Cabinet Office

Matt Dray
Data Scientist
Government Digital Service
@mattdray_gds

Can {drake} RAP?

Drake
drake()
imgflip.com

drake

github.com/ropensci/drake

tl;dr

Scale the work
you need.

Skip the work
you don't.

See evidence
of reproducibility.

Materials

This talk:

- a blog post
- code for the demo

For {drake}:

- visit the website
- read the full manual
- learn from a course
- use it in an app

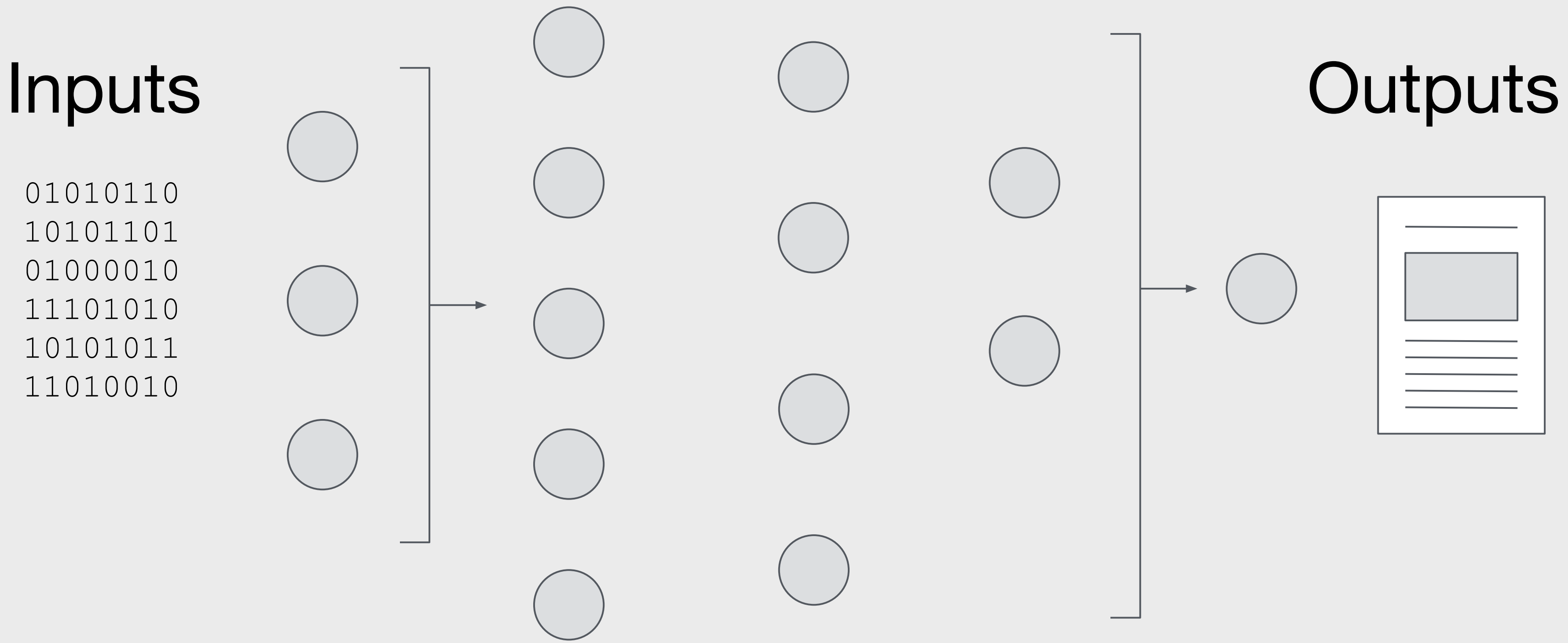
GDS

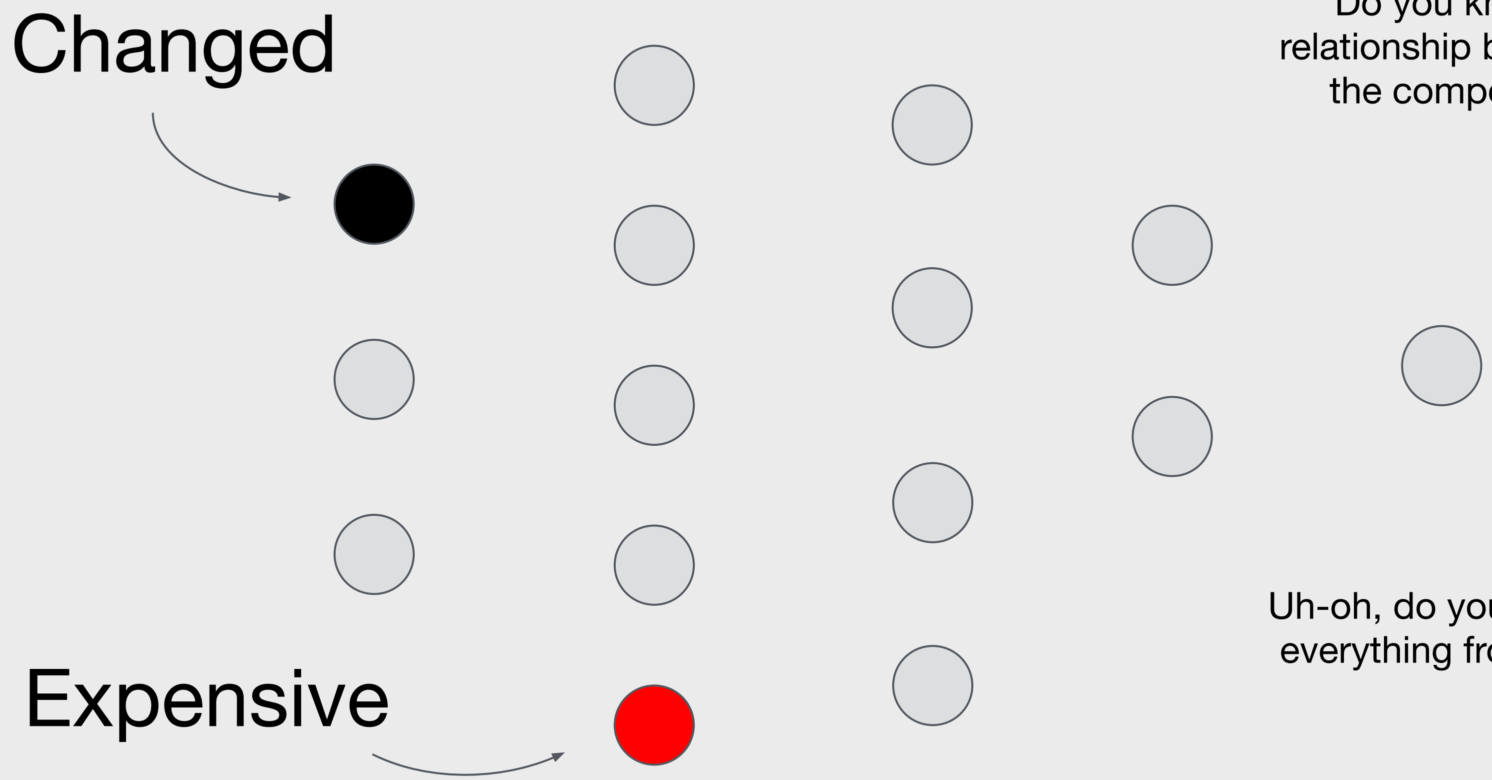
Workflows

Inputs ⟶ ‘Stuff happens’ ⟶ Outputs

01010110
10101101
01000010
11101010
10101011
11010010

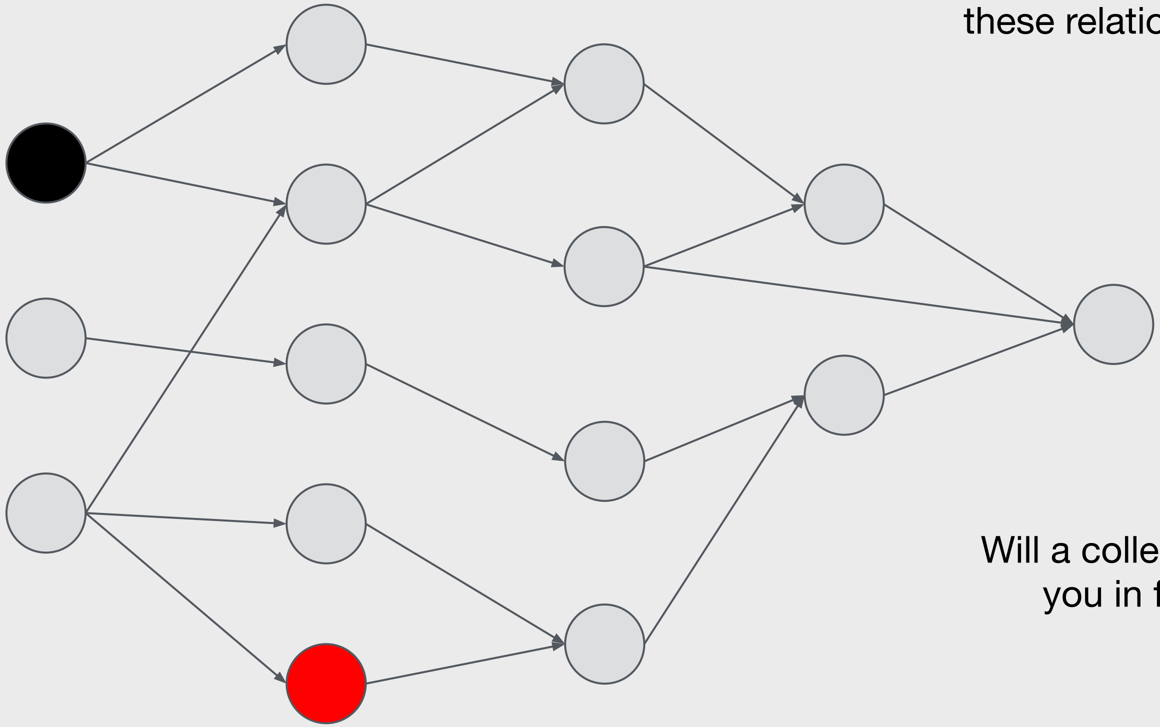
‘Stuff happens’





