

Reproduce research with remake

Daniel Falster, Rich FitzJohn, Will Cornwell

 adaptive_plant,  danielfalster.com

Two related problems

1. Making research reproducible

- ▶ Recreate outputs from a paper
- ▶ Record entire work-flow

2. Efficient R-based work-flows

- ▶ When to re-run?

R can be
irreproducible

R can be irreproducible

```
setwd("myproject/final2/works")
```

R can be irreproducible

Graphs that need manual tweaking

R can be irreproducible

Manually edit your data

R can be irreproducible

Undocumented dependencies

**R can be
reproducible**

Don't do those things

R can be reproducible

Use remake to help with rest

Remake helps



Rich FitzJohn  [rgfitzjohn](#)

Package in active development

Starts with nice code

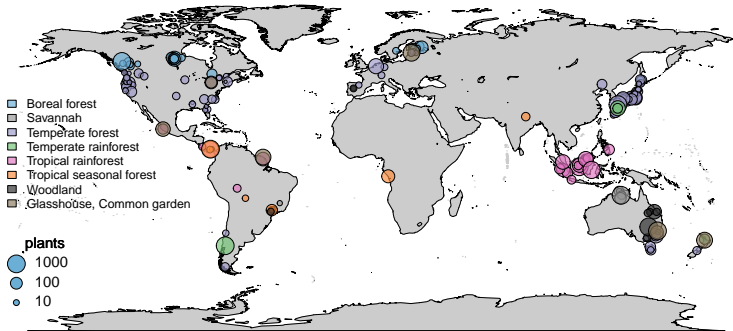
(You do this anyway right?)

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

Encourages nicer code

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

Biomass And Allometry Database (BAAD)



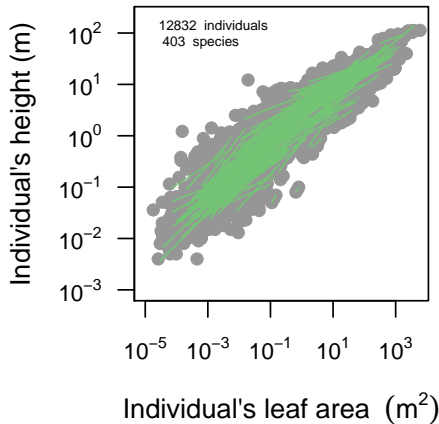
176 studies, 21084 plants, 678 species

Harvested plants

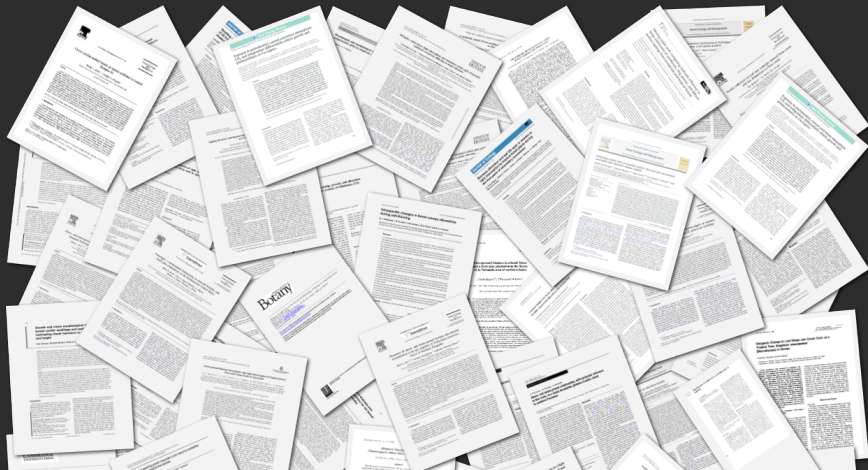


Komiyama et al., Japan

Why: We needed data



Lots of small data



#otherpeoplesdata



Andy Skelton @Da_Pineapple · Oct 20

I could deal with working on data others have collected if they'd only be consistent about how it was collected... [#otherpeoplesdata](#) [#sigh](#)

1:54 AM - 20 Oct 2015 · Details



Script everything

 Transparent, extensible

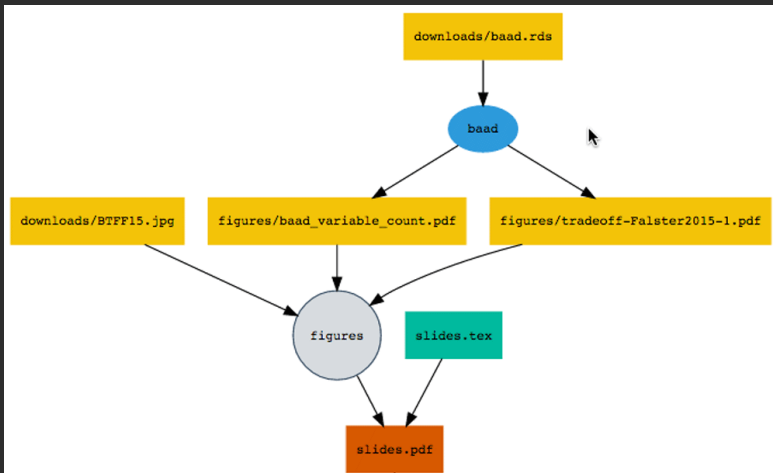
Script everything

 Transparent, extensible

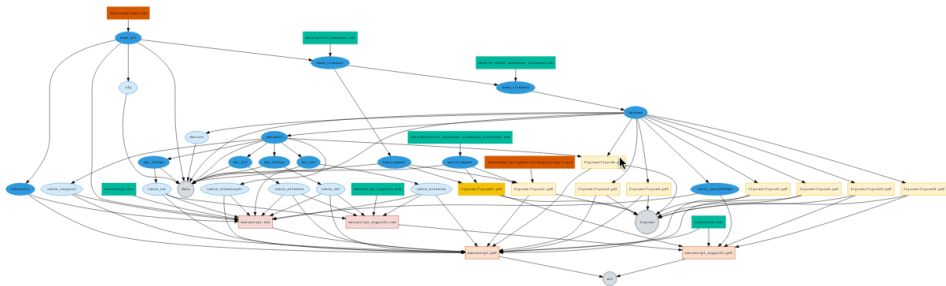
```
# Rebuild BAAD
```

```
git clone git@github.com:dfalster/baad.git  
cd baad  
remake
```

remake



remake



remake

```
# Rebuild figures from doi.org/10.1101/025361
git clone git@github.com:RemkoDuursma/baadanalysis.git
cd baadanalysis
remake
```

```
[ LOAD ]
< MAKE > all
[ BUILD ] downloads/baad.rds      | download_data("....
[ READ ]                          | # loading packages
[ BUILD ] baad_all                | baad_all <- readRDS(...)
[ BUILD ] baad_climate1          | baad_climate1 <- addWor...
[ BUILD ] dataset                | dataset <- prepare_data...
[ PLOT ] figures/Figure1.pdf     | figure1(baad_mapmat,...
[ PLOT ] figures/Figure2.pdf     | figure2(dataset)
...
[ PLOT ] figures/FigureS3.pdf    | figureS3(dataset)
[ ----- ] figures
```

remake.yml

Targets

Objects of interest

Rules

Functions returning targets

Dependencies

Upstream targets

remake.yml

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

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

figures/baad_variable_count.pdf:
  command: baad_variable_count(baad)
  plot: TRUE
```


Collaboration

R + remake + GitHub = ❤️

Today's example

(github.com/nicercode/2015.12.08-EcoStats)

1. Download some data
2. Process that file to prepare it for analysis
3. Create some plots
4. Create a table
- (5.) Create a report