Reproducible Research in R

w/ rrtools

Harnessing the power of Convention

✨

Anna Krystalli 🐦 *@annakrystalli*

3rd UK RSE Conference, 2019-09-17

👋 Hello and welcome

me: Dr Anna Krystalli

- Research Software Engineer, *University of Sheffield*

  - twitter @annakrystalli
  - github @annakrystalli
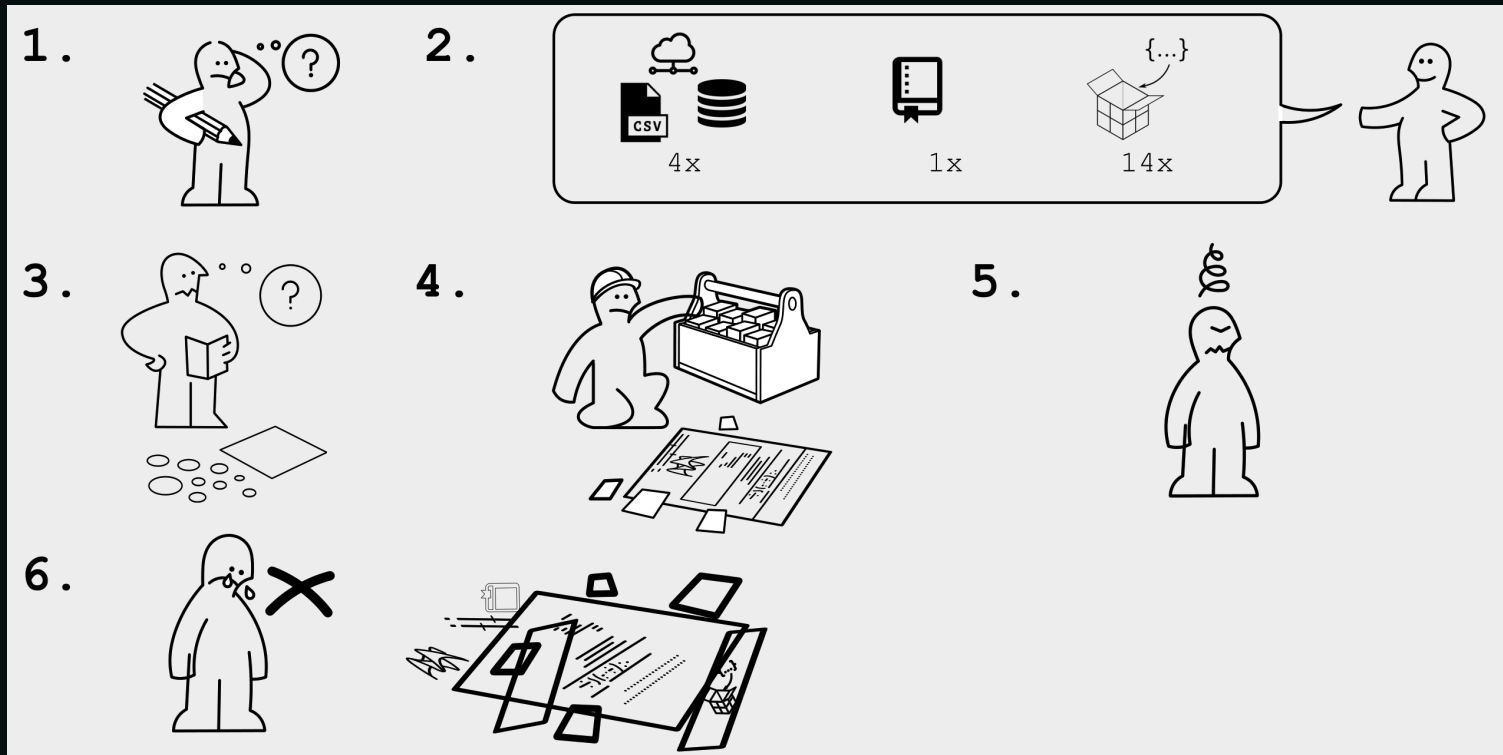 - email a.krystalli[at]sheffield.ac.uk

- Editor *rOpenSci*

Background

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)

Is published code and data enough?

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

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

The Research Compendium in R

> R packages can be used as a research compendium for organising and sharing files!