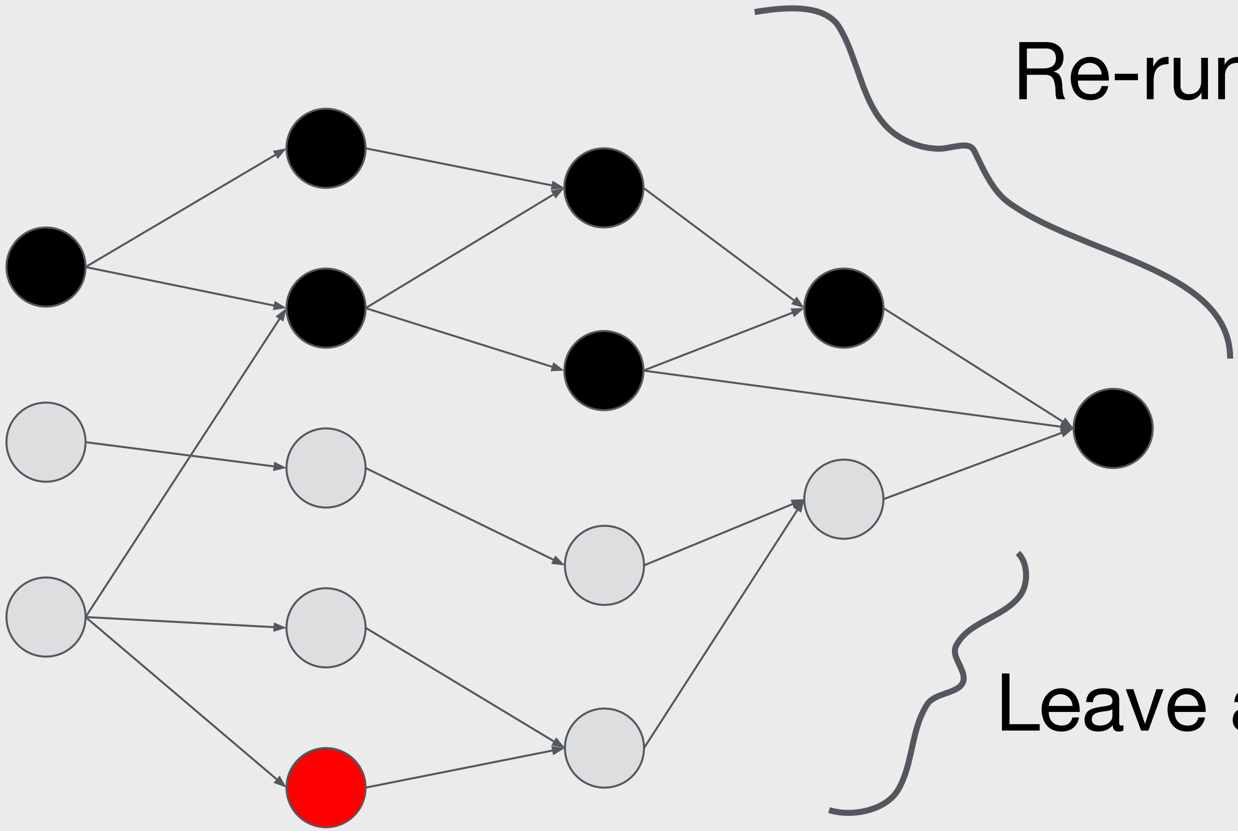
Do you know the relationship between all the components?

Uh-oh, do you have to run everything from scratch?



Can you remember these relationships?

Will a colleague? Will you in future?


Re-run
Leave alone

Solution: {drake}

- Time/computation saving
- Less for you to remember
- Better reproducibility
- Visualise dependencies
- Deals with parallelisation
- Entirely R-based

