Reproducible Research in R w/ rrtools

harnessing the power of convention

Anna Krystalli

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*

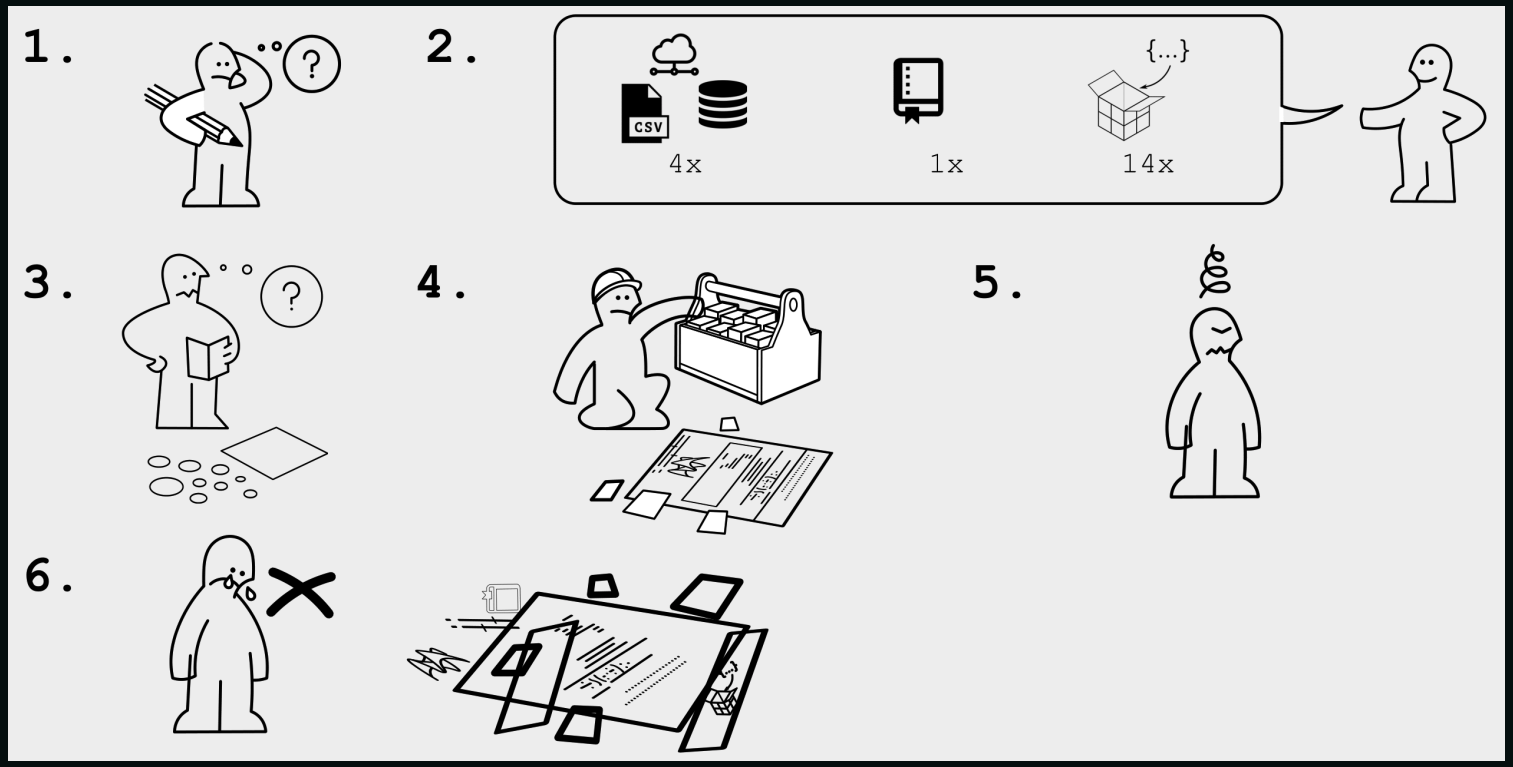
Background

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."

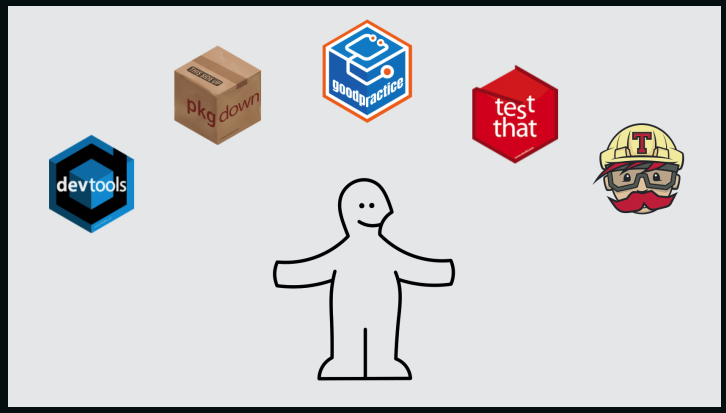
Gentleman and Temple Lang, 2004

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.

