)", "2. a lexicon table: containing further details on the total set of words in the corpus, one row for each unique word (files in dir /dir2 /shared)", and "3. a sources table: containing the metadata on the articles from which words were extracted, one row for each unique article (sources files in dir /dir3 /shared)". A note at the bottom states: "N.B. original article data held in dirs /wp/ and /text/ are currently not being incorporated into the database." The "Workflow" section follows, with a note that it's designed to run as batch scripts on iceberg/sharc due to the large size of the files and eventual database. It includes instructions for running three scripts: "a01_prep-data.R" (copying and unzipping the data into an oommind1 user directory), "a02_make-db.R" (creating the database in an oommind1 user directory), and "a03_pop-db.R" (populating the database). A code block shows the command: "openS: 'emb://user:password@ofstore.ehet.ac.uk/shared/oommind1'".

blogdown

For creating and maintaining blogs through R.

Check out <https://awesome-blogdown.com/>, a curated list of awesome #rstats blogs in **blogdown** for inspiration!

presentations

A number of existing frameworks

xaringan 📦 Presentation Ninja 幻灯忍者 · 写轮眼



Managing code

Managing analysis code

Separate function definition and application

- When a project is new and shiny, an **analysis script usually contains many lines of directly executed code.**
- As it matures, **reusable chunks get pulled into their own functions.**
- The actual analysis scripts then become relatively short, and **functions defined in separate R scripts.**



R Package Structure

Used to share functionality with the R community

- Useful **conventions**
- Useful **software development tools**
- Easy **publishing** through GitHub



Build panel

Integrated graphical user interface

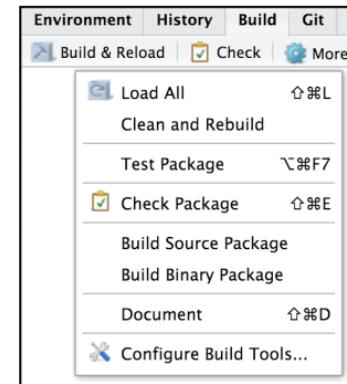
Package Writing



**File > New Project >
New Directory > R Package**

Turn project into package,
Enable roxygen documentation with
Tools > Project Options > Build Tools

Roxygen guide at
Help > Roxygen Quick Reference



R Package conventions:

- **metadata**: in the `DESCRIPTION` file
- **functions** in `.R` scripts in the `R/` folder
- **tests** in the `tests/` folder
- **Documentation:**
 - functions using **Roxygen notation**
 - workflows using `.Rmd` documents in the `vignettes/` folder

DESCRIPTION file

Package metadata

Package: gaitr

Type: Package

Title: Functions to support BMC gait analysis

Description: Helpers to analyse processed gait data.

Version: 0.0.9000

Authors@R:

```
c(person(given = "Anna",
          family = "Krystalli",
          role = c("aut", "cre"),
          email = "annakrystalli@googlemail.com"),
  person(given = "Lorenza",
          family = "Angelini",
          role = "aut",
          email = "l.angelini@sheffield.ac.uk"))
```

License: MIT + file LICENSE

citation

```
citation("gaitr")  
  
##  
## To cite package 'gaitr' in publications use:  
##  
##   Anna Krystalli and Lorenza Angelini (2019). gaitr: Functions to  
##   support BMC gait analysis. R package version 0.1.1.  
##   https://github.com/annakrystalli/gaitr  
##  
## A BibTeX entry for LaTeX users is  
##  
## @Manual{,  
##   title = {gaitr: Functions to support BMC gait analysis},  
##   author = {Anna Krystalli and Lorenza Angelini},  
##   year = {2019},  
##   note = {R package version 0.1.1},  
##   url = {https://github.com/annakrystalli/gaitr},  
## }
```

Dependency management

It's the job of the `DESCRIPTION` file to **list the packages that your code depends on.**

Imports:

```
dplyr,  
purrr,  
here,  
broom,  
tibble,  
magrittr,  
janitor,  
ggplot2
```

Suggests:

```
knitr,  
rmarkdown
```

add dependency

```
usethis::use_package("forcats", type = "Imports")
```

Functions in R/

example function script

Create a new function `.R` file in the `R/` folder

```
usethis::use_r("add")
```

```
R  
└ add.R
```

```
0 directories, 1 files
```

Document functions with Roxygen

- Create help files on build (autogenerated `.Rd` files in `man/`)
- Specify which functions are exported (autogenerated `NAMESPACE`)

```
#' Add together two numbers.  
#'  
#' @param x A number.  
#' @param y A number.  
#' @return The sum of x and y.  
#' @examples  
#' add(1, 1)  
#' add(10, 1)  
add <- function(x, y) {  
  x + y  
}
```

tests

Tests provide confidence in what the code is doing.



Example test

```
usethis::use_test("add")
```

Creates a `tests/` folder with the following files

```
tests
└── testthat
    └── test-add.R
        └── testthat.R
```

test-add.R

```
context("test-add")

test_that("add works", {
  expect_equal(add(2, 2), 4)
})
```

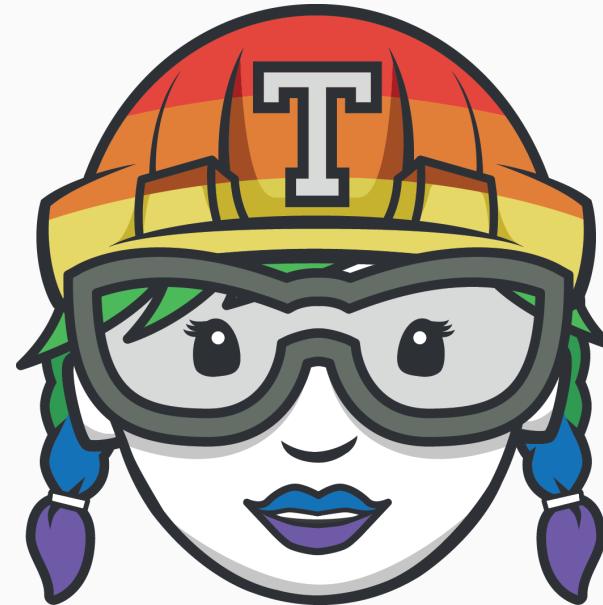
Continuous Integration w/ Travis

A cloud testing framework for automating your tests

- Monitor the effect of changes to the code
- Safe onboarding of contributions

Start with a `.travis.yml` file

```
usethis::use_travis()
```



.travis.yml

Resulting .travis.yml file template

```
language: R  
sudo: false  
cache: packages
```

instructions to enable TRAVIS CI

- ✓ Writing '.travis.yml'
- ✓ Adding '^\\.travis\\.yml\$' to '.Rbuildignore'
- Turn on travis for your repo at <https://travis-ci.org/profile/annakrystalli>
- Copy and paste the following lines into '/Users/Anna/Documents/workflows/talks/README.md'

```
[![Travis build status](https://travis-ci.org/annakrystalli/talks.svg?branch=master)](https://travis-ci.org/annakrystalli/talks)
```