{drake} workflow overview

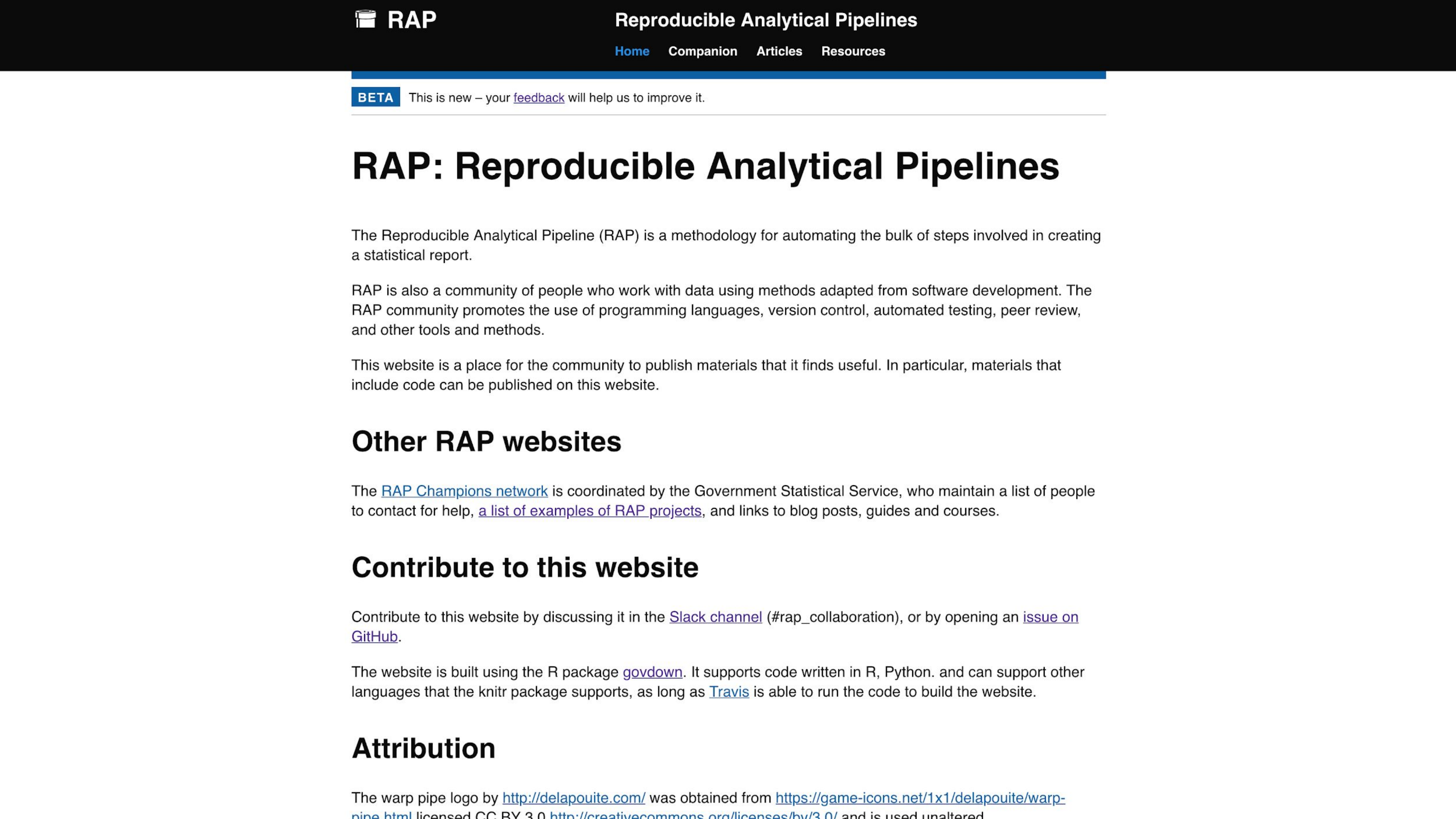
1. Create scripts and plan
2. Make the plan
3. Change stuff and re-make

When to {drake}?

```
01_read.R
02_import.R
03_tidy.R
04_clean.R
05_model.R
06_plot.R
07_report.Rmd
```

This is a good start

But why isn’t it optimal?



RAP Reproducible Analytical Pipelines

Home Companion Articles Resources

BETA This is new – your feedback will help us to improve it.

RAP: Reproducible Analytical Pipelines

The Reproducible Analytical Pipeline (RAP) is a methodology for automating the bulk of steps involved in creating a statistical report.

RAP is also a community of people who work with data using methods adapted from software development. The RAP community promotes the use of programming languages, version control, automated testing, peer review, and other tools and methods.

This website is a place for the community to publish materials that it finds useful. In particular, materials that include code can be published on this website.

Other RAP websites

The RAP Champions network is coordinated by the Government Statistical Service, who maintain a list of people to contact for help, a list of examples of RAP projects, and links to blog posts, guides and courses.

Contribute to this website

Contribute to this website by discussing it in the Slack channel (#rap_collaboration), or by opening an issue on GitHub.

The website is built using the R package govdown. It supports code written in R, Python. and can support other languages that the knitr package supports, as long as Travis is able to run the code to build the website.

Attribution

The warp pipe logo by http://delapouite.com/ was obtained from https://game-icons.net/1x1/delapouite/warp-pipe.html licensed CC BY 3.0 http://creativecommons.org/licenses/by/3.0/ and is used unaltered.

