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

```
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

Turn Rscript into data

```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)
dev.off()
```

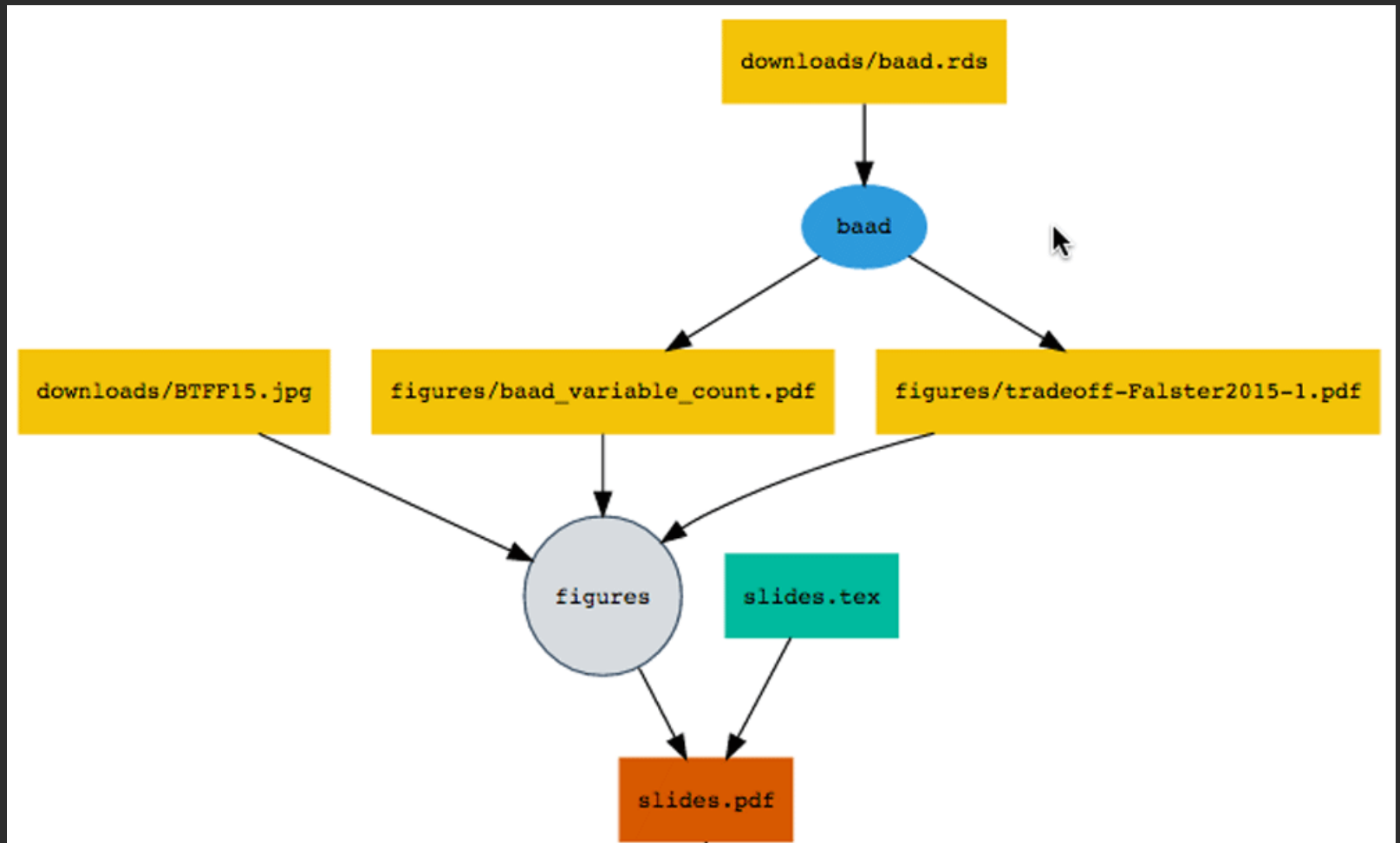
Turn Rscript into data

Targets	Objects to make
Rules	Functions building targets
Dependencies	Upstream targets

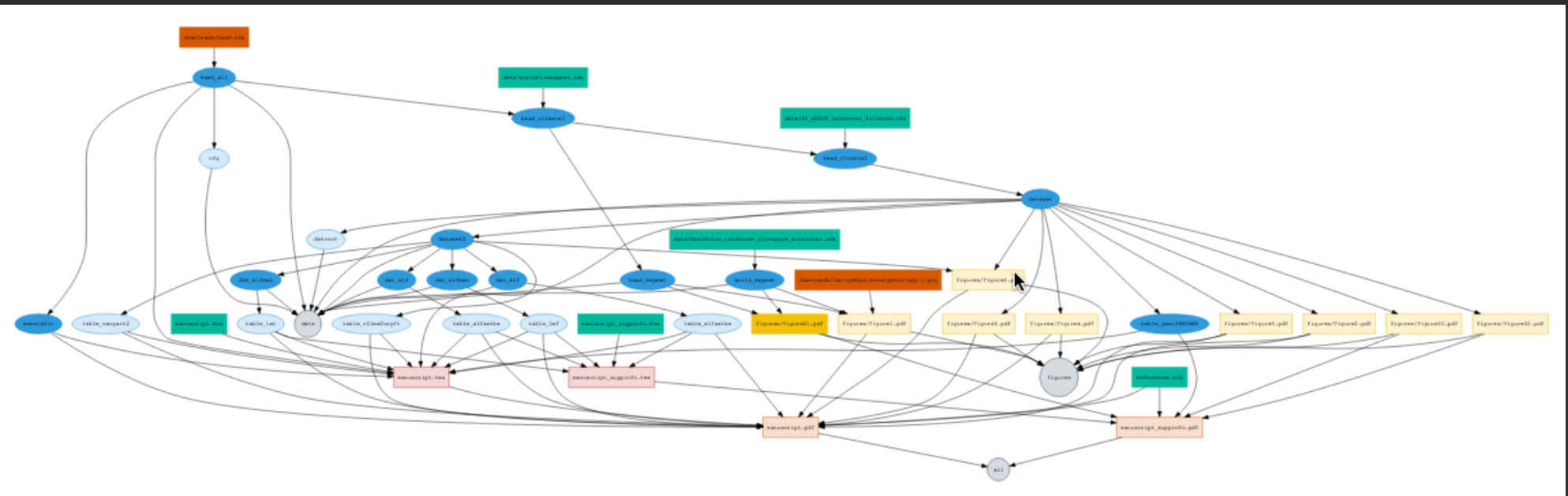
remake.yml

