

Reproducible research is still a challenge

Rich FitzJohn  | Matt Pennell | Amy Zanne | Will Cornwell

June 9, 2014

why is this still hard?

complexity

why is this still hard?

tooling

why is this still hard?

incentives + motivation

nobody cares

(Apart from people doing it out of the goodness of their hearts of course)

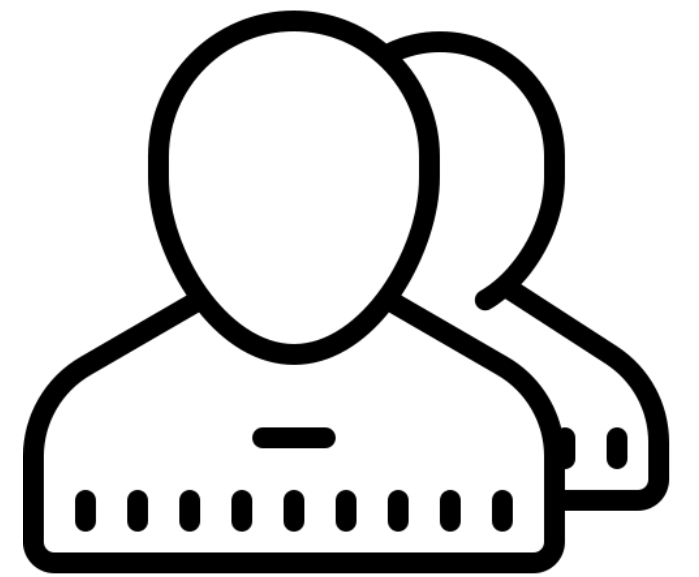
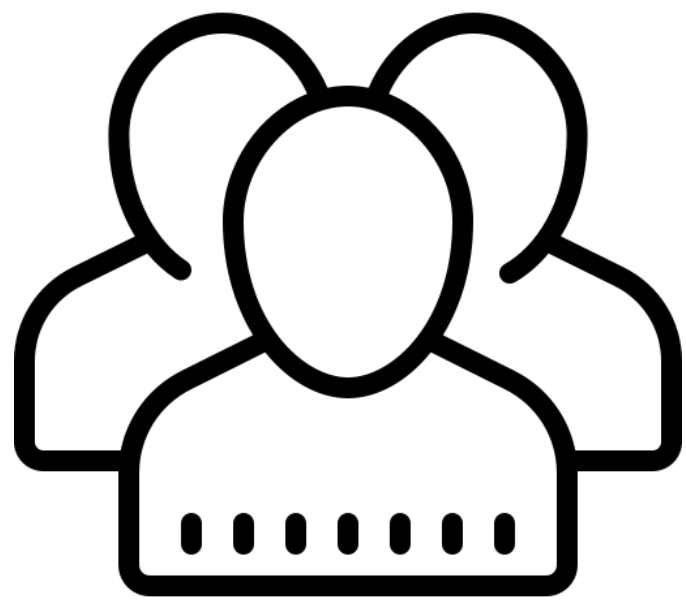
nobody cares?

(Apart from people doing it out of the goodness of their hearts of course)



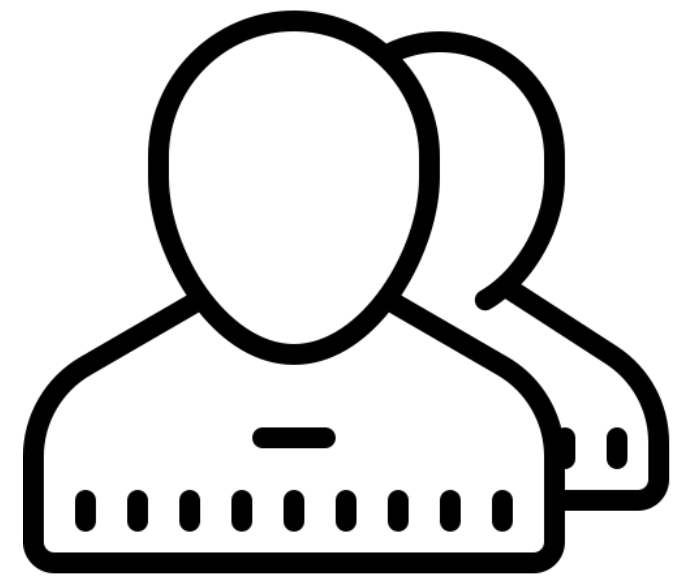
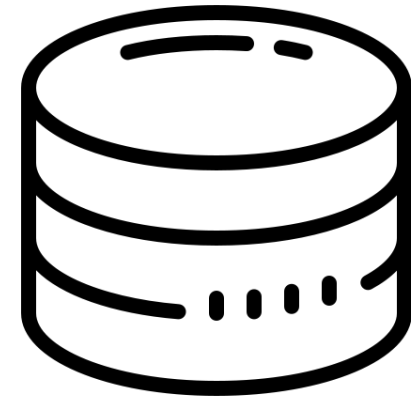
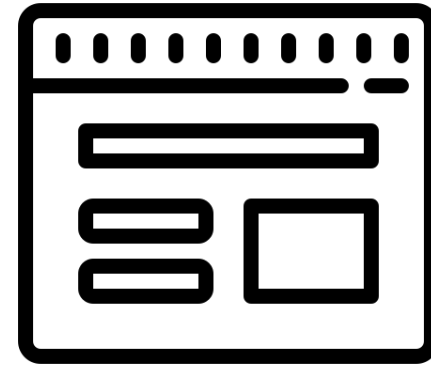
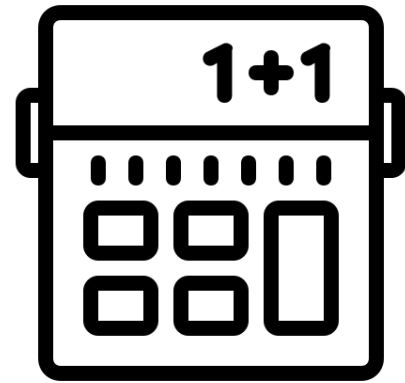
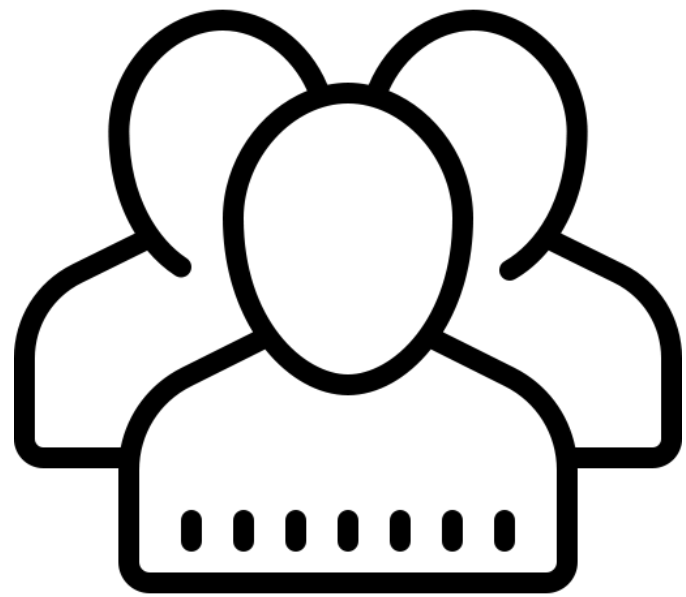
BILL & MELINDA
GATES *foundation*





