

Reproducible scientific workflows in R

Introduction to the {targets} package

Ernest Guevarra

2022-02-14

Outline

- Concepts on scientific workflows
- The `{targets}` package
- Practical session

Concepts on scientific workflows

**Concept #1: Reproducibility, reproducibility,
reproducibility!**



Karl Broman

@kwbroman



Keith Baggerly: Most important tool for [#ReproducibleResearch](#) is the *mindset*, when starting, that the end product will be reproducible.

♡ 11 4:07 PM - Nov 20, 2015



 [See Karl Broman's other Tweets](#)

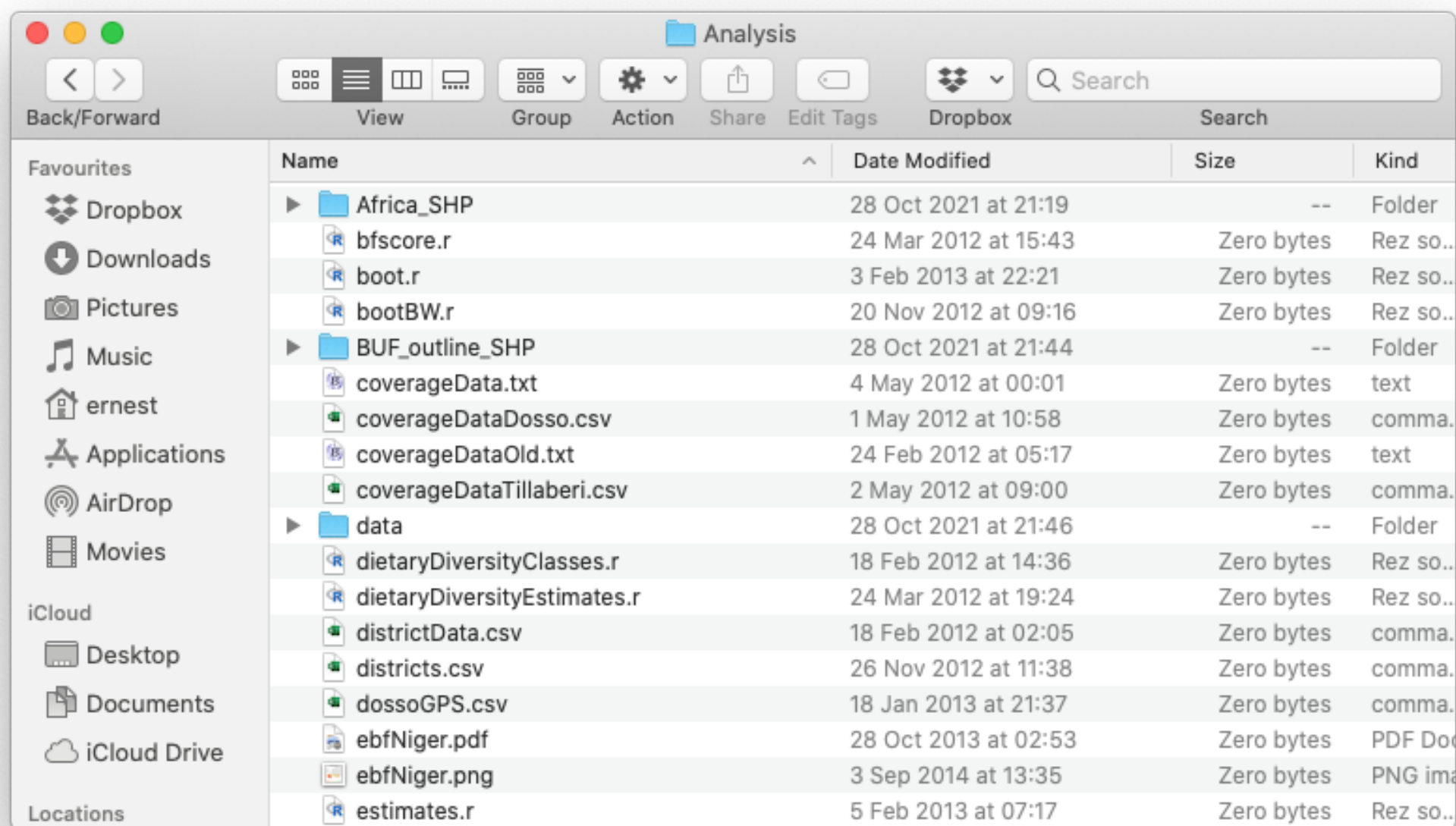


Keith Baggerly, via @kwbroman tweet

Concept #2: Organisation

File organization and naming are powerful weapons against chaos.