Research compendia

A Research compendium

The paper is the advertisement

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

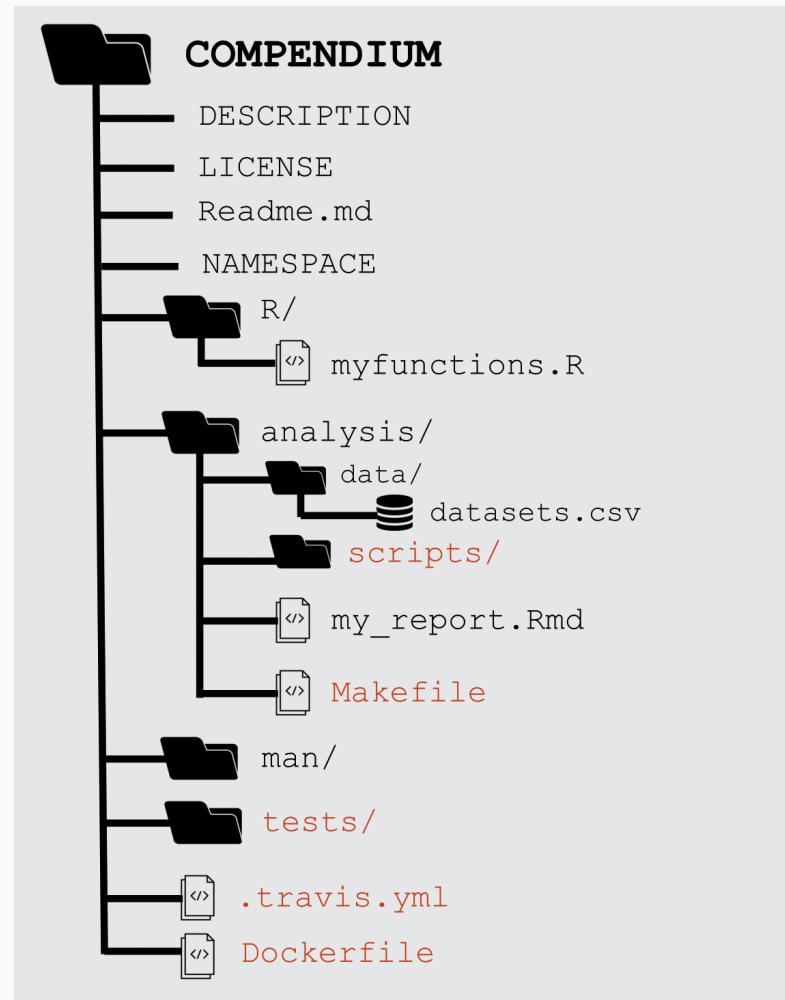
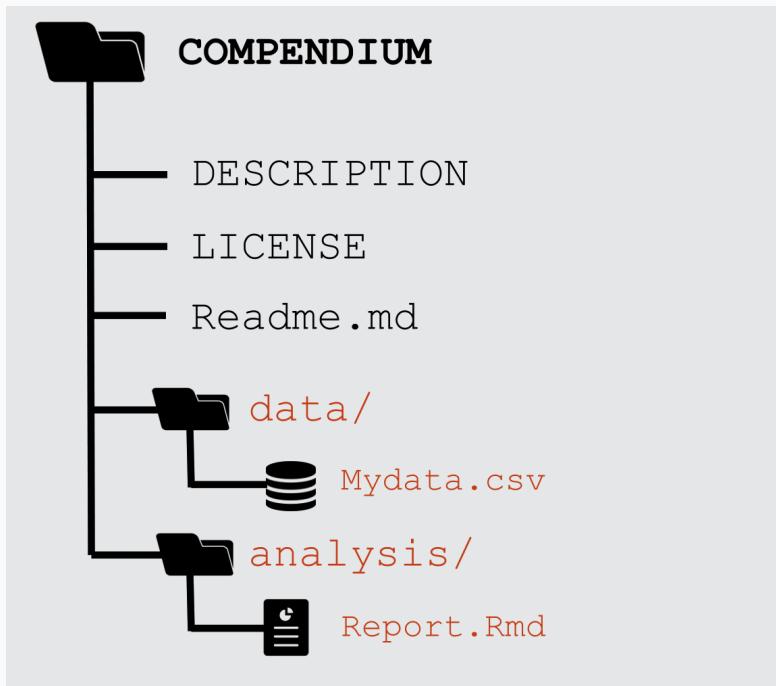
John Claerbout paraphrased in *Buckheit and Donoho (1995)*

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

Research compendia in R



Example compendium

Paper:

Boettiger, C. (2018) From noise to knowledge: how randomness generates novel phenomena and reveals information.

<https://doi.org/10.1111/ele.13085>

Compendium

cboettig/noise-phenomena: Supplement to: "From noise to knowledge: how randomness generates novel phenomena and reveals information"

<http://doi.org/10.5281/zenodo.1219780>

The screenshot shows a GitHub repository page for 'cboettig/noise-phenomena'. The repository has 119 commits, 2 branches, 4 releases, and 2 contributors. The commits list includes files like appendixA, appendixB, paper, reviews, .Rbuildignore, .drone.yml, .gitignore, .travis.yml, DESCRIPTION, Dockerfile, README.md, and noise-phenomena.Rproj, all submitted by cboettig. The repository is described as a 'Ecology Letters review paper: Noise-driven phenomena in population biology' with tags for research-compendium, reproducible-research, ecology, and stochasticity. The page also shows options to create new files, upload files, find files, clone or download, and a watch/unstar/fork button.

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

Prepare for GitHub

```
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`.

rrcompendium

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/\(DOI-to-paper\)](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](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](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/rrcompendium/archive/master.zip>

Or you can install this compendium as an R package, rrcompendium, 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, see <https://bookdown.org/yihui/bookdown/>
- Add bibliographic details of cited items to the 'references.bib' file
- For adding captions & cross-referencing in an Rmd, see <https://bookdown.org/yihui/bookdown/>
- For adding citations & reference lists in an Rmd, see <http://rmarkdown.rstudio.com/authors.html>

paper.Rmd to paper.pdf

Rmd

pdf

```
1 # ...
2 # title: "From noise to knowledge: how randomness generates novel phenomena and reveals information"
3 # author:
4 #   - name: "Carl Boettiger"
5 #   - affiliation: a
6 #     - email: "cboettig@berkeley.edu"
7 # address: b
8 #   - code: o
9 #   - address: "Dept of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley CA 94720-3114, USA"
10 #!csl: trends-journals.csl
11 #!bib: ecology-letters.bib
12 #!bib: biology-phylo.bib
13 #!output:
14 #   titles::elsevier-article:
15 #     keep-text: true
16 #   notes: true
17 #   default
18 # layout: 3p # review = doublespace, 3p = singlespace, 5p = two-column
19 # preamble: l
20 # \newcommand{\doi}[1]{\text{DOI:}\,[#1]}
21 # \usepackage{color}
22 # \journal[Ecology Letters]
23 # \linenumbers
24 # \usepackage{setspace}
25 # \doublespacing
26 # abstract:
27 # Article Type: Review & Synthesis
28 # Running title: From noise to knowledge
29 # Figures: 0
30 # Tables: 0
31 # Word count: 7430
32 # Word count in abstract: 181
33 # Number of references: 111
34 # All data (simulated data and code) is made available on GitHub (https://github.com/cboettig/noise-phenomena), and if accepted, will be archived with provided DOI in an appropriate scientific data repository.
35
36 # Keywords: stochasticity, demographic noise, environmental noise, colored noise, quasi-cycles, tipping points
37 # Abstract
38
39 # Noise, as the term itself suggests, is not often seen as a source to be reckoned with. In contrast, the notion that noise can be studied over time to reveal processes of interest is fundamental to scientific inquiry. While well-known in some fields, noise is often overlooked in its most right: noise can induce novel phenomena that could not be understood from some underlying deterministic model alone. Nor is all noise the same, and close examination of differences in frequency, color or magnitude can reveal insights that would otherwise be inaccessible. Yet with each aspect of stochasticity leading to some new or unexpected behavior, the time is right to move beyond the familiar refrain of "everybody is important" (#BJorntstad2003). Stochastic phenomena can suggest new ways of inferring causality, can reveal new insights about biological systems, and can even change our perspective about what is possible. While this might feel compelling examples, however, we observe that the study of stochastic phenomena is only beginning to make this translation into empirical inference. There are rich opportunities at this interface in the years ahead.
```

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)
```

Further Helpers



rticles

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



citr

RStudio Add-in to Insert Markdown Citations

The screenshot shows the RStudio interface with the 'paper.Rmd' file open. The code editor displays a section of text discussing stochastic population models, mentioning Bartlett (1960). A modal dialog titled 'Insert citation' is centered over the editor, containing the citation text 'Bartlett (1960). Stochastic population models in ecology and epidemiology.' and the corresponding BibTeX entry '[@Bartlett1960]'. A checked checkbox labeled 'In parentheses' is present. Below the dialog, the R Markdown editor shows the YAML front matter with a bibliography file specified as 'refs.bib'. The R console at the bottom shows the command 'citr::insert_citation()' being run twice, resulting in the message 'Listening on http://127.0.0.1:7464'.