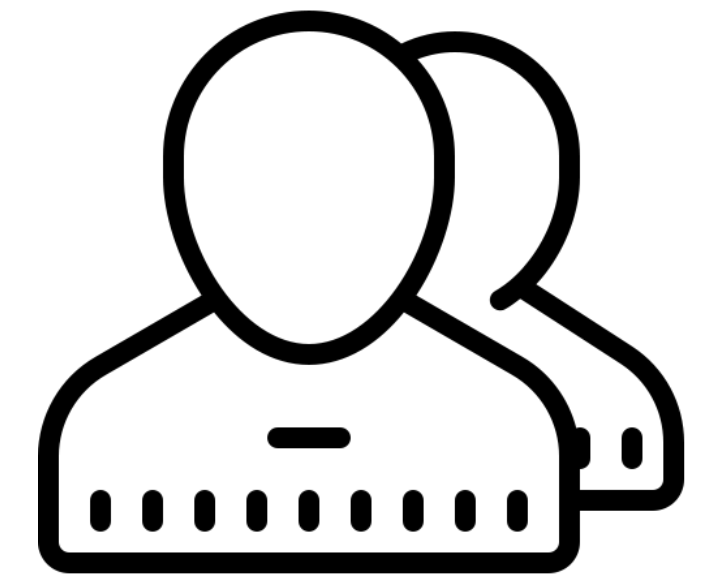
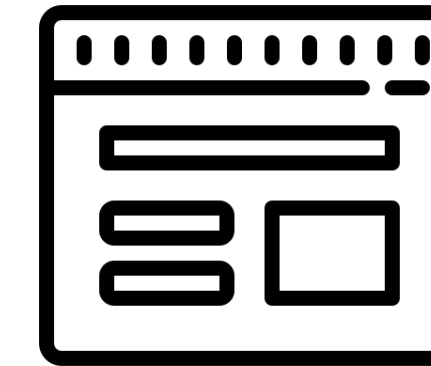
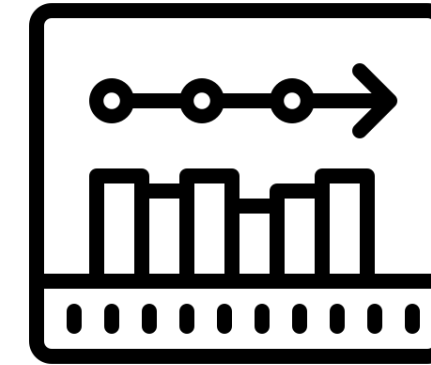
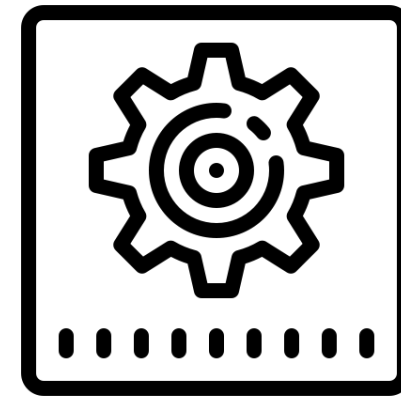
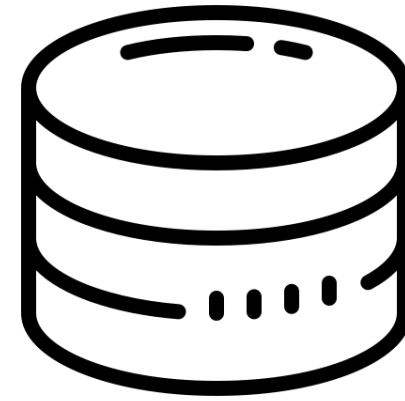
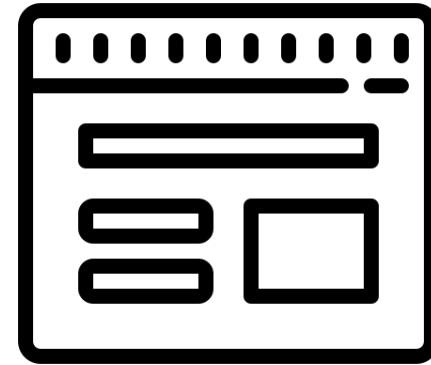
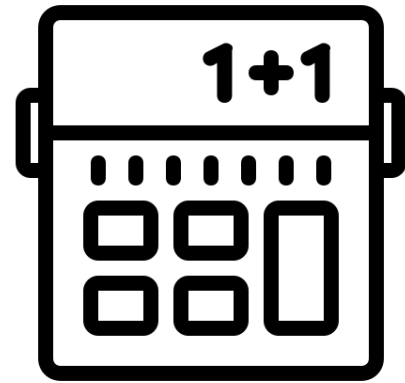
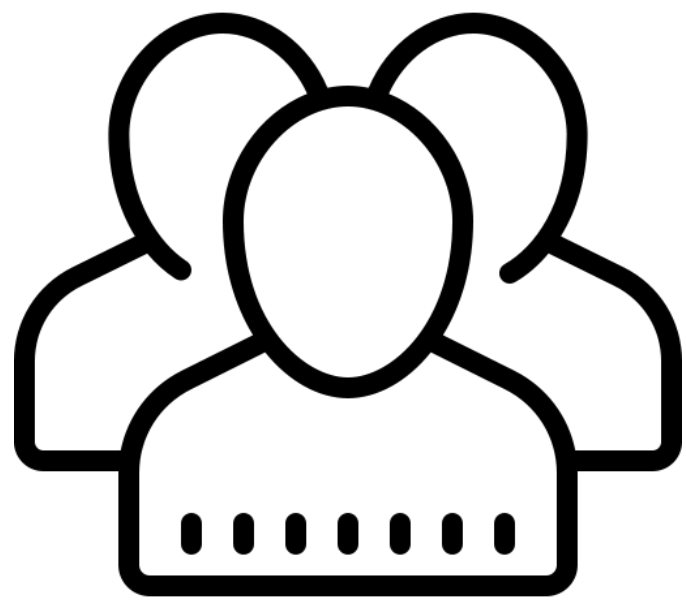
BILL & MELINDA
GATES *foundation*





BILL & MELINDA
GATES *foundation*

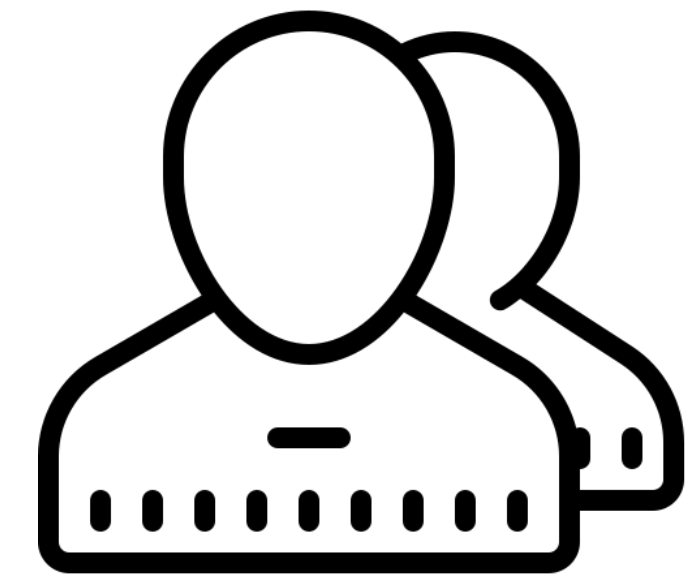
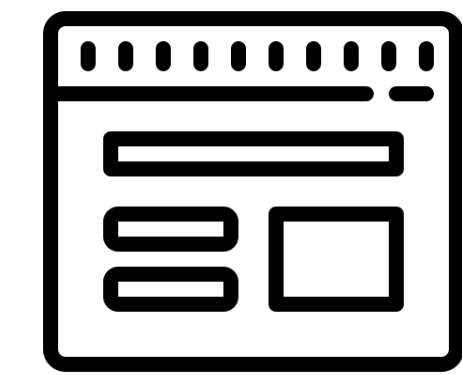
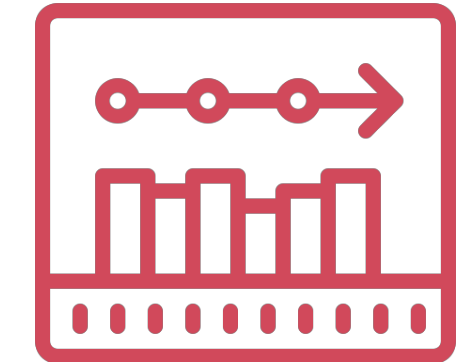
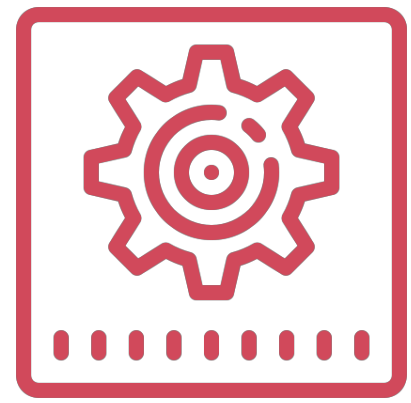
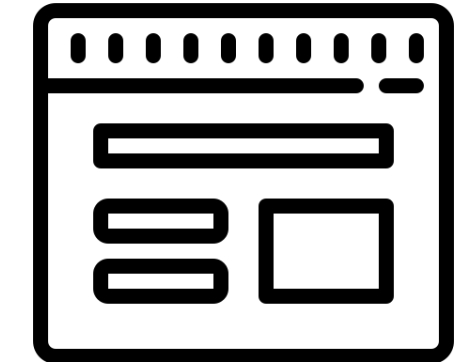
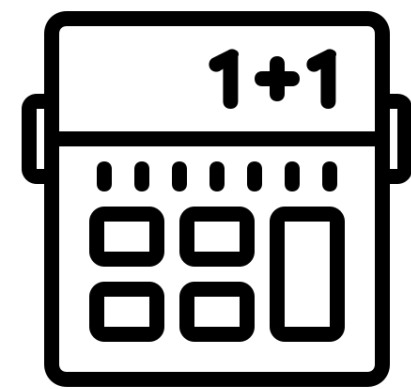
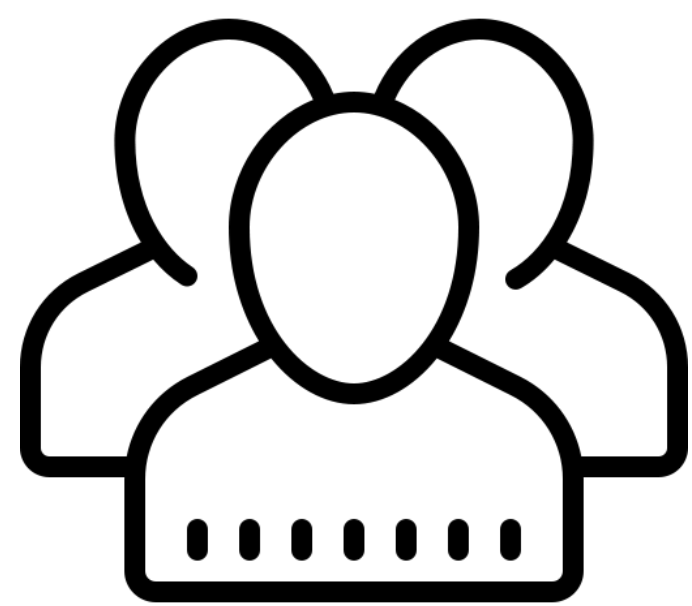