```
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



Map dependencies



Map dependencies

→ Many advantages

One-command build

```
> remake::make("all")
<  MAKE > all
[ BUILD ] baad                          |  baad <- baad_data("
[  KNIT ] report.md                     |  knitr::knit("report
[ BUILD ] report.html                   |  render("report.md",
[  PLOT ] output/map.pdf                |  figure_map(baad)
[  PLOT ] output/fig_count.pdf          |  figure_variable_cou
[  PLOT ] output/fig_allometry.pdf      |  fig_allometry(baad)
[ BUILD ] summary                       |  summary <- table_su
[ BUILD ] output/summary.csv            |  write.csv(summary,
[ ----- ] all
```

Determines when to rerun

```
> remake::make("all")
<  MAKE > all
[ BUILD ] baad                          |  baad <- baad_data("
[  KNIT ] report.md                     |  knitr::knit("report
[ BUILD ] report.html                   |  render("report.md",
[  PLOT ] output/map.pdf                |  figure_map(baad)
[  PLOT ] output/fig_count.pdf          |  figure_variable_cou
[  PLOT ] output/fig_allometry.pdf      |  fig_allometry(baad)
[ BUILD ] summary                       |  summary <- table_su
[ BUILD ] output/summary.csv            |  write.csv(summary,
[ ----- ] 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
```

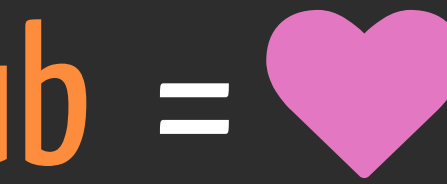
Determines when to rerun

```
> 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
[  PLOT ] output/fig_count.pdf         |  figure_count(baad)
[    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("re
(   DEL ) report.html                   |  file.remove("re
(   DEL ) output/map.pdf                |  file.remove("ou
(   DEL ) output/fig_count.pdf          |  file.remove("ou
(   DEL ) output/fig_allometry.pdf      |  file.remove("ou
(   DEL ) output/summary.csv            |  file.remove("ou
```

Getting started

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

Resources

Slides	danielfalster.com
Package	github.com/richfitz/remake
Tutorials	github.com/ropenscilabs/remake-tutorial
	github.com/nicercode/2015.12.08-EcoStats
Examples	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