Reproducible Computational Environments

Why isn't sharing code enough?

Case Study: Sharing a Geospatial Analysis in R

On a computer without System Library **GDAL** 

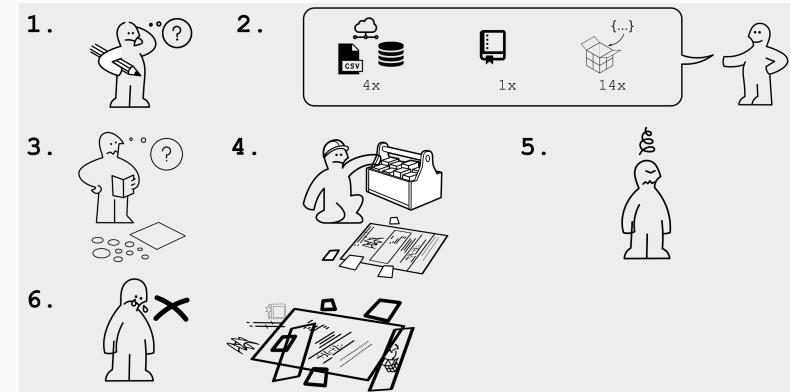
package 'rgdal' successfully unpacked
and MD5 sums checked

```
configure: gdal-config: gdal-config  
checking gdal-config usability ... ./conf  
line 1353: gdal-config: command not foun  
no
```

```
Error: gdal-config not found
```

```
...
```

```
ERROR: configuration failed for  
package 'rgdal'
```



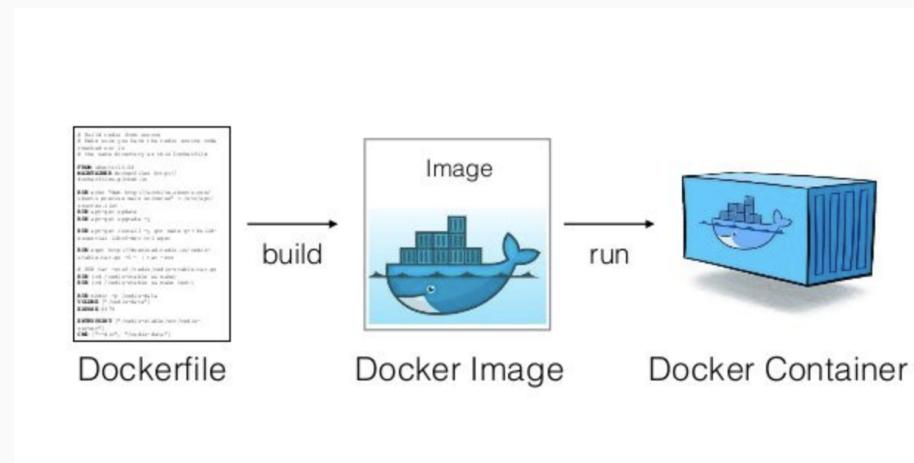
slide: *Karthik Ram: rstudio::conf 2019 talk*

What are Docker containers?

standardized units of software

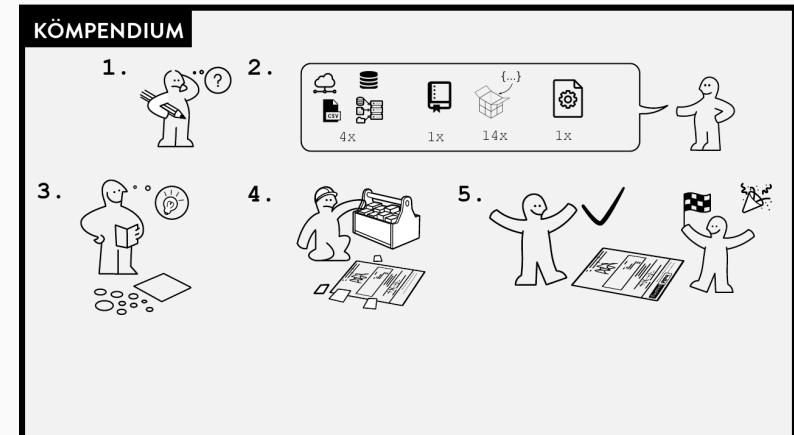
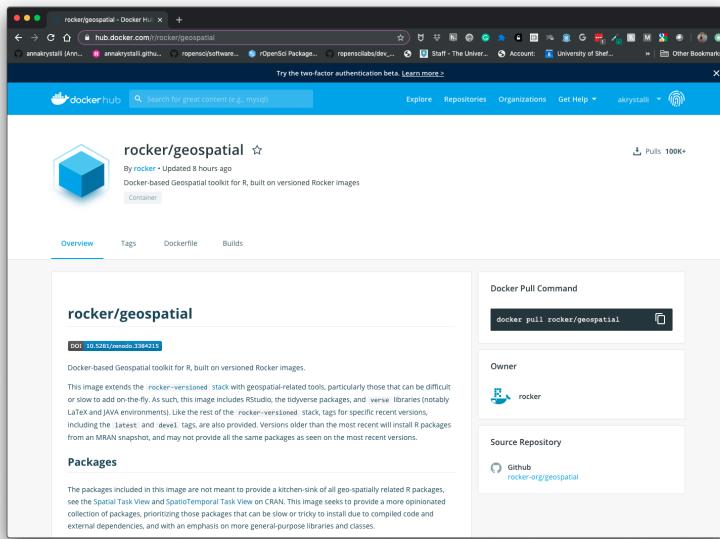
package up everything needed to run an application: code, runtime, system tools, system libraries and settings in a lightweight, standalone, executable package

- **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 libraries that are required for the R package to run
- * Check the last line of the dockerfile to specify which Rmd should be rendered in the container

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/rrcompendiumDTB.svg?branch=master>)

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

* To connect Docker, go to <https://travis-ci.org/>, and add your environment variables: DOCKER_EMAIL, DOCKER_USER, DOCKER_PASS to enable pushing to the 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 $TRAVIS_BRANCH-$TRAVIS_BUILD_NUMBER; 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

Environment Variables

Customize your build using environment variables. For secure tips on generating private keys [read our documentation](#)

DOCKER_PASS	 <small>Secure variable</small>	Available to all branches	
DOCKER_USER	 <small>Secure variable</small>	Available to all branches	

Travis build passes!