@JennyBryan



Concept #3: DRY - Don't repeat yourself

**Don't repeat yourself. It's not only repetitive,
it's redundant, and people have heard it before.**

Lemony Snicket

```
# Overlay maps of Niger and Nigeria to clean-up the map
par(new=TRUE)
plot(nigeria, axes = FALSE, xlim = mapXLimits, ylim = mapYLimits, border = "white", col = "white")

par(new = TRUE)
plot(boundaries, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.5, border = "black")

par(new = TRUE)
plot(n1, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.25, col = "blue")

par(new = TRUE)
plot(n4, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.25, col = "blue")

par(new = TRUE)
plot(n5, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.25, col = "blue")

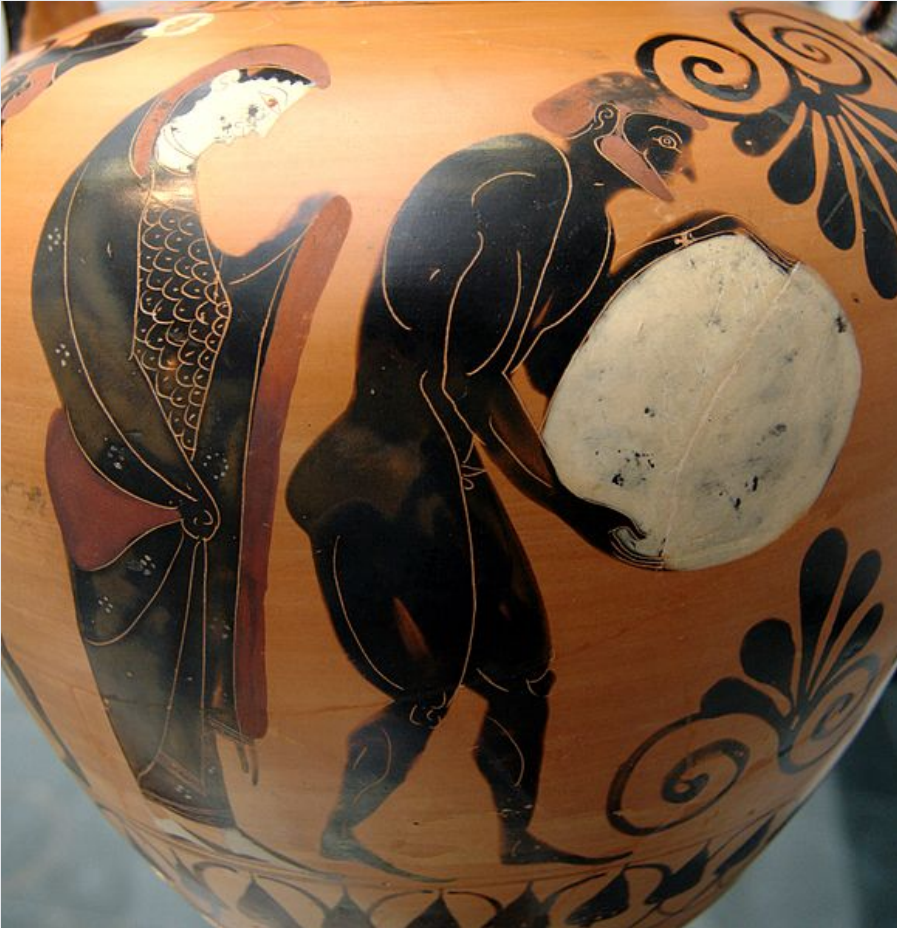
par(new = TRUE)
plot(n6, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.25, col = "blue")

par(new = TRUE)
plot(n6.27, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.25, col = "blue")

par(new = TRUE)
plot(n7, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.25, col = "blue")

par(new = TRUE)
plot(n14, axes = FALSE, xlim = mapXLimits, ylim=mapYLimits, lwd = 0.25, col = "blue")
```

Sisyphean loop



1. Launch the code.
2. Wait while it runs.
3. Discover an issue.
4. Restart from scratch.

Concept #3: Frequency reduces difficulty



Jenny Bryan ✓

@JennyBryan



If it hurts, do it more often. — @martinfowler

We shared this idea in a recent workshop and it resonated with many. Applies to all sorts of things: git tasks (pushing & pulling), keeping your s/w stack current, sharing your work with others, etc.