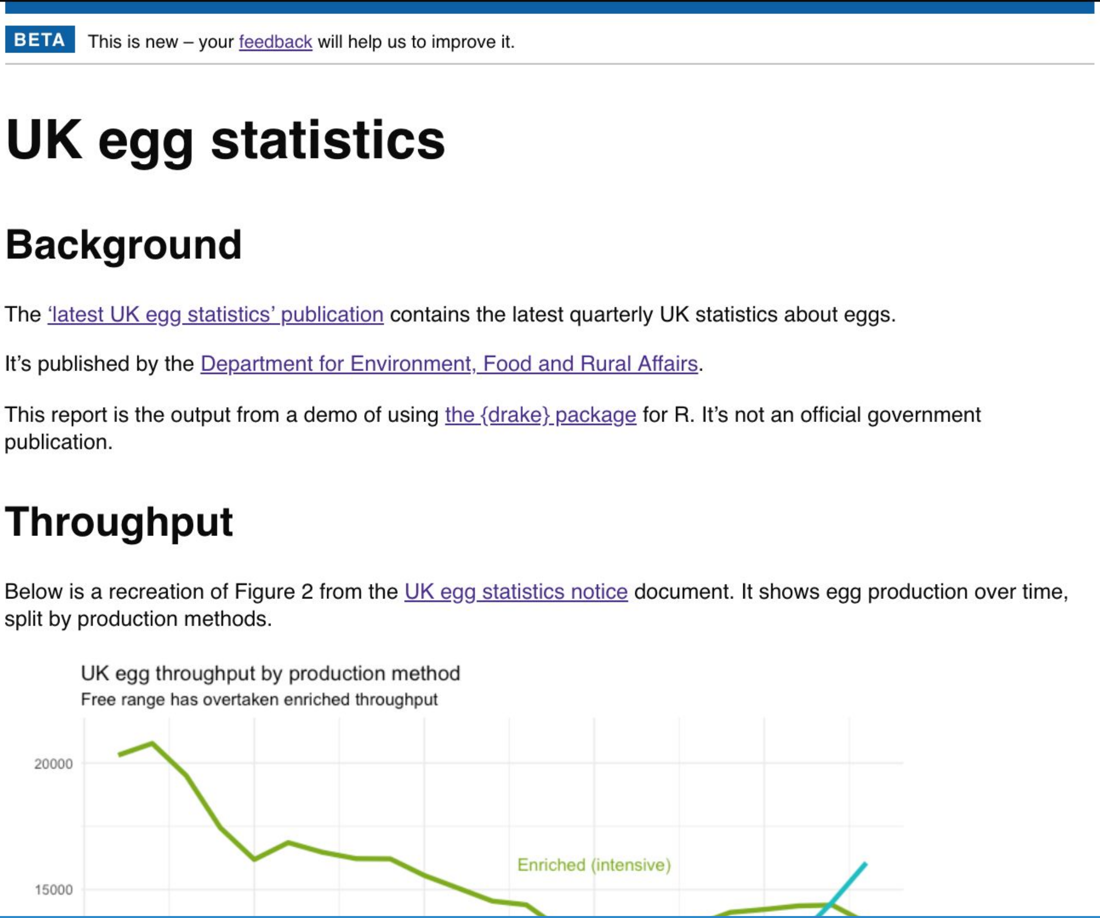
RAP is for:

- reproducibility
- automation
- minimising error
- doing it faster
- building trust

Egg stats:

- publication
- my code
- my report

Demo: {drake} for RAP



BETA This is new – your feedback will help us to improve it.

UK egg statistics

Background

The 'latest UK egg statistics' publication contains the latest quarterly UK statistics about eggs.

It's published by the Department for Environment, Food and Rural Affairs.

This report is the output from a demo of using the {drake} package for R. It's not an official government publication.

Throughput

Below is a recreation of Figure 2 from the UK egg statistics notice document. It shows egg production over time, split by production methods.

UK egg throughput by production method
Free range has overtaken enriched throughput
20000
15000
Enriched (intensive)

GDS

Live demo

Access the demo code in RStudio in your browser:

launch binder

Step-by-step

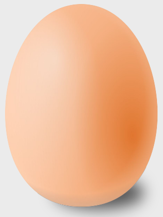
{drake} workflow overview

1. Create scripts and plan
2. Make the plan
3. Change stuff and re-make

{drake} workflow overview

1. Create scripts and plan
2. Make the plan
3. Visualise
4. Change stuff
5. Check changes
6. Re-make

1. Create scripts and plan



GDS

‘Ingredients’

Files that set up your analysis

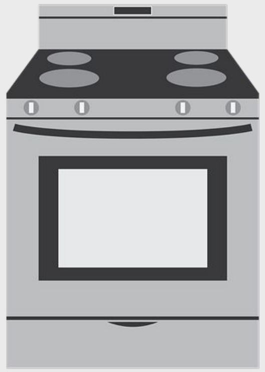
```
# packages.R

library(drake)
library(dplyr)
library(readODS)
library(ggplot2)
…

# functions.R

clean_data <- function(raw_data) { … }
create_plot <- function(data) { … }
```

```r
# plan.R

plan <- drake_plan(

  raw_data = read_ods("data.csv"),

  data = clean_data(raw_data),

  plot = create_plot(data),

  report = rmarkdown::render(
    knitr_in("report.Rmd"),
    output_file = file_out("report.html"),
  )

)
```

Write the ‘recipe’