BILL & MELINDA
GATES *foundation*







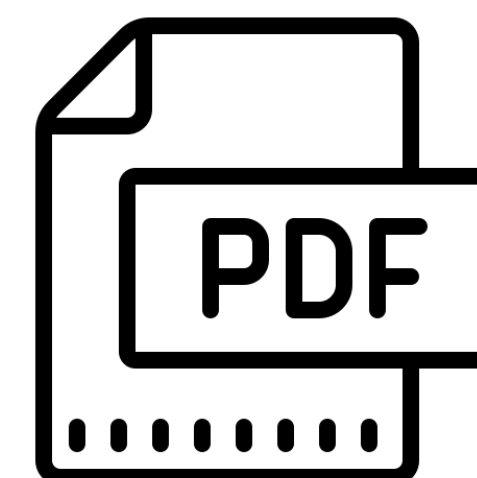
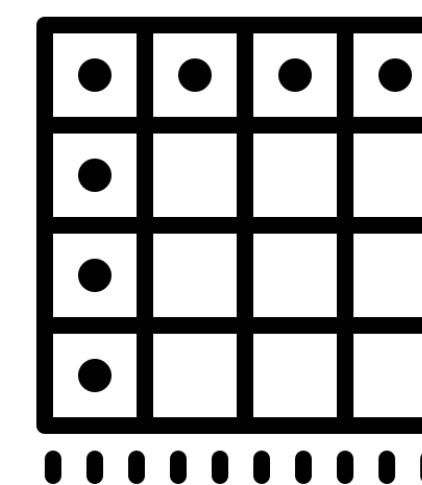
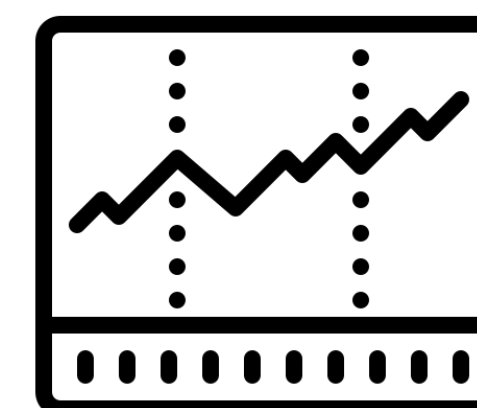
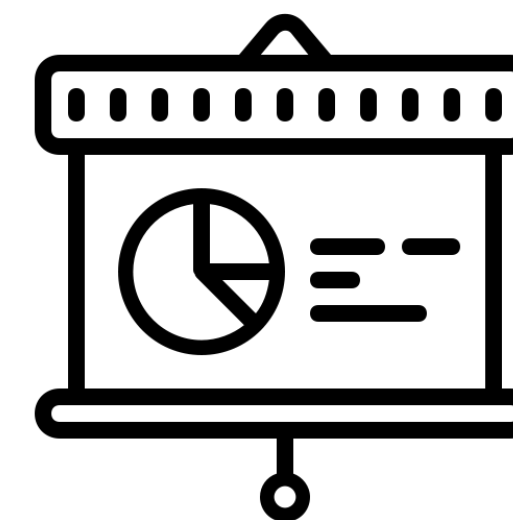
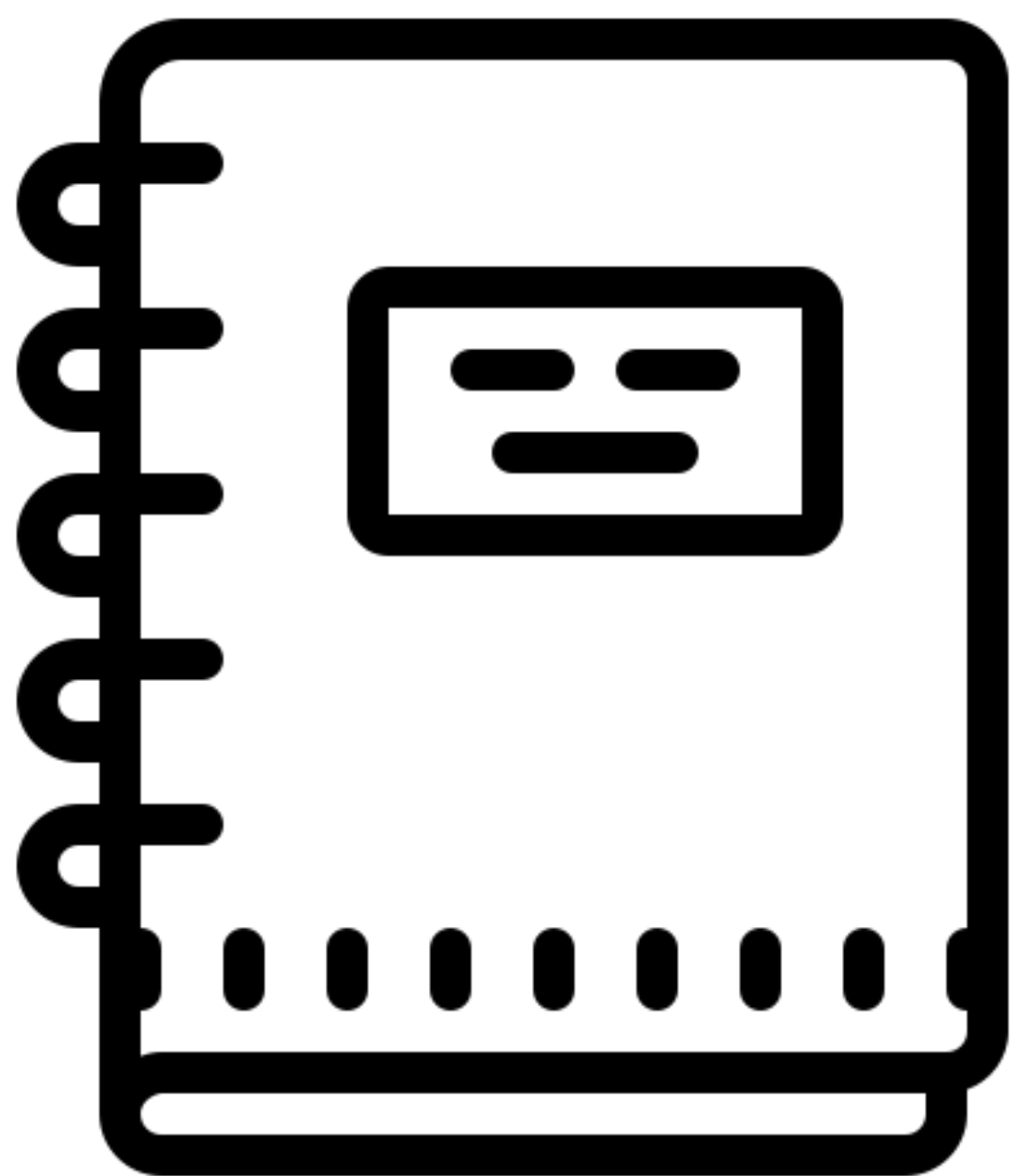
BILL & MELINDA
GATES *foundation*