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

✨ Harnessing the power of Convention! ✨

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

Enter `rrtools`

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

`rrtools` demo

Create a research compendium and recreate a research paper from data, analysis code and text using `rrtools` and friends (eg `usethis`).

demo materials - source: `cboettig/noise-phenomena`

Subset of materials from the published compendium of code, data, and author's manuscript:

Carl Boettiger. (2018, April 17). *cboettig/noise-phenomena: Supplement to: "From noise to knowledge: how randomness generates novel phenomena and reveals information"* (Version revision-2). *Zenodo.* http://doi.org/10.5281/zenodo.1219780

accompanying the publication:

Carl Boettiger [iD]. *From noise to knowledge: how randomness generates novel phenomena and reveals information*. Published in *Ecology Letters*, 22 May 2018 https://doi.org/10.1111/ele.13085
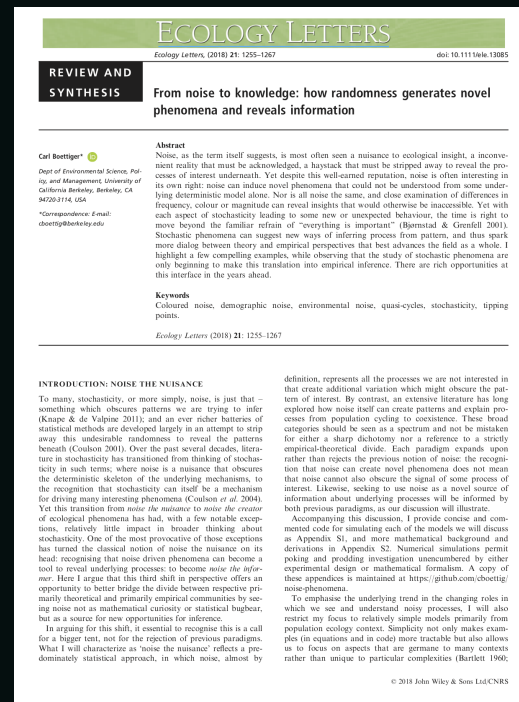
(Semi-) Live demo

In the interest of time, some copying & pasting
will be used

Setup

install LaTeX

Only required if you don't have LaTeX installed

```r
install.packages('tinytex')
tinytex::install_tinytex()
```

On Windows

You might need to install Rtools

installed workflow dependencies

```r
source("install.R")
```

```r
install.packages(c(
    # accesing remote repositories
    "remotes",
    #' analysis
    "dplyr", "ggplot2", "ggthemes", "here",
    #' bibliographic / publishing
    "citr", "rticles", "rmarkdown",
    #' documentation
    "roxygen2",
    #' graphics
    "Cairo"))
```

installed `rrtools`

```r
remotes::install_github("benmarwick/rrtools")
```

Opened Rstudio

in project **rrtools-rse19** (bit.ly/rrtools-rse19)

loaded some libraries

```
library(rrtools)
library(usethis)
library(testthat)
```

Got materials

```r
usethis::use_course("bit.ly/rrtools_wks")
```

```
~/Desktop/rrtools-wkshp-materials-master/
├── README.md
├── analysis.R
├── gillespie.csv
├── paper.pdf
├── paper.txt
└── refs.bib
```

analysis.R

Load Libraries

```r
library(dplyr)
library(readr)
library(ggplot2)
library(ggthemes)
```