[martinfowler.com/bliki/Frequenc...](https://martinfowler.com/bliki/Frequency.html)

♡ 155 7:32 PM - Jan 22, 2019



💬 50 people are talking about this



@martinfowler, via @JennyBryan tweet

The {targets} package

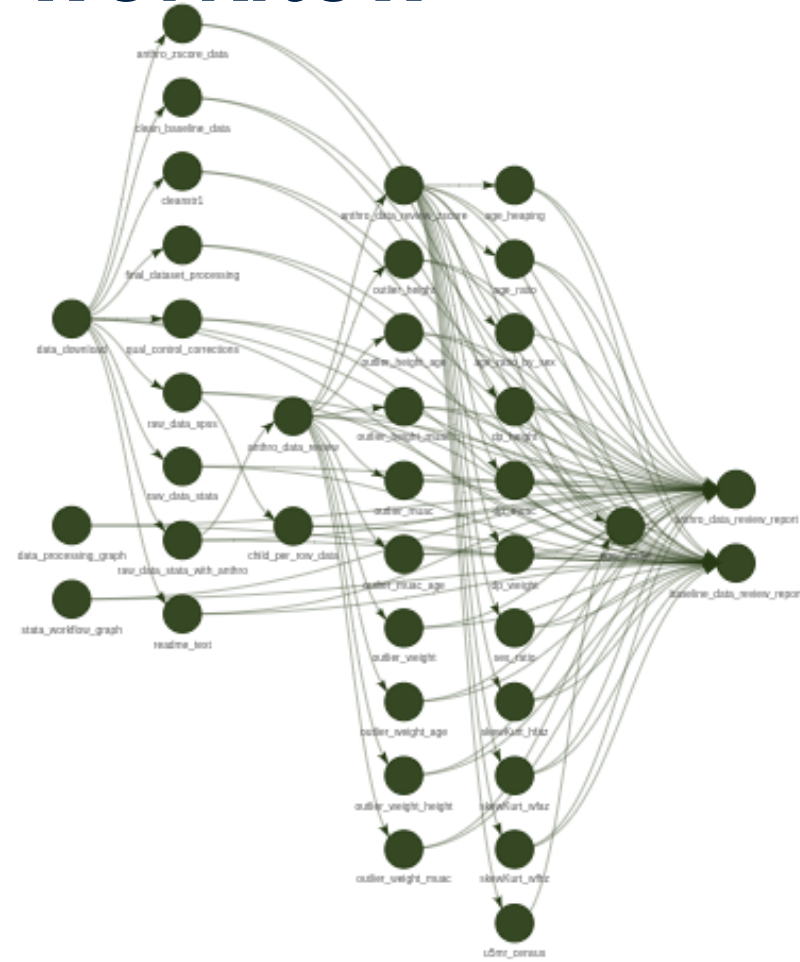


- a pipeline toolkit for Statistics and data science in R
- maintain a reproducible workflow without repeating yourself
- learns how your workflow fits together
- skips costly runtime for tasks that are already up-to-date
- runs only the necessary computation
- supports implicit parallel computing
- abstracts files as R objects
- shows tangible evidence that the results match the underlying code and data

{targets} file organisation

{targets} script file

{targets} workflow



Questions?

Practical session

We will all continue to go through Exercise #1 in the
Practical R for Epidemiologists book

Questions?

Thank you!

Slides can be viewed at <https://OxfordIHTM.github.io/open-reproducible-science/session4.html>

PDF version of slides can be downloaded at <https://OxfordIHTM.github.io/open-reproducible-science/pdf/session4-reproducible-scientific-workflows.pdf>

R scripts for slides available [here](#)