annakrystalli / rrcompendiumDTB

build passing

Current Branches Build History Pull Requests More options

master correct compendium name on README

#12 passed

Commit 90157b3 ↗ Ran for 5 min 25 sec

Compare 3ea62f9..90157b3 ↗ about 2 hours ago

Branch master ↗

Anna Krystalli

no language set

AMD64

build passing

Image on Dockerhub

The screenshot shows the Dockerhub profile of user **akrystalli**. The profile includes a blue fingerprint icon, the username **akrystalli**, a link to [Edit profile](#), a status as a **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 repository list displays 3 of 3 repositories, with the first one being **akrystalli/rrcompendiumdtb**, created by **akrystalli** and updated a few seconds ago. It is categorized as a **Container** and has 1 download. A small grey 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>

Working with the Docker Image

Pull it from Dockerhub.

```
docker pull akrystalli/rrcompendiumdtb:latest
```

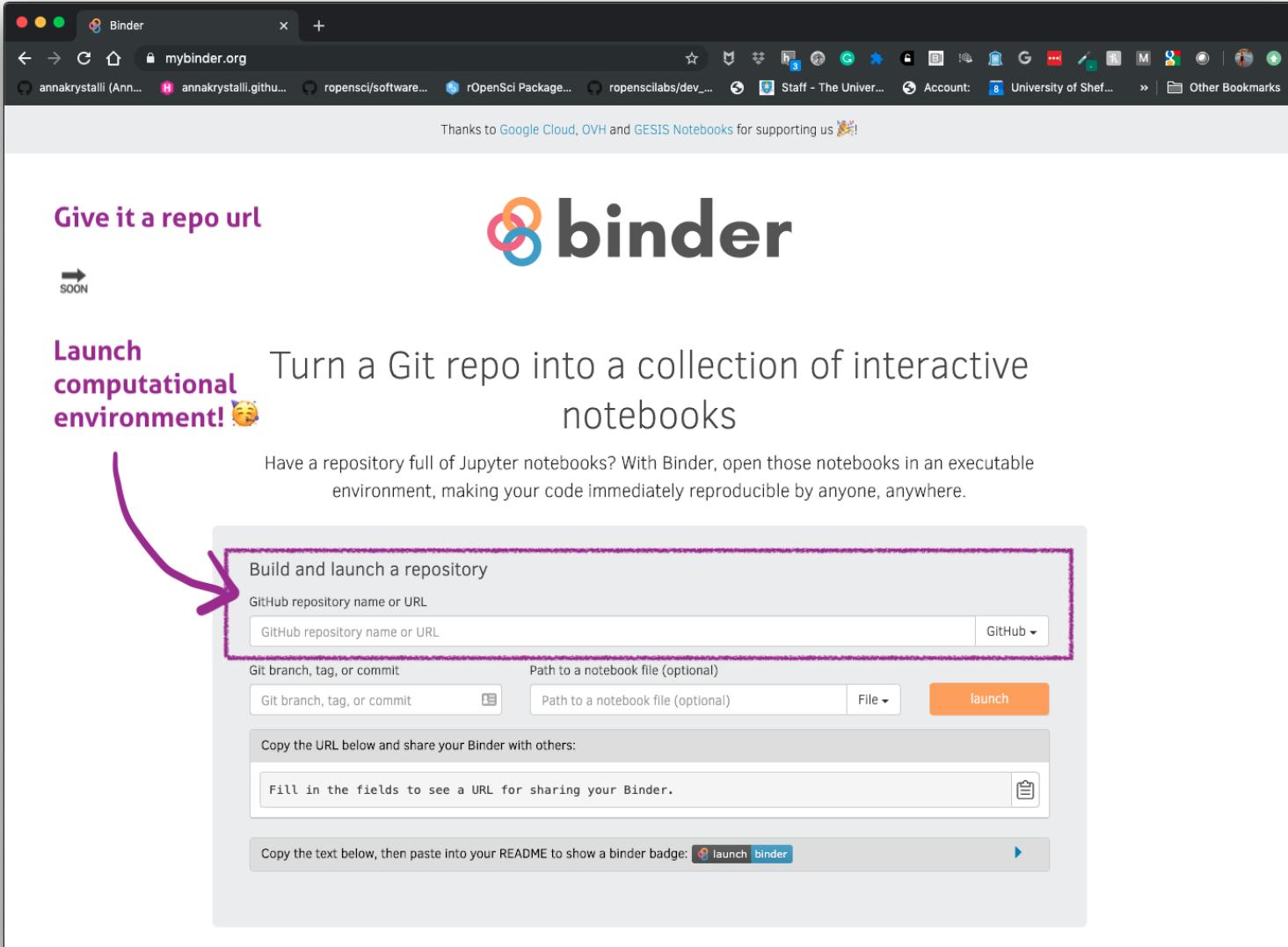
Run in Rstudio in your browser

More on [using the RStudio image](#)

```
docker container run -e PASSWORD=password -e USERID=$UID -p  
8787:8787 --detach --name rrcompendiumdtb  
akrystalli/rrcompendiumdtb:latest
```

Binder

What is binder? <https://mybinder.org/>



The screenshot shows the mybinder.org homepage. At the top, there's a navigation bar with tabs like 'mybinder.org', 'Staff - The Univer...', 'Account:', and 'University of Shef...'. Below the navigation, it says 'Thanks to Google Cloud, OVH and GESIS Notebooks for supporting us!'. The main heading is 'Give it a repo url' with a 'SOON' button next to it. To the right is the 'binder' logo. Below the heading, there's a call-to-action: 'Launch computational environment!' followed by a smiling face emoji. The text explains: 'Turn a Git repo into a collection of interactive notebooks'. A purple arrow points from the left towards the 'GitHub repository name or URL' input field, which is highlighted with a purple border. Below this field are options for 'Git branch, tag, or commit' and 'Path to a notebook file (optional)'. There's also a 'launch' button. Further down, there's a section for sharing the URL and another for adding a binder badge to a README.

Give it a repo url

SOON

Launch computational environment!

Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

Build and launch a repository

GitHub repository name or URL

Git branch, tag, or commit Path to a notebook file (optional)