Set theme

```r
theme_set(theme_grey())
```

analysis.R

Load and process data

```r
# create colour palette
colours <- ptol_pal()(2)

# load-data
data <- read_csv(here::here("gillespie.csv"), col_types = "cdiddd")

# recode-data
data <- data %>%
  mutate(system_size = recode(system_size,
                               large = "A. 1000 total sites",
                               small = "B. 100 total sites"))
```
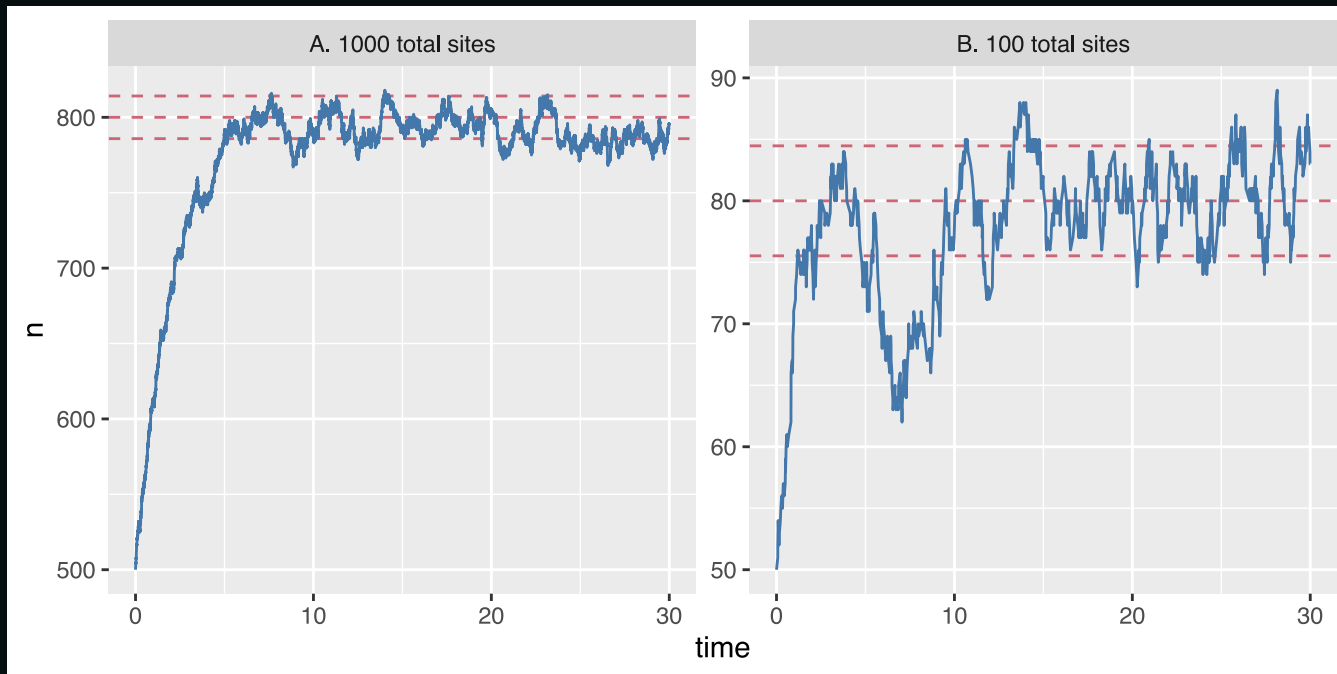
analysis.R

plot

```r
# plot-gillespie
data %>%
  ggplot(aes(x = time)) +
  geom_hline(aes(yintercept = mean), lty=2, col=colours[2]) +
  geom_hline(aes(yintercept = minus_sd), lty=2, col=colours[2]) +
  geom_hline(aes(yintercept = plus_sd), lty=2, col=colours[2]) +
  geom_line(aes(y = n), col=colours[1]) +
  facet_wrap(~system_size, scales = "free_y")
```

analysis.R

plot

global user settings

.Rprofile

```r
usethis::edit_r_profile()
```

```r
options(usethis.full_name = "Anna Krystalli",
        servr.daemon = TRUE,
        pkgType = "binary",
        usethis.description =
            list(`Authors@R` =
                    'person(
                        given = "Anna",
                        family = "Krystalli",
                        role = c("aut", "cre"),
                        email = "annakrystalli@googlemail.com",
                        comment = c(ORCID = "0000-0002-2378-4915"))',
                Version = "0.0.0.9000"
            ),
        usethis.protocol  = "https"
)
```

.Renviron

Get GitHub PAT

```
usethis::browse_github_pat()
✔ Opening URL 'https://github.com/settings/tokens/new?scopes=repo,gis
● Call `usethis::edit_r_environ()` to open '.Renviron'.
● Store your PAT with a line like:
  GITHUB_PAT=xxxyyyzzz
  [Copied to clipboard]
● Make sure '.Renviron' ends with a newline!
```

Edit file

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

```
GITHUB_PAT=f0f766313811965a5064174bd919bc770e067ce6
```

Not this one 😉

Let's go! 🏁

follow along at `workflow.R`

Demo Links

demo materials

bit.ly/rrtools_wks

workflow code

bit.ly/rrtools-workflow

outputs

- Example compendium `rrcompendiumRSE19`
- Example Docker image of compendium

`rrtools` GH repo

benmarwick/rrtools

walkthrough GH repo

annakrystalli/rrtools-rse19

Thanks for 👀 👋

Acknowledgements