`rrtools` demo

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

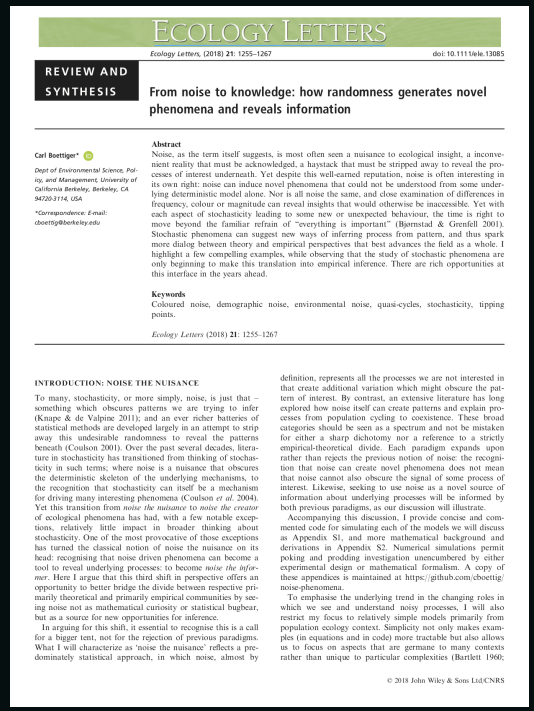
demo materials - source: `cboettig/noise-phenomena`

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. *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



ECOLOGY LETTERS

Ecology Letters, (2018) **21**: 1255–1267 doi: 10.1111/ele.13085

REVIEW AND SYNTHESIS

From noise to knowledge: how randomness generates novel phenomena and reveals information

Carl Boettiger*

Dept of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, CA 94720-3114, USA

*Correspondence: E-mail: cboettig@berkeley.edu

Abstract

Noise, as the term itself suggests, is most often seen a nuisance to ecological insight, a inconvenient reality that must be acknowledged, a haystack that must be stripped away to reveal the processes of interest underneath. Yet despite this well-earned reputation, noise is often interesting in its own 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, colour or magnitude can reveal insights that would otherwise be inaccessible. Yet with each aspect of stochasticity leading to some new or unexpected behaviour, the time is right to move beyond the familiar refrain of "everything is important" (Bjørnstad & Grenfell 2001). Stochastic phenomena can suggest new ways of inferring process from pattern, and thus spark more dialog between theory and empirical perspectives that best advances the field as a whole. I highlight a few compelling examples, while observing that the study of stochastic phenomena are only beginning to make this translation into empirical inference. There are rich opportunities at this interface in the years ahead.

Keywords

Coloured noise, demographic noise, environmental noise, quasi-cycles, stochasticity, tipping points.

Ecology Letters (2018) **21**: 1255–1267

INTRODUCTION: NOISE THE NUISANCE

To many, stochasticity, or more simply, noise, is just that – something which obscures patterns we are trying to infer (Knape & de Valpine 2011); and an ever richer batteries of statistical methods are developed largely in an attempt to strip away this undesirable randomness to reveal the patterns beneath (Coulson 2001). Over the past several decades, literature in stochasticity has transitioned from thinking of stochasticity in such terms; where noise is a nuisance that obscures the deterministic skeleton of the underlying mechanisms, to the recognition that stochasticity can itself be a mechanism for driving many interesting phenomena (Coulson *et al.* 2004). Yet this transition from *noise the nuisance* to *noise the creator* of ecological phenomena has had, with a few notable exceptions, relatively little impact in broader thinking about stochasticity. One of the most provocative of those exceptions has turned the classical notion of noise the nuisance on its head: recognising that noise driven phenomena can become a tool to reveal underlying processes: to become *noise the informer*. Here I argue that this third shift in perspective offers an opportunity to better bridge the divide between respective primarily theoretical and primarily empirical communities by seeing noise not as mathematical curiosity or statistical bugbear, but as a source for new opportunities for inference.

In arguing for this shift, it essential to recognise this is a call for a bigger tent, not for the rejection of previous paradigms. What I will characterize as 'noise the nuisance' reflects a predominately statistical approach, in which noise, almost by definition, represents all the processes we are not interested in that create additional variation which might obscure the pattern of interest. By contrast, an extensive literature has long explored how noise itself can create patterns and explain processes from population cycling to coexistence. These broad categories should be seen as a spectrum and not be mistaken for either a sharp dichotomy nor a reference to a strictly empirical-theoretical divide. Each paradigm expands upon rather than rejects the previous notion of noise: the recognition that noise can create novel phenomena does not mean that noise cannot also obscure the signal of some process of interest. Likewise, seeking to use noise as a novel source of information about underlying processes will be informed by both previous paradigms, as our discussion will illustrate.

Accompanying this discussion, I provide concise and commented code for simulating each of the models we will discuss as Appendix S1, and more mathematical background and derivations in Appendix S2. Numerical simulations permit poking and prodding investigation unencumbered by either experimental design or mathematical formalism. A copy of these appendices is maintained at https://github.com/cboettig/noise-phenomena.

To emphasise the underlying trend in the changing roles in which we see and understand noisy processes, I will also restrict my focus to relatively simple models primarily from population ecology context. Simplicity not only makes examples (in equations and in code) more tractable but also allows us to focus on aspects that are germane to many contexts rather than unique to particular complexities (Bartlett 1960;

(Semi-) Live demo

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

Skipped setup

install LaTeX

Only required if you don't have LaTeX installed

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

On Windows

You might need to install Rtools

installed workflow dependencies

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

installed `rrtools`

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

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

Got materials

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

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

Opened Rstudio

in project `rrtools-rse19`

loaded some libraries

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

Relevant global user settings

.Rprofile

```
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"
)
```

Edit file

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

.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! 

Demo Links

demo materials

bit.ly/rrtools_wks

workflow code

bit.ly/rrtools-workflow

outputs

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

Thanks for 👀 👋