Prepare steps into a dataframe

2. Make the plan



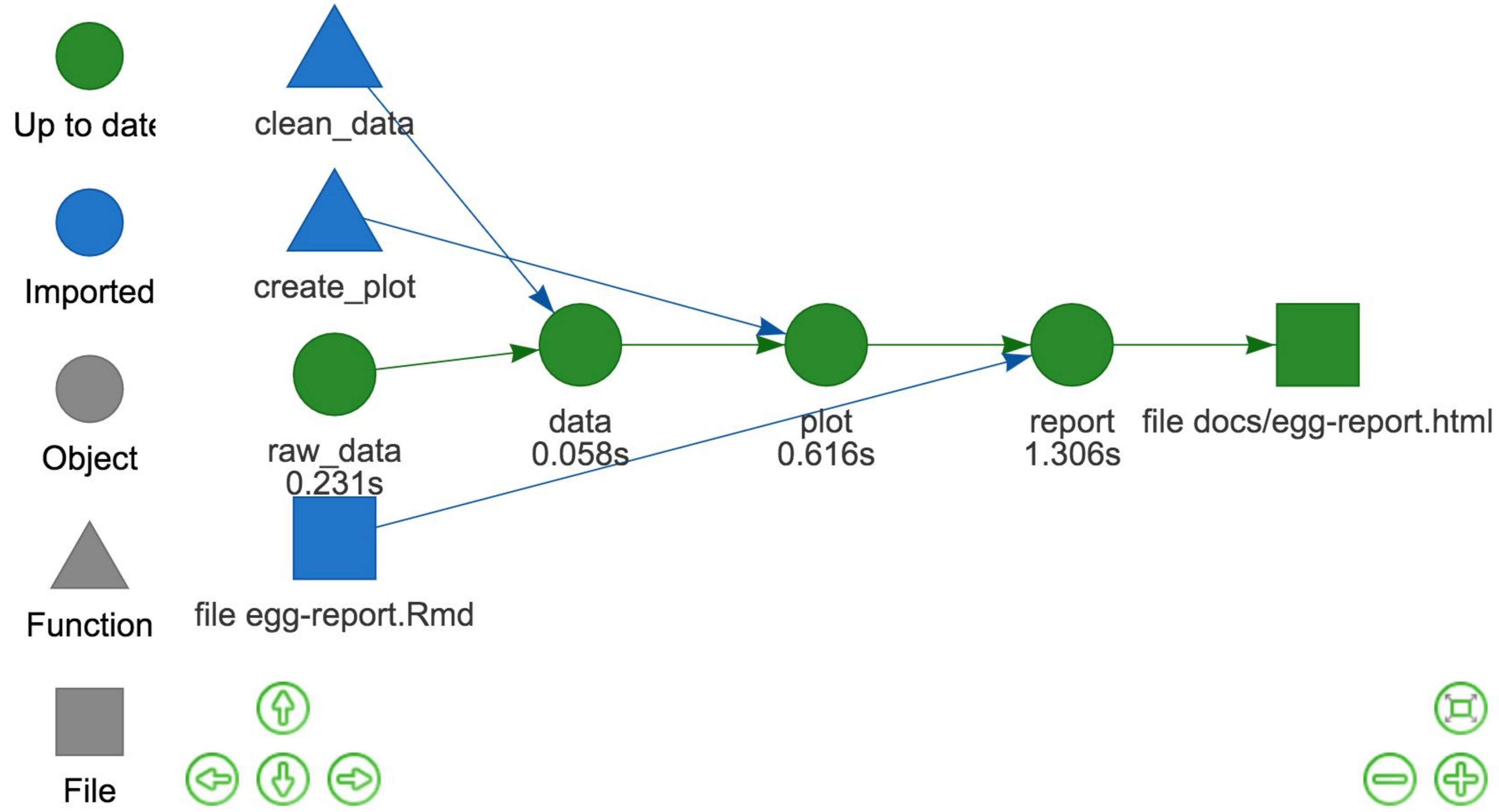
```
# make.R

source(packages.R)
source(functions.R)
source(plan.R)

make(plan)
```

‘Bake the cake’

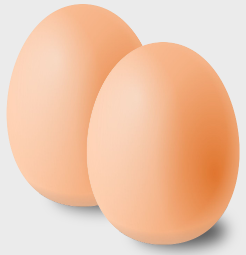
Prepare ingredients, fetch recipe

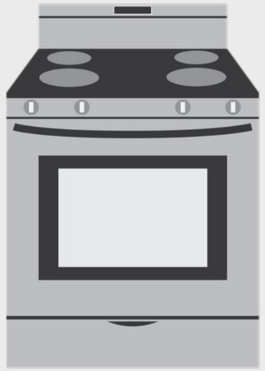
Make the recipe

3. Visualise


Up to date
Imported
Object
Function
File
clean_data
create_plot
raw_data
0.231s
file egg-report.Rmd
data
0.058s
plot
0.616s
report
1.306s
file docs/egg-report.html