File

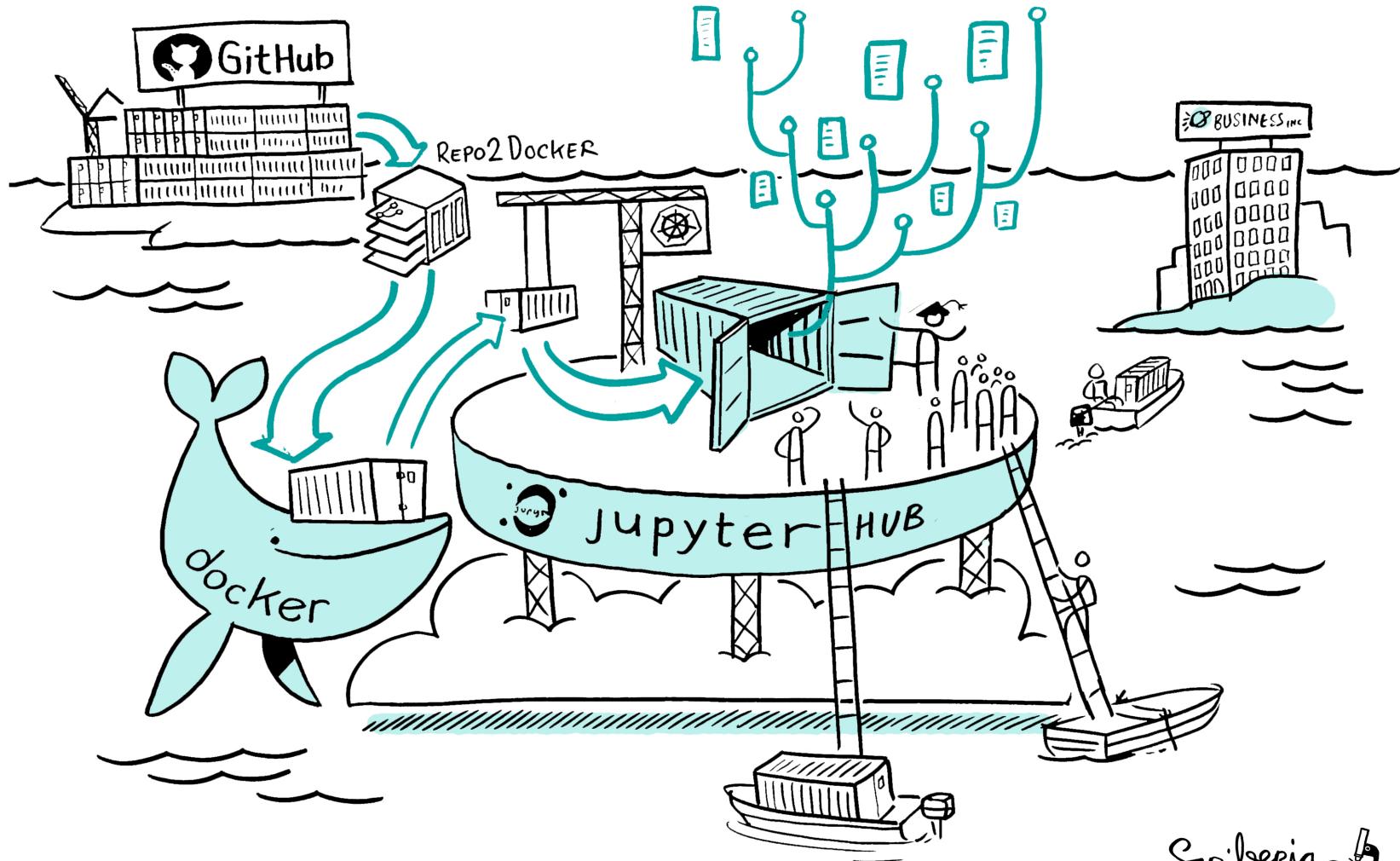
Copy the URL below and share your Binder with others:

Fill in the fields to see a URL for sharing your Binder.

Copy the text below, then paste into your README to show a binder badge:

Binderhub Ecosystem

This image was created by Scriberia for The Turing Way community and is used under a CC-BY licence.



Binder for R

The diagram illustrates the workflow for creating a Binder repository from an R project:

- Top Left:** A screenshot of an RStudio desktop interface showing an R console window. The console output includes R version 3.6.0 details, a note about no warranty, and information about being a collaborative project. A large yellow "BAM!" comic-style explosion graphic is overlaid on the left side of the screen.
- Top Right:** A screenshot of an RStudio desktop interface showing the file browser. It displays a directory structure for a repository named "karthik/friday-test/master". The files listed include .gitignore, analysis.R, binder-test.Rproj, DESCRIPTION, DESCRIPTION.I, kitematic, and README.md. An "OMG!" comic-style explosion graphic is overlaid on the right side of the screen.
- Bottom Left:** A screenshot of a GitHub repository page for "karthik/friday-test". The repository description is "Example repo for dataRetrieval package". A "launch binder" button is highlighted with a blue arrow. The "README.md" file is shown. The "Build logs" section shows the server launching.
- Bottom Center:** A list of steps to create a Binder repository:
 1. Click Use this template to the right of the "Use this template" button.
 2. Give this repo a new name and create a new repo in your account.
 3. Click the Clone or download button, copy the URL.
 4. in RStudio Desktop, click the Project drop down on the top right, Choose New Project > Version Control > Git, and

R repository options for Binder

slide: **Karthik Ram: rstudio::conf 2019 talk**

Basic	free	Premium	free	Pro	free
<i>install.r</i> <i>runtime.txt</i> <i>apt.txt</i>		<i>Dockerfile</i> <i>install.r</i>		<i>Dockerfile</i> <i>DESCRIPTION</i>	
Slow but easy to setup. Recommended for beginners		Faster launch		Best for compendia	



launch binder



launch binder



launch binder

Binderise your R projects w/ holepunch

<https://github.com/karthik/holepunch>

```
remotes::install_github("karthik/holepunch")
```

Create .binder/Dockerfile

```
holepunch::write_dockerfile(maintainer = "Anna Krystalli")
```

```
→ Setting R version to 3.6.0
→ Locking packages down at 2019-11-09
✓ Dockerfile generated at ./binder/Dockerfile
```

.binder/Dockerfile

```
FROM rocker/binder:3.6.0
LABEL maintainer='Anna Krystalli'
USER root
COPY . ${HOME}
RUN chown -R ${NB_USER} ${HOME}
USER ${NB_USER}

RUN wget https://github.com/annakrystalli/rrcompendiumDTB/raw/master/DESCRIPTION && R
devtools::install_deps(); devtools::install(); tinytex::install_tinytex()

RUN rm DESCRIPTION.1; exit 0
```

Add binder README badge

```
holepunch::generate_badge()
```

- Copy and paste the following lines into '[/Users/Anna/Documents/workflows/compendia/rrcompendiumDB/README.md](#)'
[![[Launch Rstudio Binder]](http://mybinder.org/badge_logo.svg)](<https://mybinder.org/badge.html?url=https://github.com/krystalli/rrcompendiumDB>)
[Copied to clipboard]



README.md

rrcompendiumDB

[build](#) passing [launch](#) binder

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 09 Nov 2019. Online at <https://doi.org/{DOI-to-compendium}>

Launched Binderised Compendium

The image shows a desktop interface with three main windows:

- RStudio:** On the left, the RStudio interface is visible. It has a top menu bar with File, Edit, Code, View, Plots, Session, Build, Debug, Project, Help, and Tools. Below the menu is a toolbar with icons for file operations like Open, Save, and Print. A sidebar on the left contains a tree view of files and a search bar labeled "Go to file/function". The main workspace shows an R Markdown file named "paper.Rmd". The code includes a YAML header with metadata (title, author, address) and R code for knitting the document. The R console at the bottom shows the R version (3.6.0), copyright information, and a welcome message about the software being free and redistributable.
- PDF Viewer:** In the center, a window titled "RStudio: View PDF" displays a multi-page document. The first page is a title page with the title "From noise to knowledge: how randomness generates novel phenomena and reveals information" by Carl Boettiger. Subsequent pages show abstracts and text from the document, with some sections highlighted in purple. The PDF viewer has a standard toolbar with icons for zooming, navigating, and saving.
- File Browser:** On the right, a file browser window titled "Inbox (10) - a.krys..." is open. It shows a list of files with columns for Name, Size, and Modified. The files listed are:

Name	Size	Modified
21.6 KB	Nov 9, 2019, 1:53 PM	
603 B	Nov 9, 2019, 1:53 PM	
6 KB	Nov 9, 2019, 1:53 PM	
512 B	Nov 9, 2019, 2:00 PM	
308.8 KB	Nov 9, 2019, 2:00 PM	
13.4 KB	Nov 9, 2019, 1:53 PM	
0 B	Nov 9, 2019, 2:00 PM	
18.2 KB	Nov 9, 2019, 1:59 PM	
53.8 KB	Nov 9, 2019, 1:53 PM	

