Reproducible research in R with

remake

Daniel Falster, adaptive_plant, danielfalster.com

Sydney Bioinformatics Research Symposium, 22 Jul 2016
Who am I?
Two related problems

Making research reproducible
  ► Reproducibility crisis

Efficient R-based work-flows
  ► When to re-run?
R can be irreproducible
R can be irreproducible

```r
setwd("C:/Users/jdoe/myproject/")
```
R can be irreproducible

Undocumented dependencies
R can be irreproducible

Manually edit your data
R can be irreproducible

Graphs that need manual tweaking
R can be reproducible
R can be reproducible

Don’t do those things
Remake helps

Rich FitzJohn – @rgfitzjohn
Package in active development
github.com/richfitz/remake

Thanks Rich for great package!
Demands good habits

1. Script everything
2. Self contained project
3. Document dependencies
4. Use functions
5. No global variables
Demands good habits

(You do this anyway right?)

1. Script everything
2. Self contained project
3. Document dependencies
4. Use functions
5. No global variables
Be your best self
Be your best self
+ better work flows
```r
url <- "https://github.com/dfalster/..."
download(url, "downloads/baad.rds", mode="wb")
baad <- readRDS("downloads/baad.rds")
pdf("figures/fig_count.pdf")
figure_count(baad)
de.v.off()
```

Turn R script into data

```
url <- "https://github.com/dfalster/..."
download(url, "downloads/baad.rds", mode="wb")
baad <- readRDS("downloads/baad.rds")
pdf("figures/fig_count.pdf")
figure_count(baad)
de.v.off()
```
Turn Rscript into data

- Targets: Objects to make
- Rules: Functions building targets
- Dependencies: Upstream targets
remake.yml

downloads/baad.rds:
  download: https://github.com/...

baad:
  command: readRDS("downloads/baad.rds")

figures/fig_count.pdf:
  command: fig_count(baad)
  plot: TRUE
Map dependencies

- downloads/baad.rds
- baad
- downloads/BTFP15.jpg
- figures/baad_variable_count.pdf
- figures/tradeoff-Falster2015-1.pdf
- figures
- slides.tex
- slides.pdf
Map dependencies
Map dependencies → Many advantages
One-command build

> remake::make("all")
< MAKE > all
[ BUILD ] baad
[ KNIT ] report.md
[ BUILD ] report.html
[ PLOT ] output/map.pdf
[ PLOT ] output/fig_count.pdf
[ PLOT ] output/fig_allometry.pdf
[ BUILD ] summary
[ BUILD ] output/summary.csv
[ ----- ] all

| baad <- baad_data("...
| knitr::knit("report...
| render("report.md",
| figure_map(baad)
| figure_variable_cou
| fig_allometry(baad)
| summary <- table_su
| write.csv(summary,

Determines when to rerun

```r
> remake::make("all")
< MAKE > all
[ BUILD ] baad
[ KNIT ] report.md
[ BUILD ] report.html
[ PLOT ] output/map.pdf
[ PLOT ] output/fig_count.pdf
[ PLOT ] output/fig_allometry.pdf
[ BUILD ] summary
[ BUILD ] output/summary.csv
[ ----- ] all
```
Determines when to rerun

> remake::make("all")
< MAKE > all
[ OK ] baad
[ OK ] report.md
[ OK ] report.html
[ OK ] output/map.pdf
[ OK ] output/fig_count.pdf
[ OK ] output/fig_allometry.pdf
[ OK ] summary
[ OK ] output/summary.csv
[ ----- ] all
Determines when to rerun

> remake::make("all")
< MAKE > all
[ OK ] baad
[ OK ] report.md
[ OK ] report.html
[ OK ] output/map.pdf
[ PLOT ] output/fig_count.pdf | figure_count(baad)
[ OK ] output/fig_allometry.pdf
[ OK ] summary
[ OK ] output/summary.csv
[ ----- ] all
Determines when to rerun

> remake::make("all")
< MAKE > all
[ OK ] baad
[ OK ] report.md
[ OK ] report.html
[ OK ] output/map.pdf
[ OK ] output/fig_count.pdf
[ OK ] output/fig_allometry.pdf
[ OK ] summary
[ OK ] output/summary.csv
[ ----- ] all
Collaboration

R + remake + GitHub = ❤️
Detect collaborators changes

> remake::make("all")
< MAKE > all
[  OK  ] baad
[  KNIT  ] report.md | knitr::knit("report...
[  BUILD  ] report.html | render("report.md",...
[  OK  ] output/map.pdf | figure_count(baad)
[  PLOT  ] output/fig_count.pdf
[  OK  ] output/fig_allometry.pdf
[  OK  ] summary
[  OK  ] output/summary.csv
[ -----  ] all
Runs in clean environment
Runs in clean environment
Start over

> remake::make("clean")
( DEL ) baad | rm("baad")
( DEL ) summary | rm("summary")
[ CLEAN ] clean
( DEL ) report.md | file.remove("report.md")
( DEL ) report.html | file.remove("report.html")
( DEL ) output/map.pdf | file.remove("output/map.pdf")
( DEL ) output/fig_count.pdf | file.remove("output/fig_count.pdf")
( DEL ) output/fig_allometry.pdf | file.remove("output/fig_allometry.pdf")
( DEL ) output/summary.csv | file.remove("output/summary.csv")
Getting started

1. Embrace good habits
2. Work from examples
3. Easier in new project?
Resources

**Slides**
danielfalster.com

**Package**
github.com/richfitz/remake

github.com/ropenscilabs/remake-tutorial

github.com/nicercode/2015.12.08-EcoStats

**Tutorials**
github.com/traitecoevo/remake_baad_example

github.com/dfalster/baad

github.com/RemkoDuursma/baadanalysis

github.com/jscamac/Alpine_Shrub_Experiment

github.com/AnaisGibert/Growth_trait_metaanalysis
Acknowledgements

Funding
Macquarie University
Australian Research Council

Inspiration
Rich FitzJohn for remake & his talk github.com/richfitz/reproducibility-2014

Design
Mike Bostock: bost.ocks.org/mike/d3/workshop