4. Change stuff
5. Check changes
6. Remake





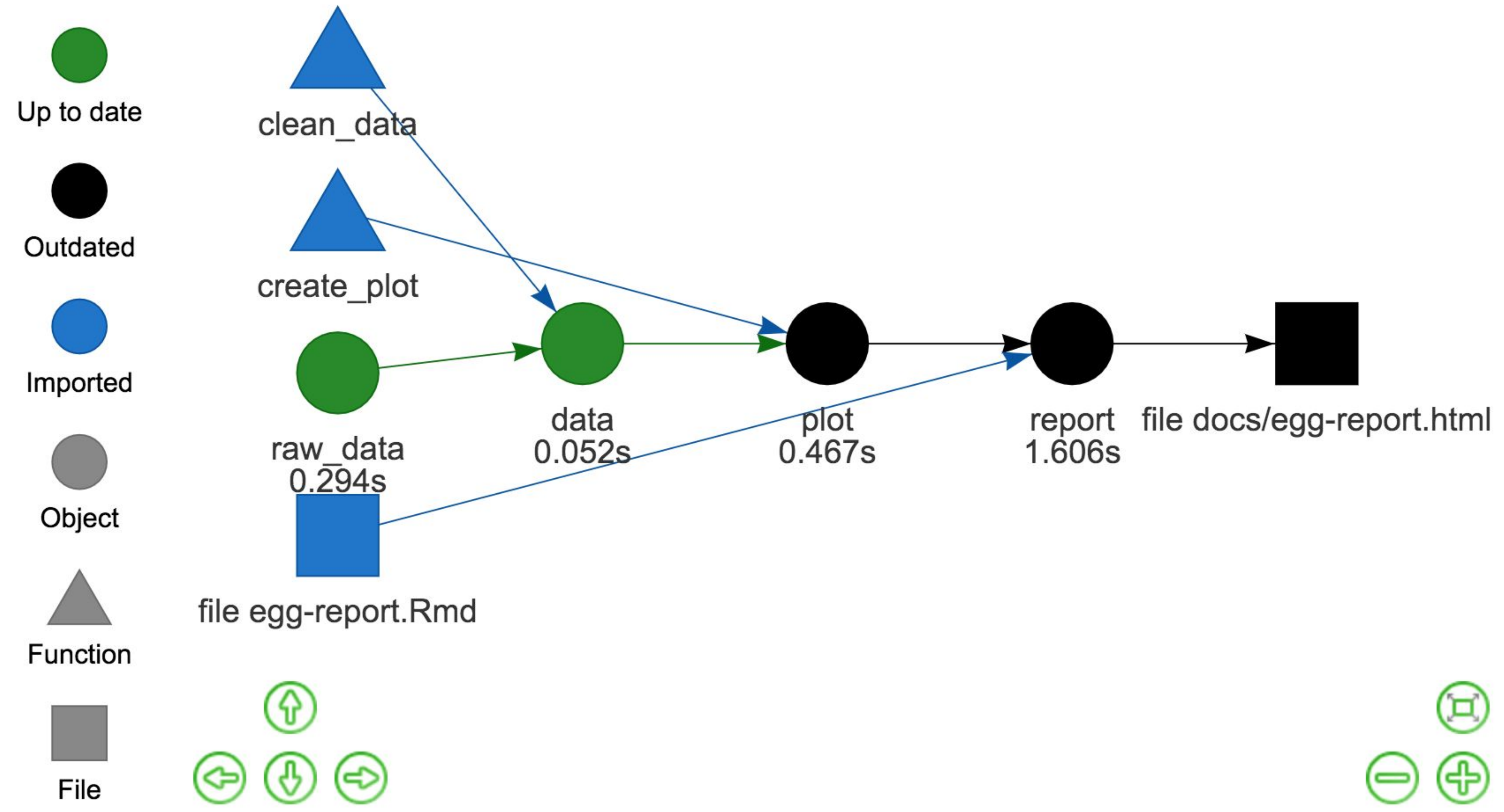
GDS

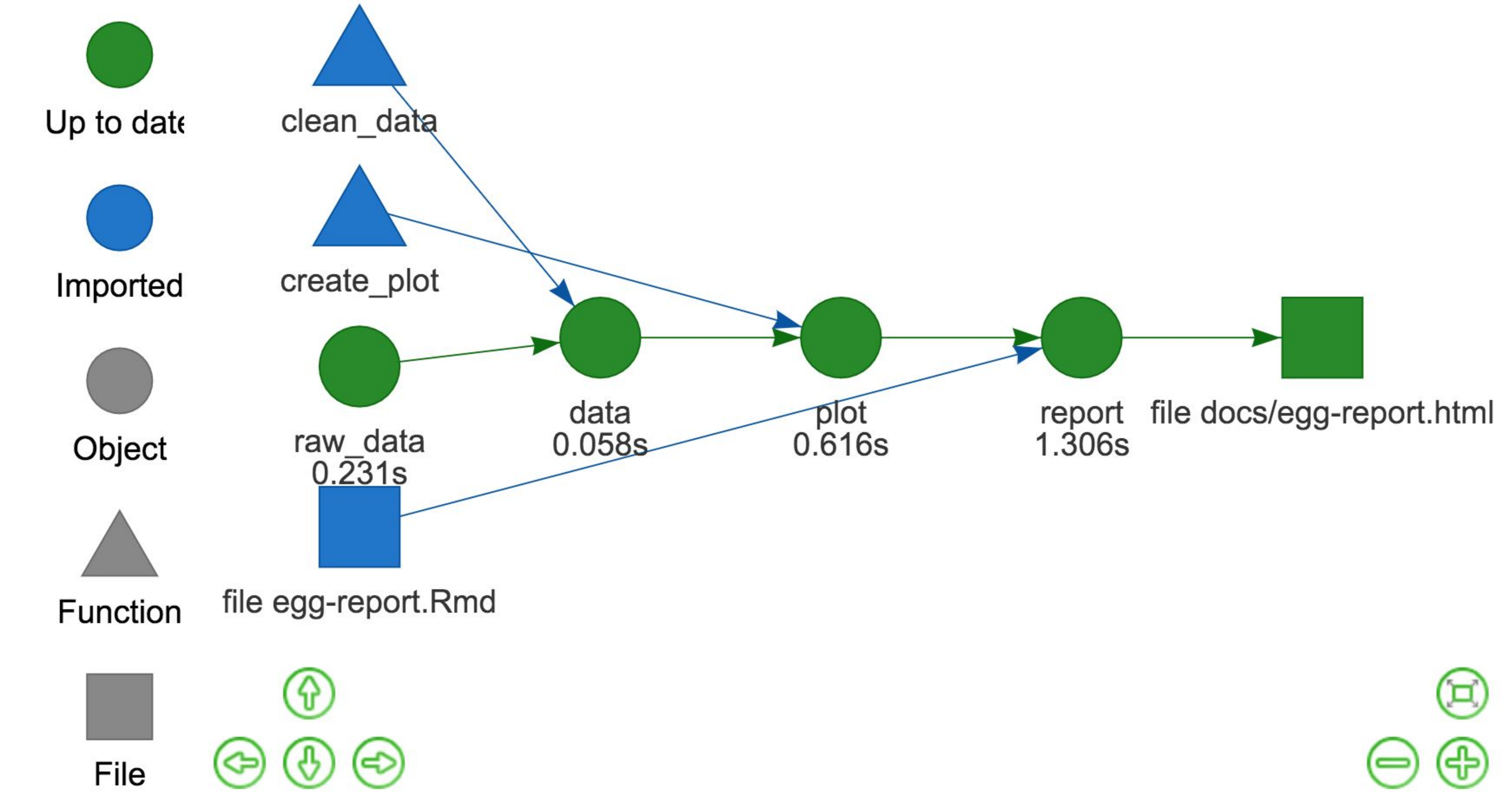
```
# change something, then:

source(functions.R)

outdated(config)

config <- drake_config(plan)
vis_drake_graph(egg_config)

make(plan.R)
```

Change your data/code

See what's out of date

Update!


Up to date
Outdated
Imported
Object
Function
File
clean_data
create_plot
raw_data
0.294s
file egg-report.Rmd
data
0.052s
plot
0.467s
report
1.606s
file docs/egg-report.html


Up to date
Imported
Object
Function
File
clean_data
create_plot
raw_data
0.231s
file egg-report.Rmd
data
0.058s
plot
0.616s
report
1.306s
file docs/egg-report.html

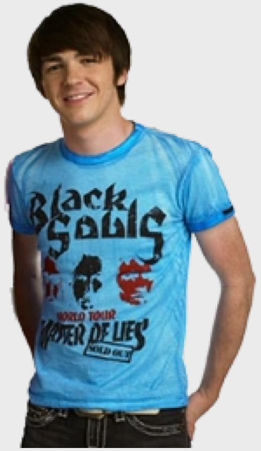
Folder view

```
drake-egg-rap/
├── R/
│   ├── functions.R
│   ├── packages.R
│   ├── egg-report.Rmd
│   └── plan.R
├── make.R
├── data/
│   └── eggs-packers-02may19a.ods
├── docs/
└── .drake/
```

```
drake-egg-rap/
├── R/
│   ├── functions.R
│   ├── packages.R
│   ├── egg-report.Rmd
│   └── plan.R
├── make.R
├── data/
│   └── eggs-packers-02may19a.ods
├── docs/
└── .drake/
```

```
drake-egg-rap/
├── R/
│   ├── functions.R
│   ├── packages.R
│   ├── egg-report.Rmd
│   └── plan.R
├── make.R
├── data/
│   └── eggs-packers-02may19a.ods
├── docs/
└── .drake/
```

Hall of fame


Black Souls
World Tour
Master of Lies
Sold Out

drake
github.com/ropensci/drake

Cabinet Office

Matt Dray
@mattdray_gds
matthew.dray@digital.cabinet-office.gov.uk