Reproducibility in Practice

ReproHacks

one day reproducibility hackathons

Mission: Reproduce papers from code and data

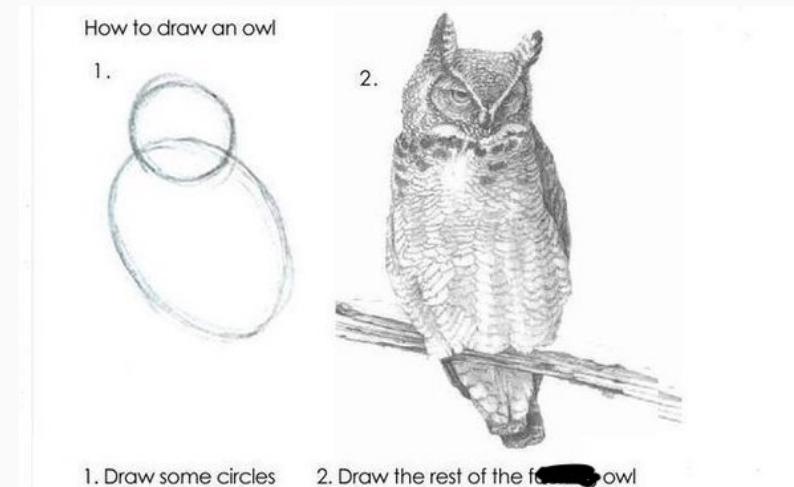
- Record experiences and feedback to authors
- Available soon: Publish Reproducibility Report in ReScience C



ReproHack Benefits

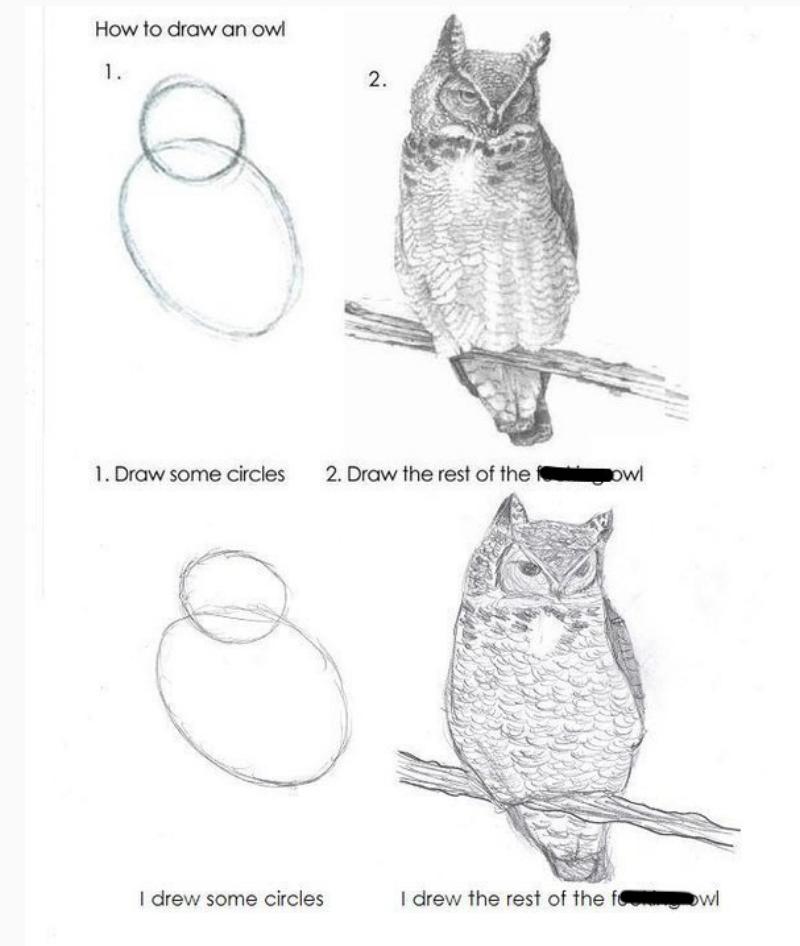
Participants

1. Practical reproducibility they can implement in their own work



Participants

1. Practical reproducibility they can implement in their own work



Participants

1. Practical reproducibility they can implement in their own work
2. Inspiration from working with other people's code and data.

CHIBIRD

Participants

1. Practical reproducibility they can implement in their own work
2. Inspiration from working with other people's code and data.
3. Reproduction as community value



Authors

1. Useful feedback on the reproducibility of their work

OpenCon ReproHack feedback form

Name of participant
Anna Krystalli, Marios Georgiou

Which paper did you attempt? *

Paper #3. The archaeology, chronology and stratigraphy of Madjedbebe (Malakunjanja II): a site in northern Australia with early occupation

Team issue URL *

https://github.com/OpenCon-London/OpenCon_London-Doathon/issues/4

Did you manage to reproduce it? *

Yes
 No
 Almost

On a scale of 1 to 10, how much of the paper did you manage to reproduce? *

1 2 3 4 5 6 7 8 9 10

None of it All of it

Authors

1. Useful feedback on the reproducibility of their work
2. Appreciation for their efforts in making their work reproducible



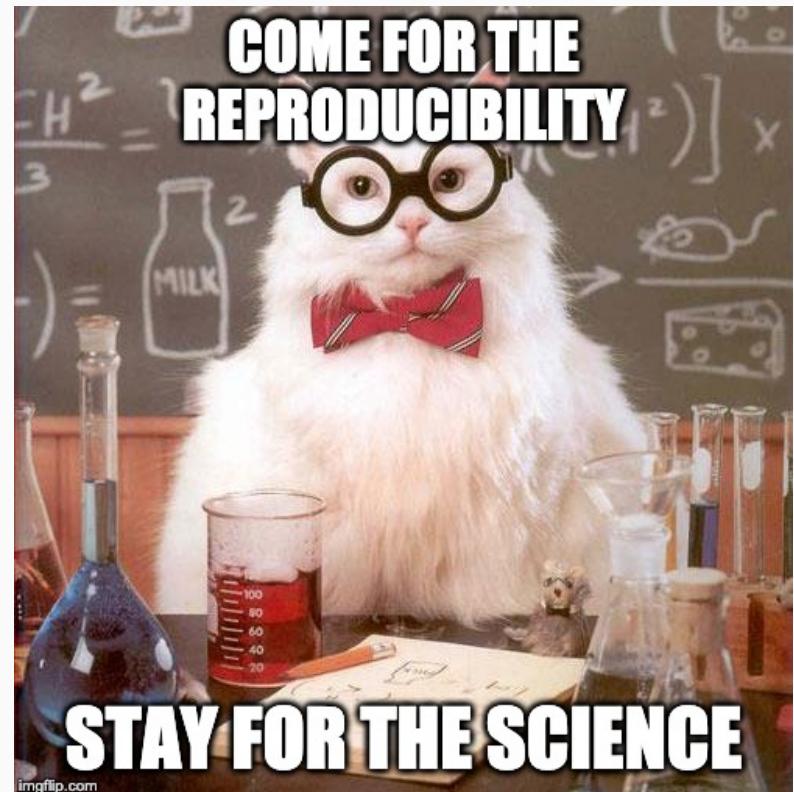
Authors

1. Useful feedback on the reproducibility of their work
2. Appreciation for their efforts in making their work reproducible