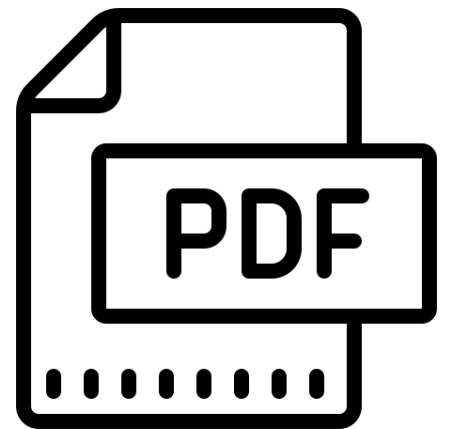
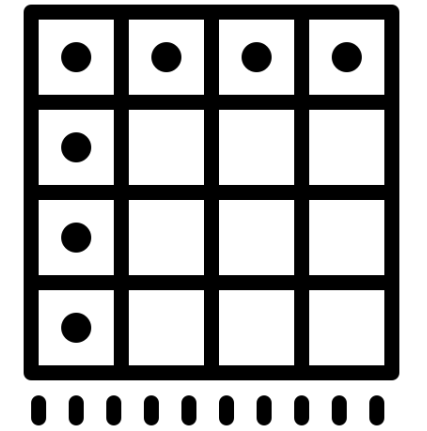
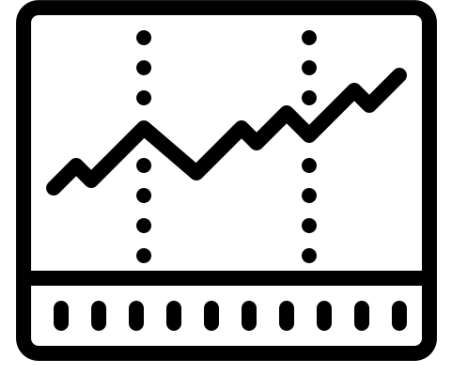
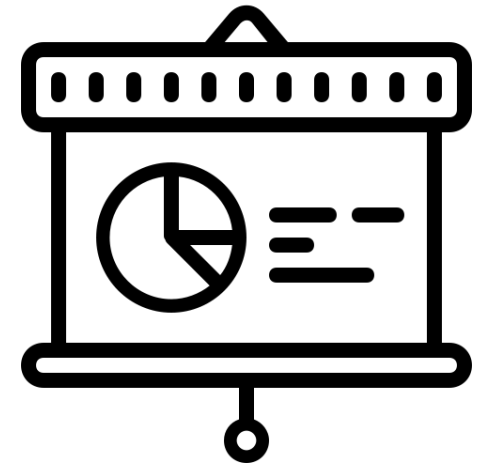
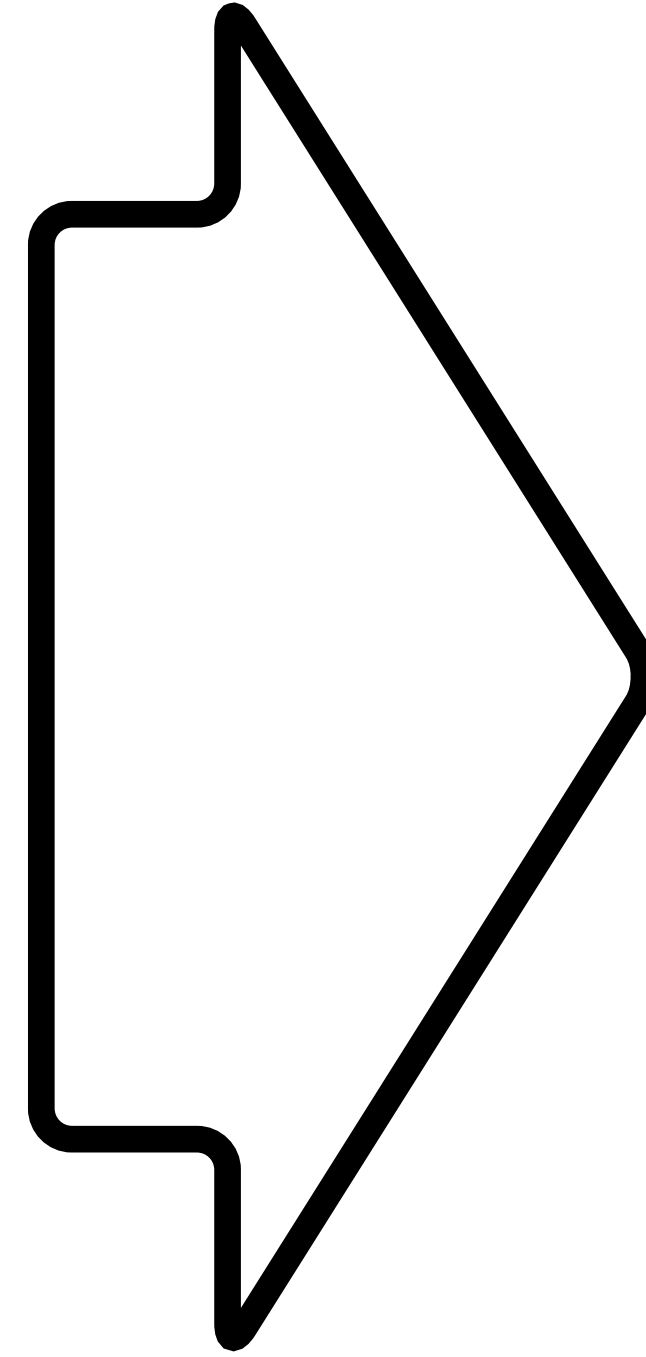
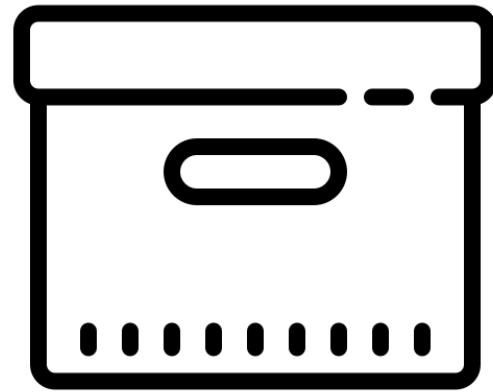
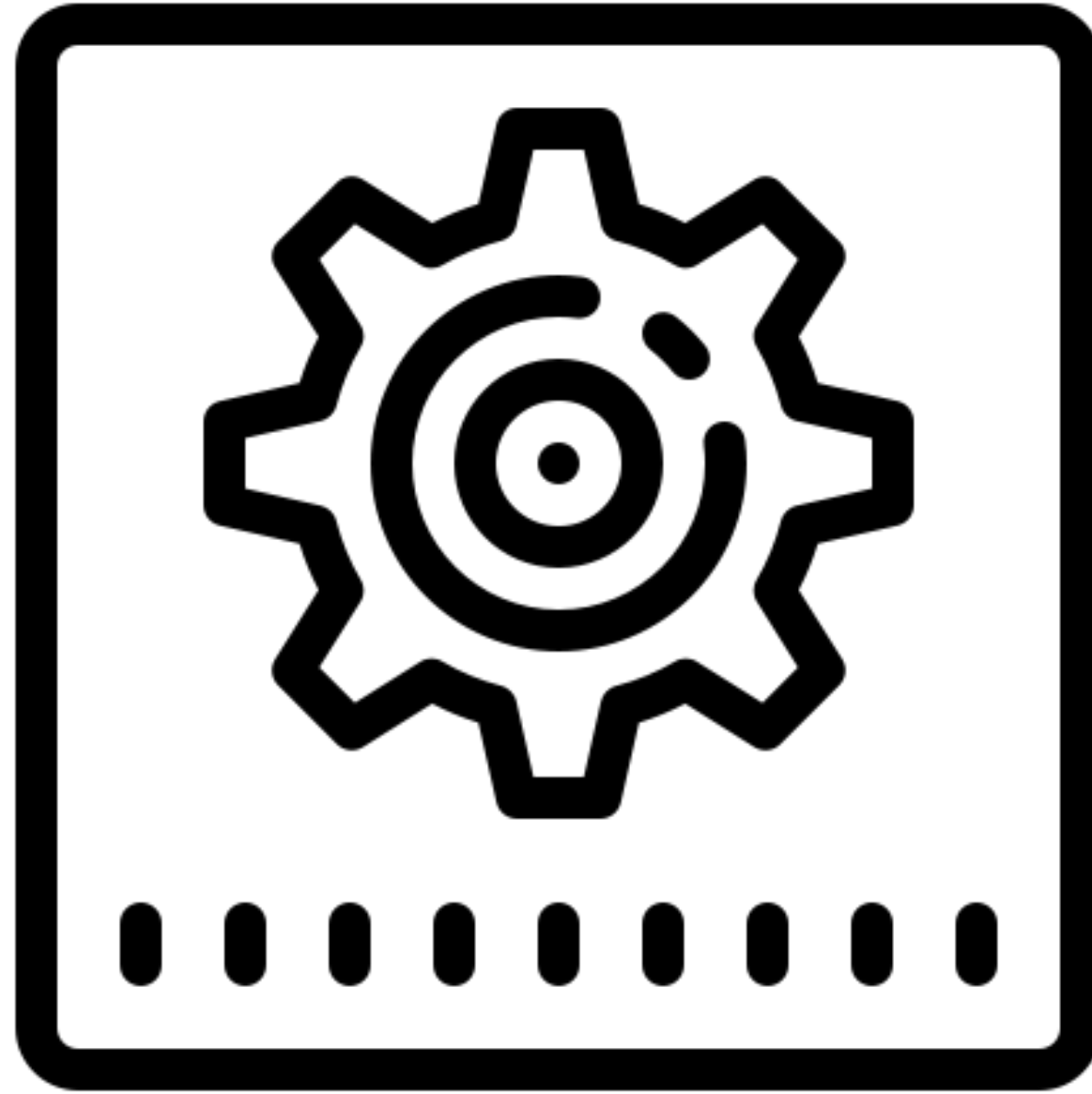
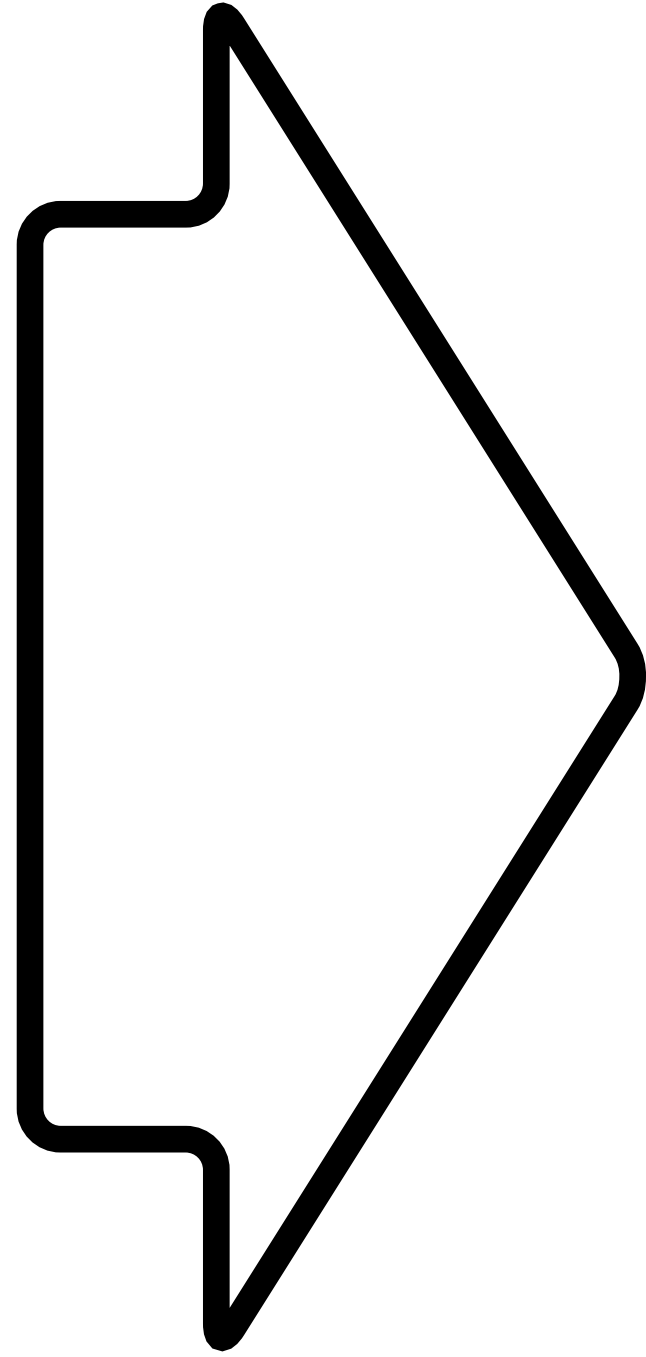
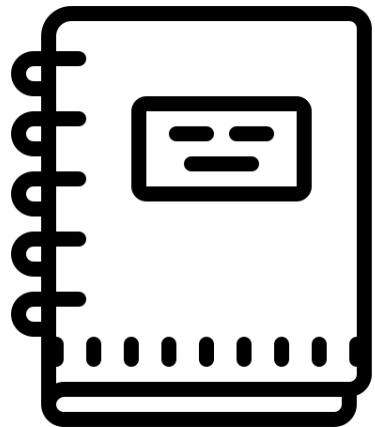
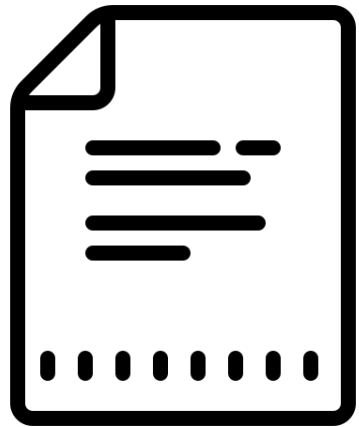
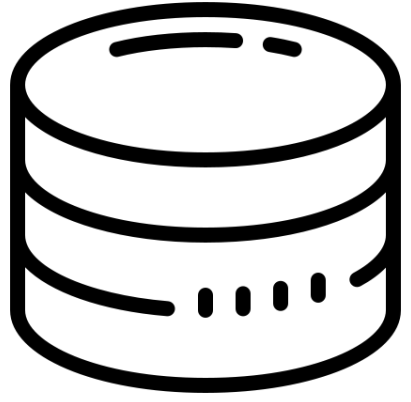
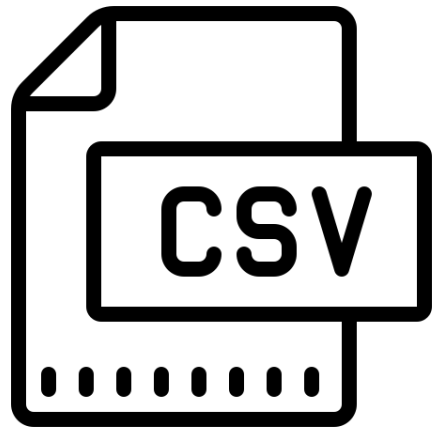