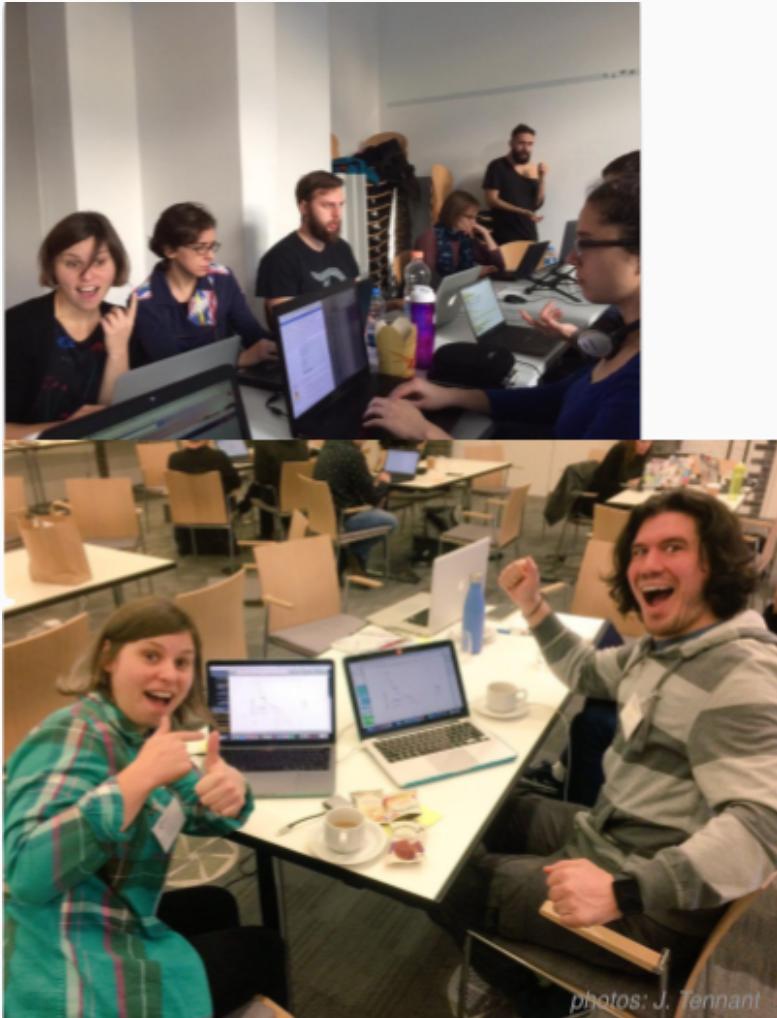


Authors

1. Useful feedback on the reproducibility of their work
2. Appreciation for their efforts in making their work reproducible
3. An opportunity to engage others with their research.



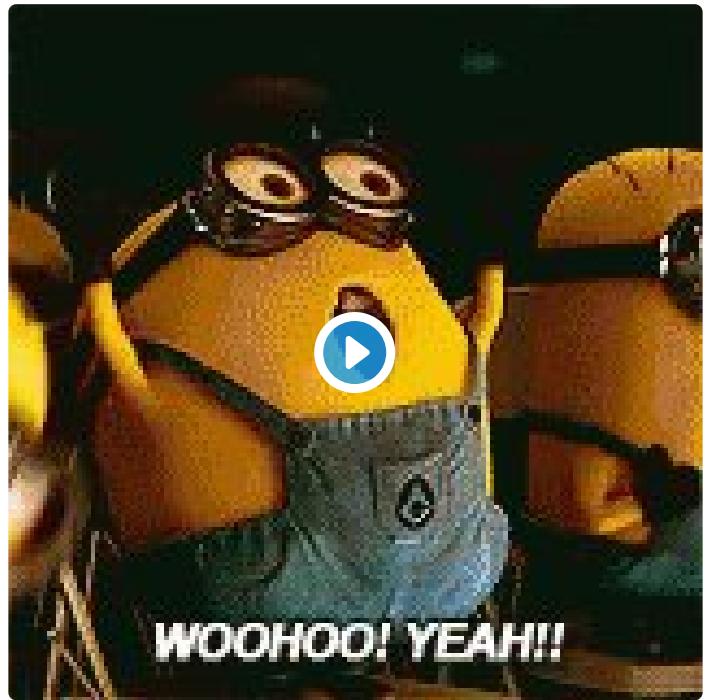
Reproducing papers is fun!!



Jessica Ward
@JKRWard



Huge thanks to @EnviroKaty for submitting a fab
🦋🦋🦋 paper to the #CCMcr19 #ReproHack!
I had loads of fun reproducing the analysis for
this really cool paper doi.org/10.3390/insect...



♡ 12 3:41 PM - Jun 27, 2019

102 / 112

Upcoming ReproHacks!

 **ReproHack** 
@ReproHack

This thing is really happening!



Twitter icon

Heart icon 16 11:48 AM - Sep 20, 2019

i icon

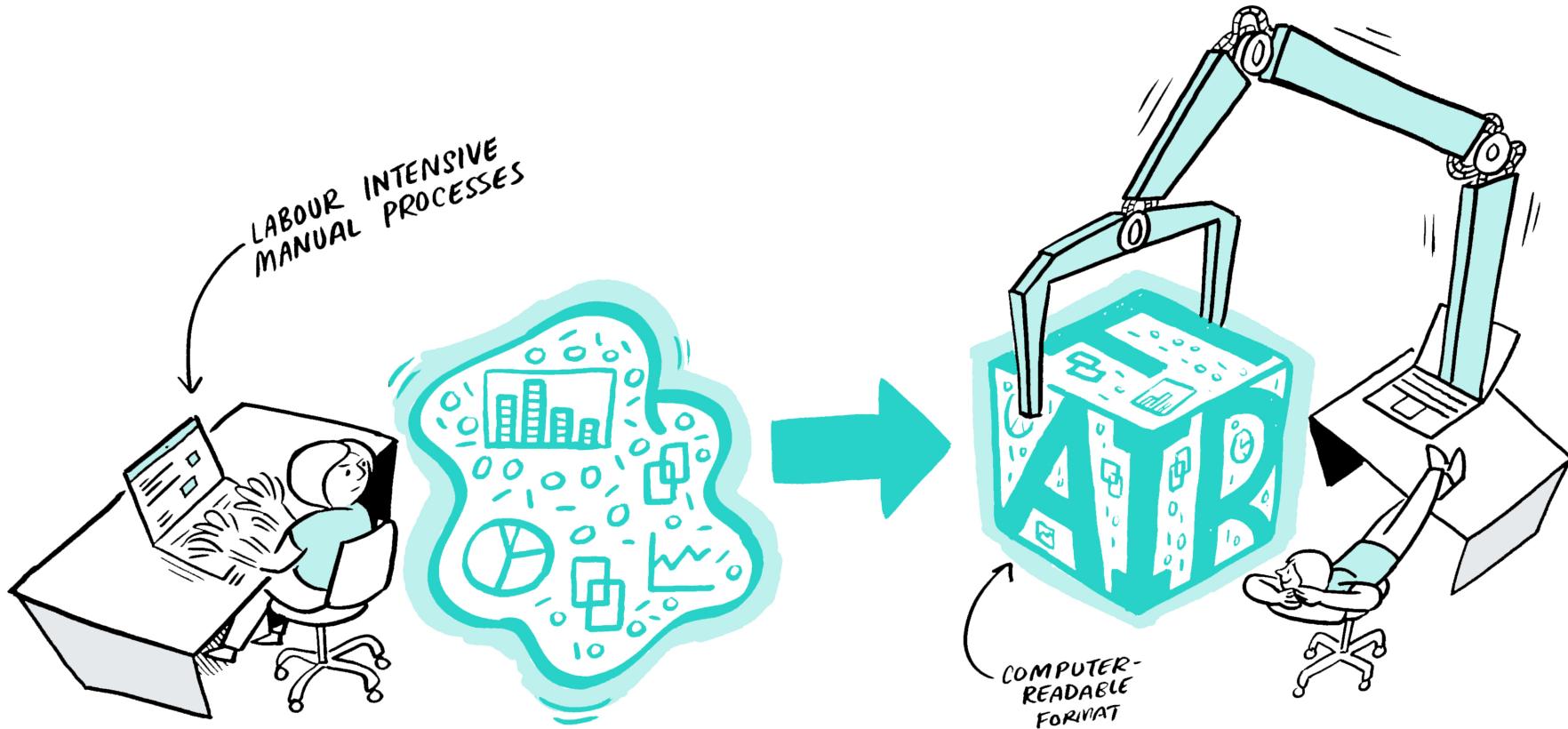
See ReproHack's other Tweets >



Take home

Following conventions ➔

This image was created by Scriberia for The Turing Way community and is used under a CC-BY licence.



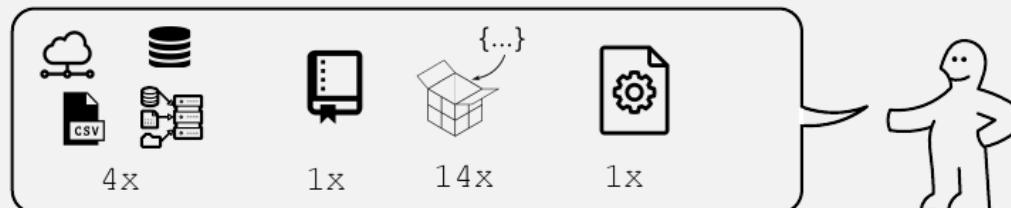
Successful Reproducibility ➔

KÖMPENDIUM

1.



2.



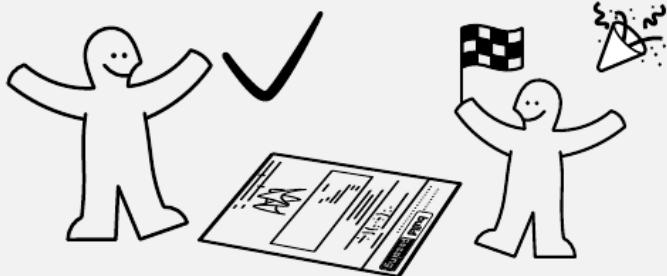
3.



4.



5.

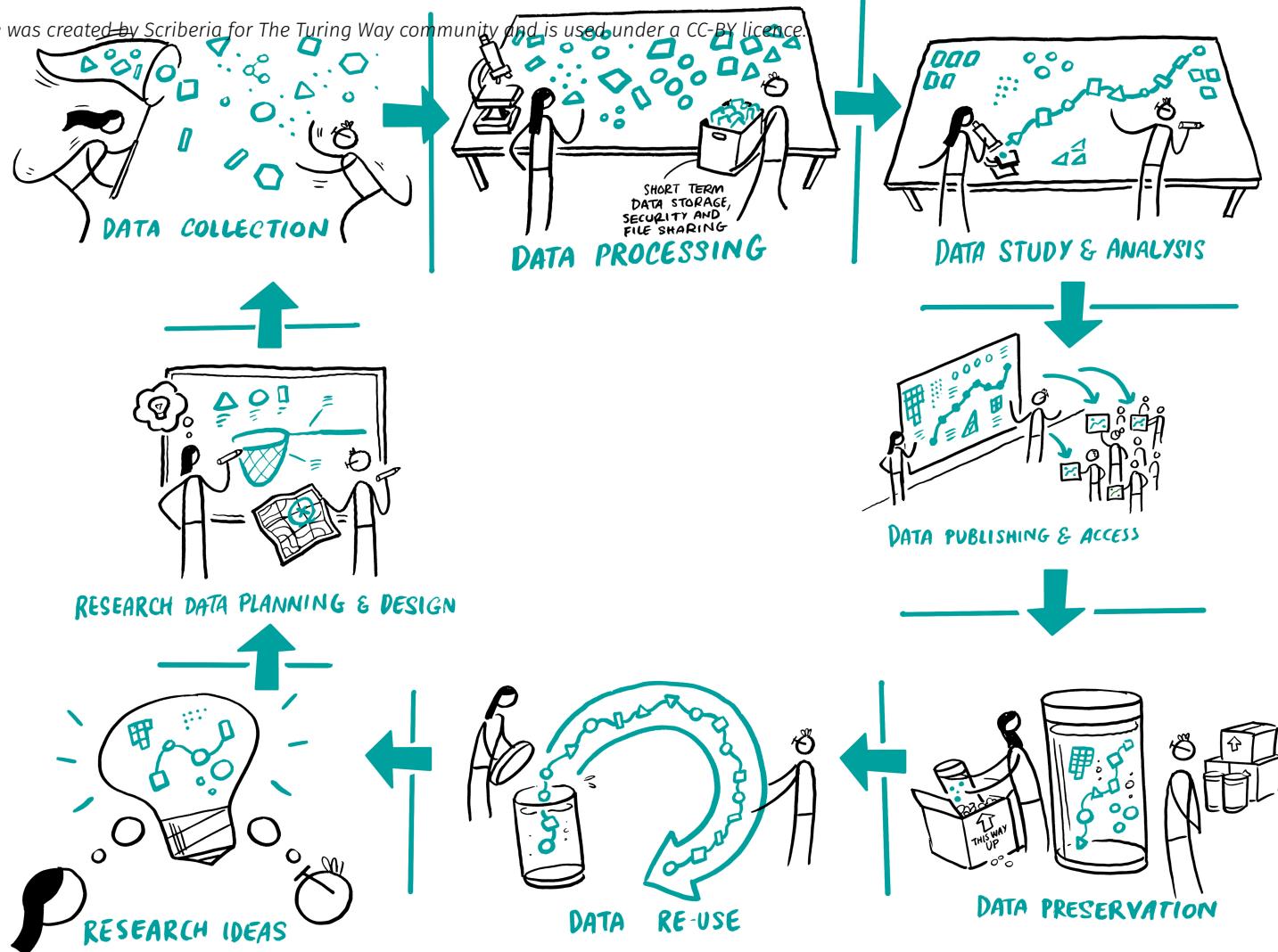


slide: **Karthik Ram: rStudio::conf 2019 talk**

Enhanced Research Cycle



This image was created by Scriberia for The Turing Way community and is used under a CC-BY licence.



Reproducibility as standard



This image was created by Scriberia for The Turing Way community and is used under a CC-BY licence.



Scriberia

Resources

The Turing Way

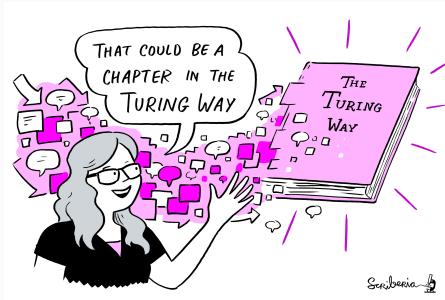
Book

a lightly opinionated guide to reproducible
data science

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

workshops

- **Boost Your Research Reproducibility with Binder** [materials](#)
- **Build a binderhub** [materials](#)



<https://github.com/alan-turing-institute/the-turing-way>

Reproducibility in R

Version Control

- Happy Git and GitHub for the useR

RMarkdown

- R Markdown: The Definitive guide
- RMarkdown Driven Development (RmdDD): Blog post by Emily Riederer

R Packages

- R packages by Hadley Wickham and Jenny Bryan

Research Compendia

- Karthik Ram: *rstudio::conf 2019 talk*

Docker & Binder

- Getting started with binder [docs](#)
- rOpenSci [Docker tutorial](#)

Tutorials

- Rstudio Essentials Webinar series
- rrresearch: ACCE DTP course on Research Data & Project Management

ReproHacks

rephack-hq repository

Check out our [issues](#)

Chat to us:

[chat on gitter](#)

Sign up to our Newsletter

Enter your email address

[Subscribe](#)

powered by [TinyLetter](#)

Questions?