Our goal

Code written by domain scientists
runs on multiple machines, first time

Trace back to origins of changes
in any report





orderly

data:

```
summary: SELECT * FROM ...
```

resources:

- support.R
- metadata.csv

packages:

- ggplot2
- knitr

script: script.R

artefacts:

- report:

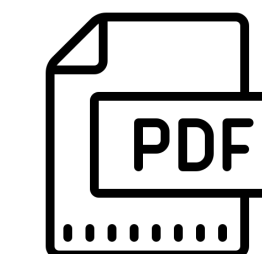
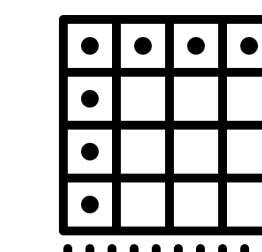
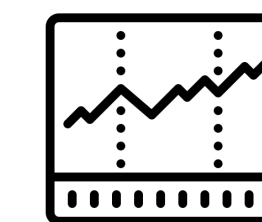
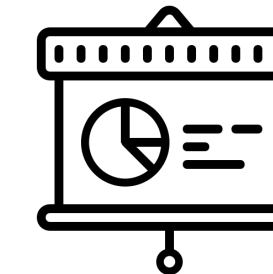
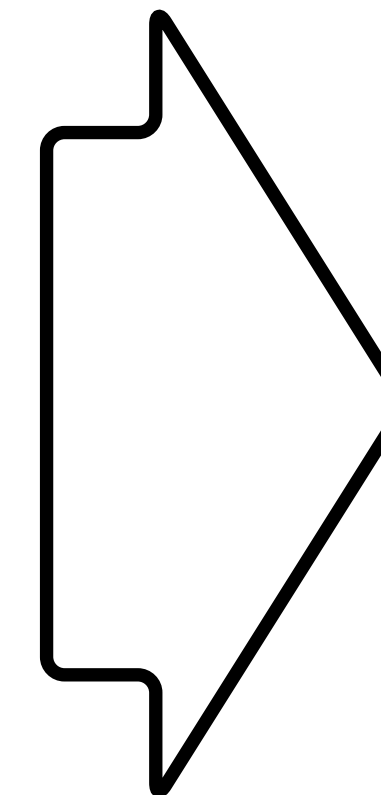
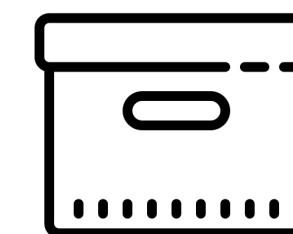
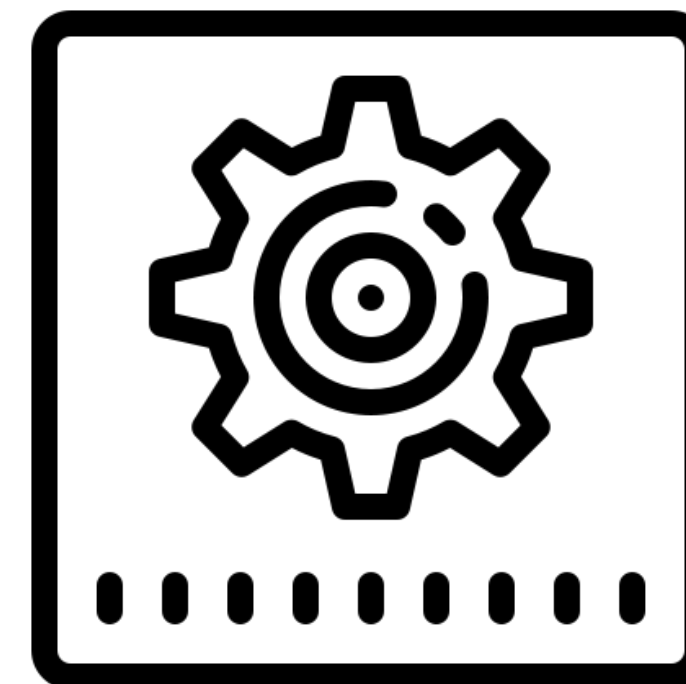
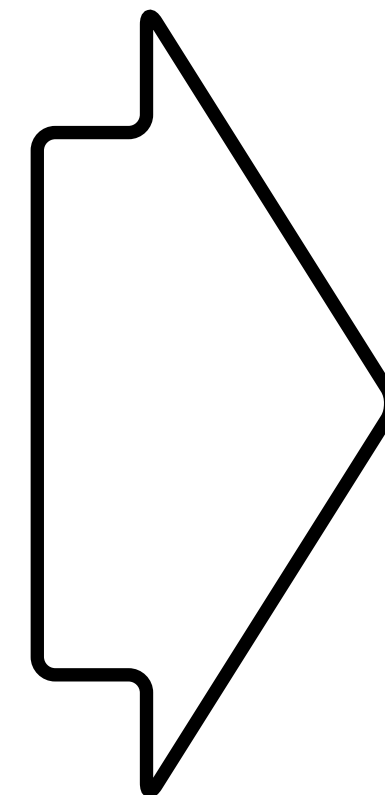
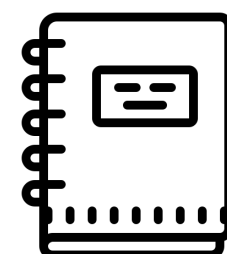
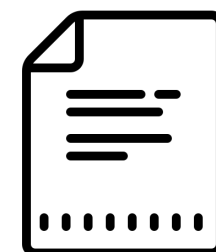
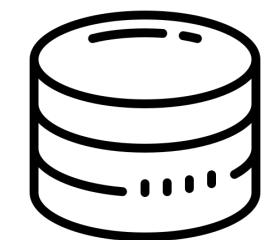
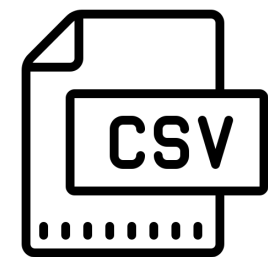
 - description: Summary of results

 - filenames: summary.pdf

- data:

 - description: Processed data for further use

 - filenames: data.csv



data:

```
summary: SELECT * FROM ...
```

resources:

- support.R
- metadata.csv

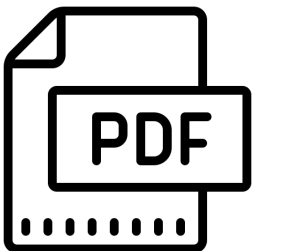
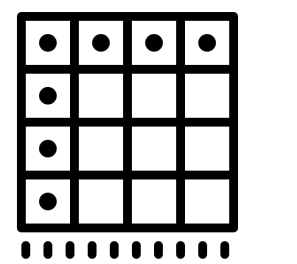
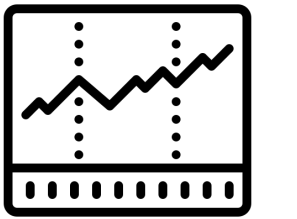
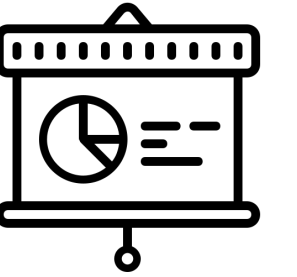
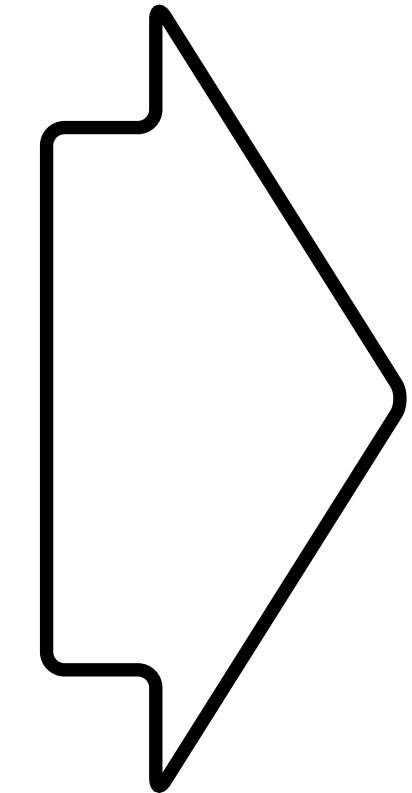
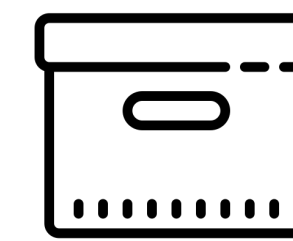
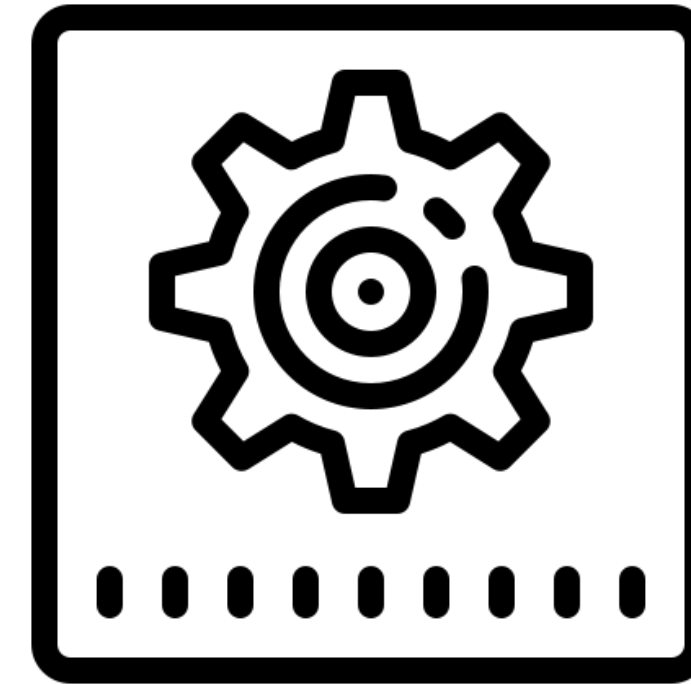
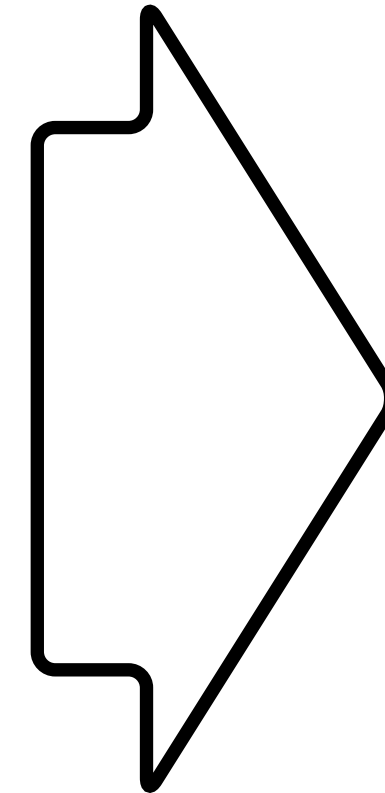
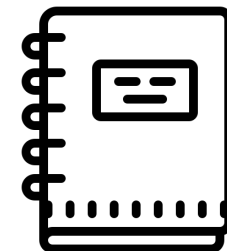
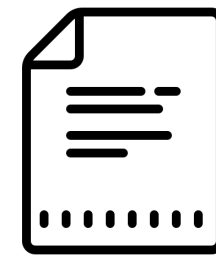
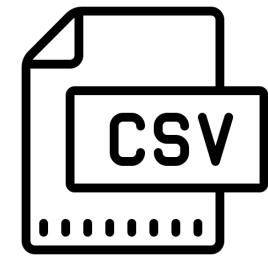
packages:

- ggplot2
- knitr

script: script.R

artefacts:

- report:
 - description: Summary of results
 - filenames: summary.pdf
- data:
 - description: Processed data for further use
 - filenames: data.csv



data:

```
summary: SELECT * FROM ...
```

resources:

- support.R
- metadata.csv

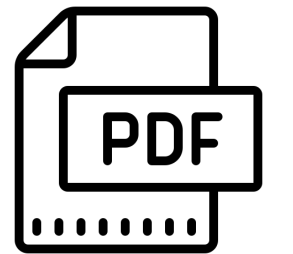
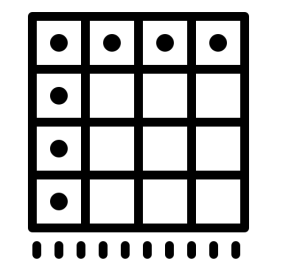
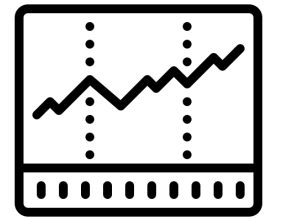
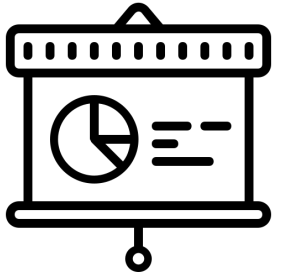
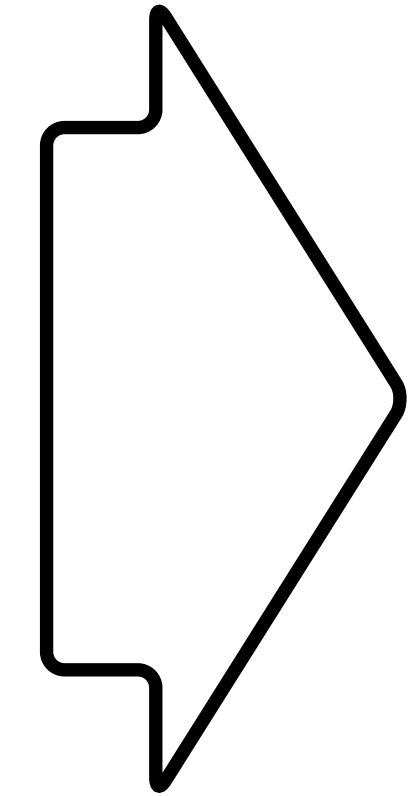
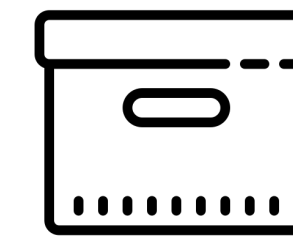
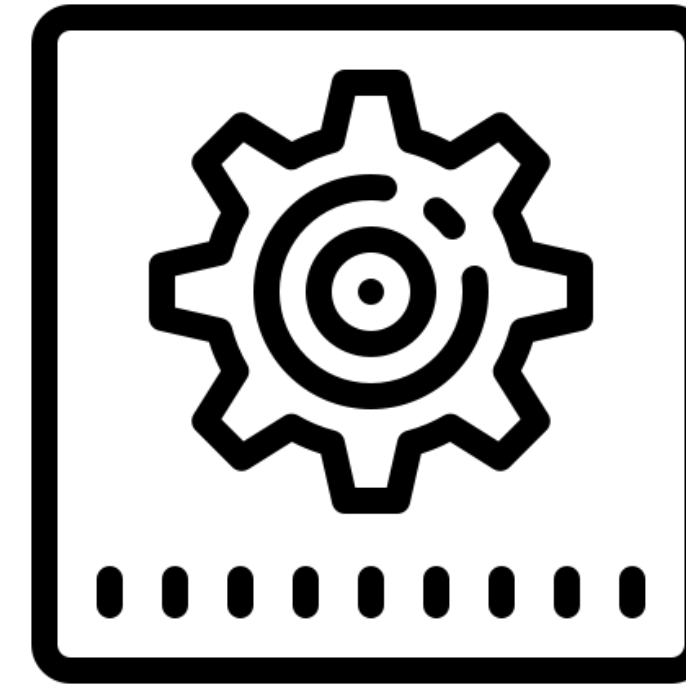
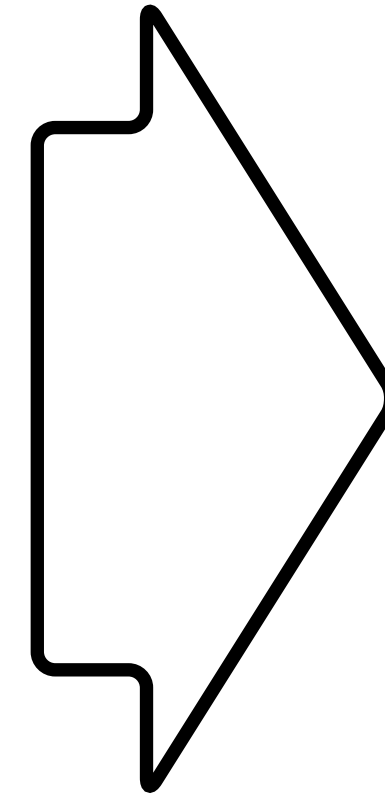
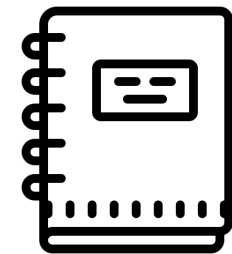
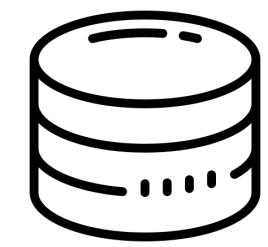
packages:

- ggplot2
- knitr

script: script.R

artefacts:

- report:
 - description: Summary of results
 - filenames: summary.pdf
- data:
 - description: Processed data for further use
 - filenames: data.csv



data:

```
summary: SELECT * FROM ...
```

resources:

- support.R
- metadata.csv

packages:

- ggplot2
- knitr

script: script.R

artefacts:

- report:

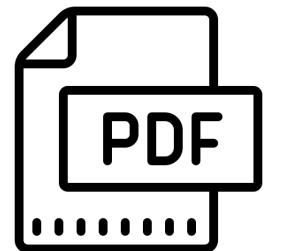
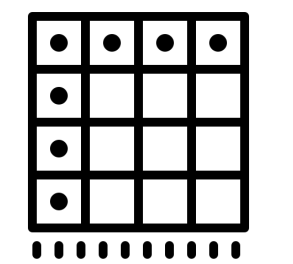
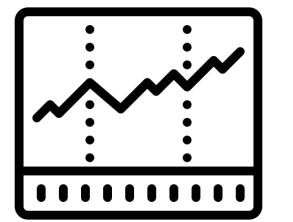
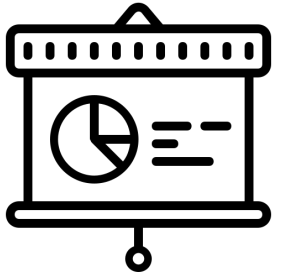
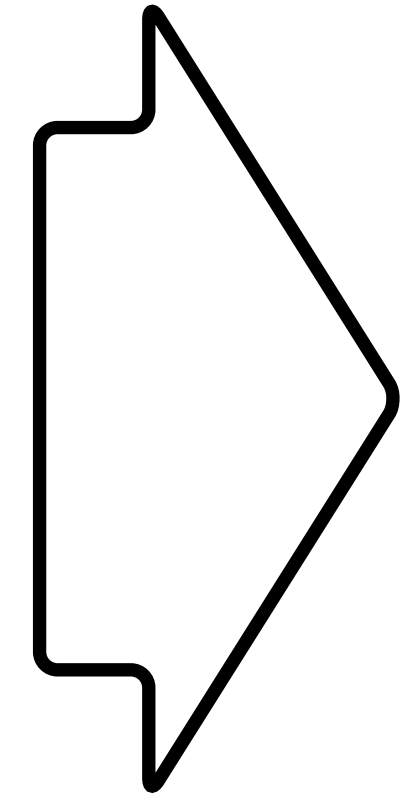
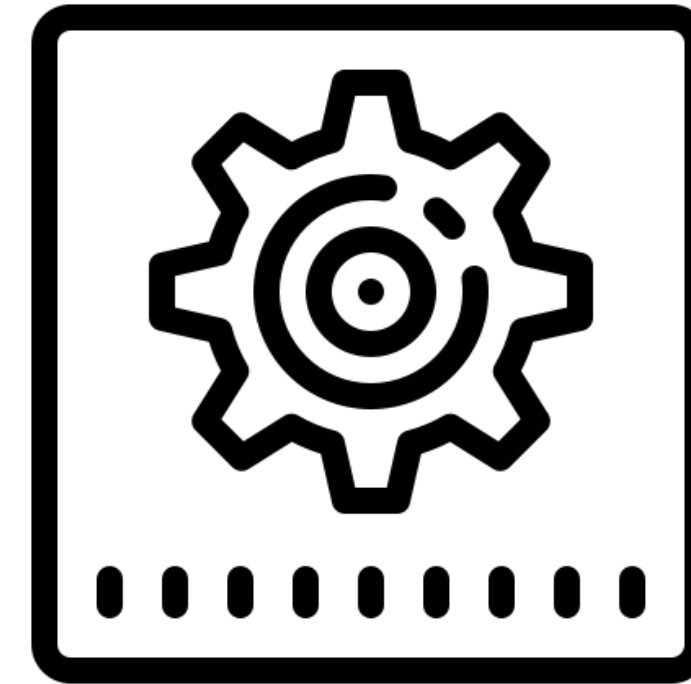
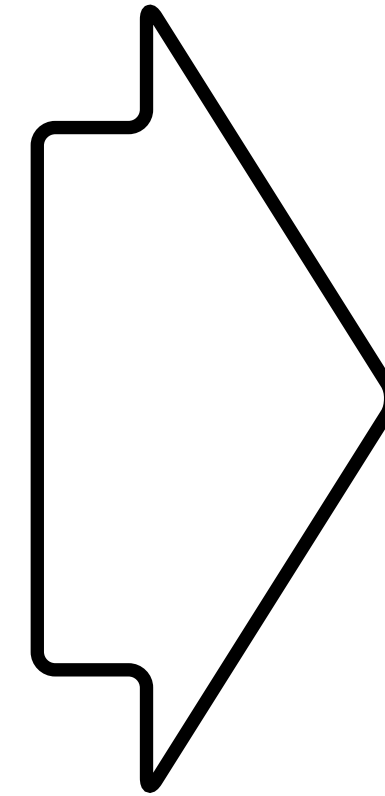
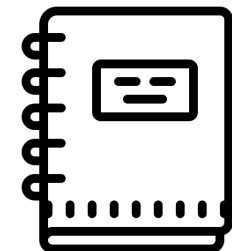
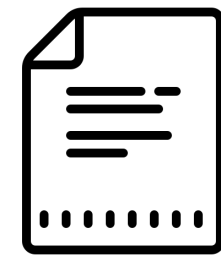
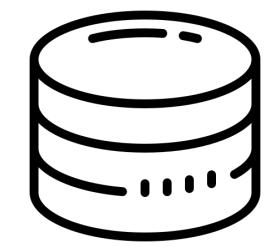
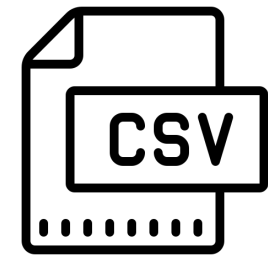
description: Summary of results

filenames: summary.pdf

- data:

description: Processed data for further use

filenames: data.csv



data:

```
summary: SELECT * FROM ...
```

resources:

- support.R
- metadata.csv

packages:

- ggplot2
- knitr

script: script.R

artefacts:

- report:

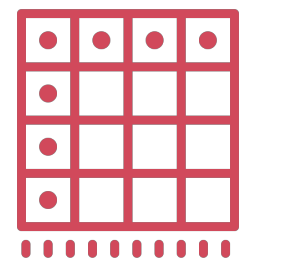
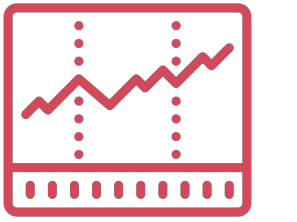
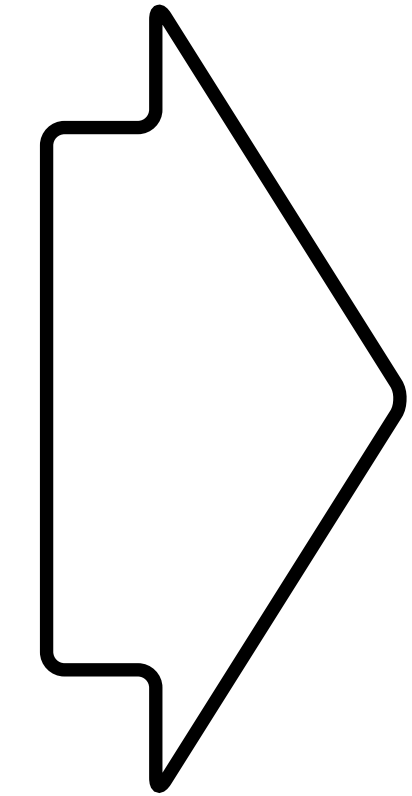
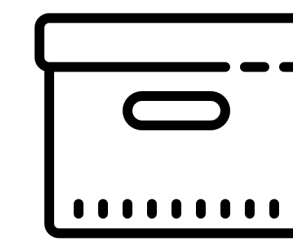
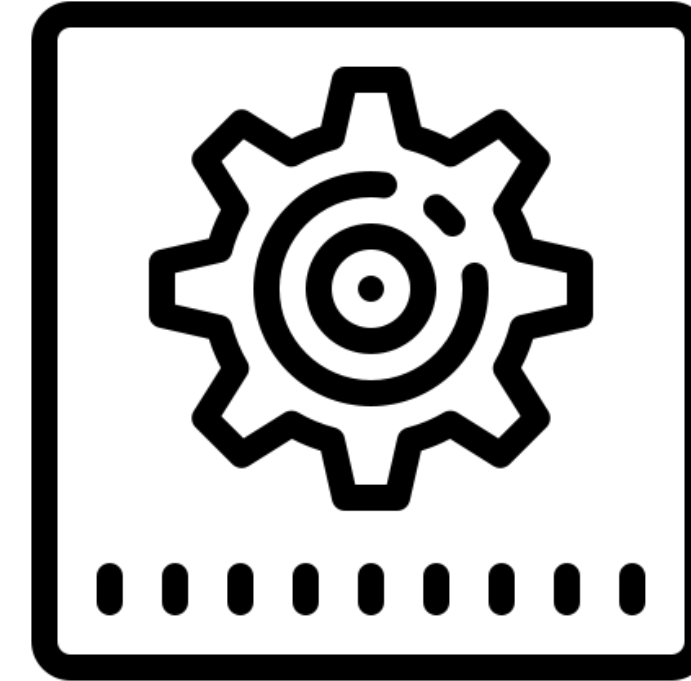
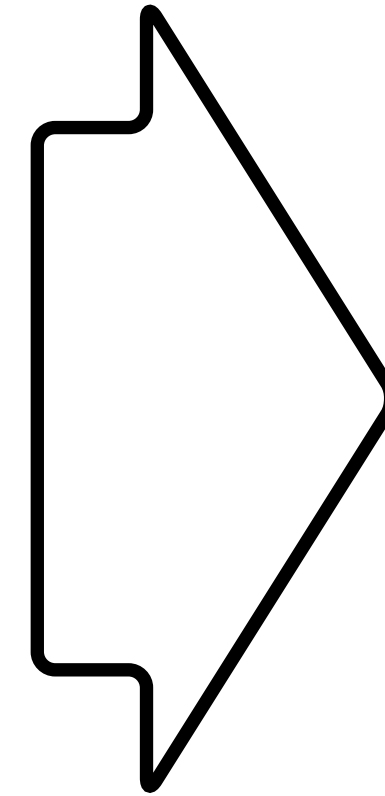
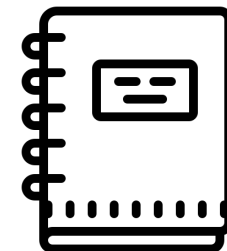
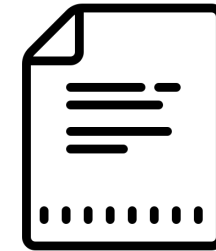
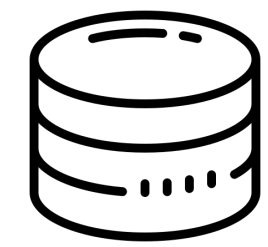
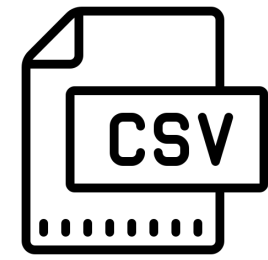
description: Summary of results

filenames: summary.pdf

- data:

description: Processed data for further use

filenames: data.csv



orderly_config.yml

src/

myreport/

orderly.yml

script.R

support.R

metadata.csv

orderly_config.yml

src/

myreport/

orderly.yml

script.R

support.R

metadata.csv

archive/

myreport/

20190204-143204-f5aa3bc9/

orderly.yml

script.R

support.R

metadata.csv

summary.pdf

data.csv

orderly_run.rds

orderly run myreport

orderly_config.yml

src/

myreport/

orderly.yml

script.R

support.R

metadata.csv

archive/

myreport/

20190204-143204-f5aa3bc9/

20190204-192249-3bc9f5aa/

orderly.yml

script.R

support.R

metadata.csv

summary.pdf

data.csv

orderly_run.rds

orderly run myreport

Automate the boring bits

This is embarrassingly
simple

Interface for Stakeholders

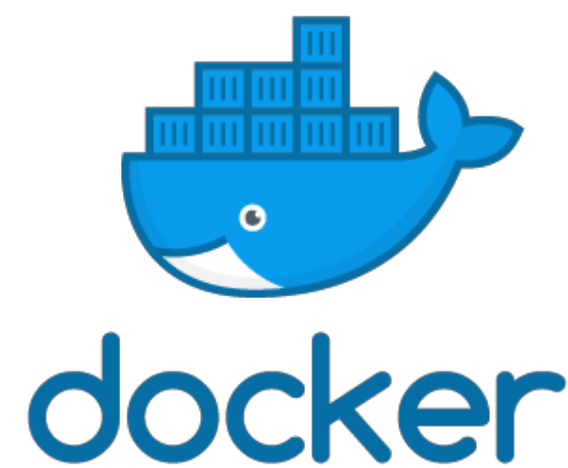
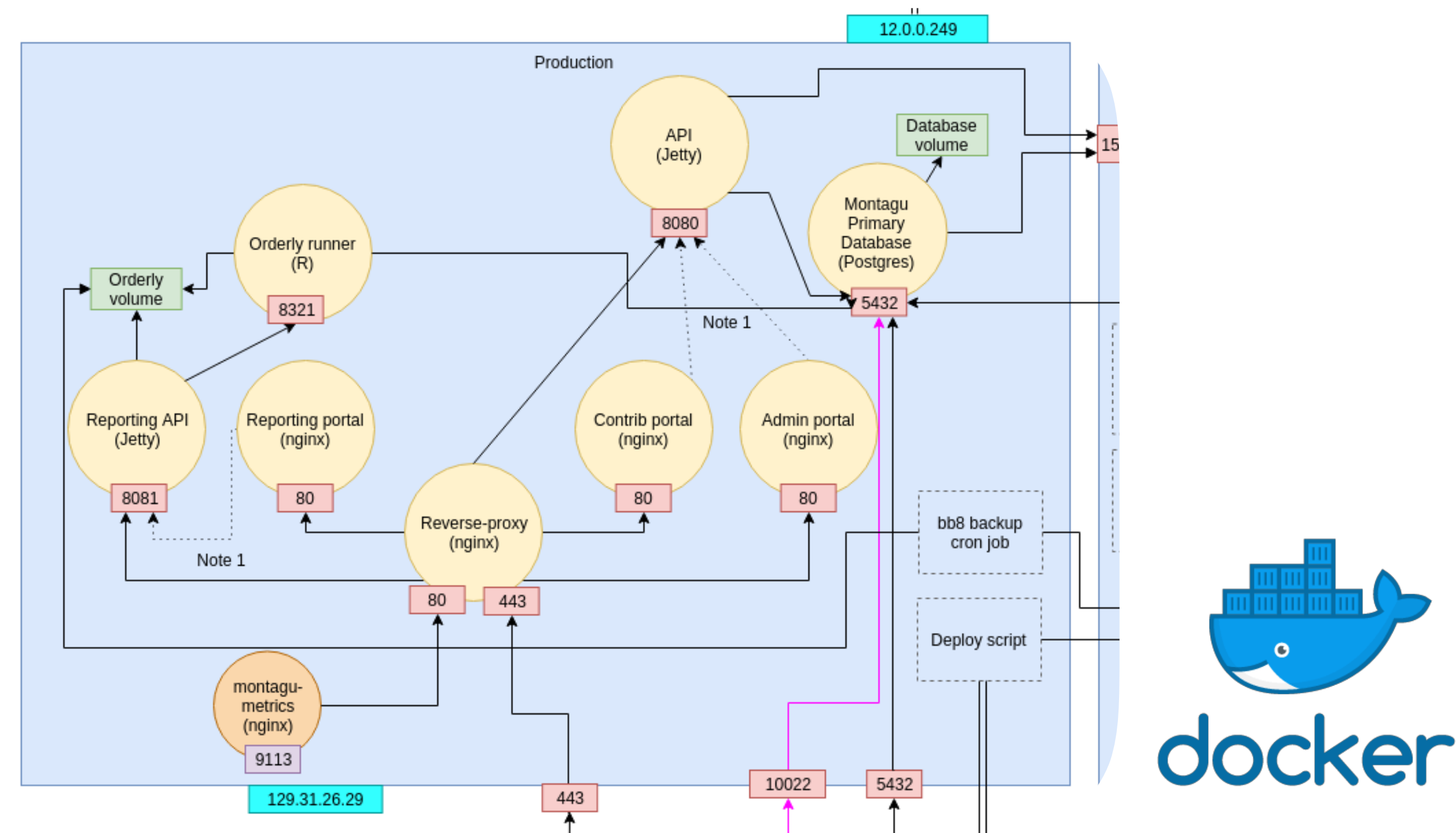
Find a report

Click on a column heading to sort by that field. Hold shift to multi-sort.

[Collapse all reports](#) / [Expand all reports](#)

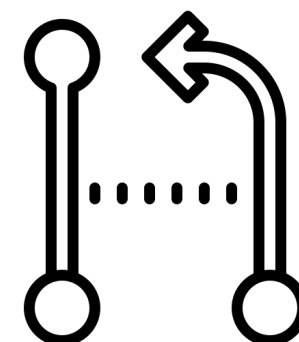
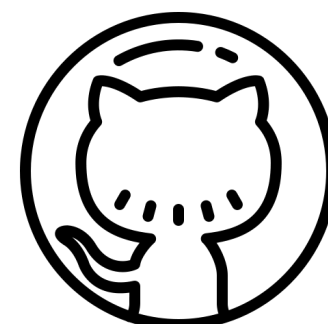
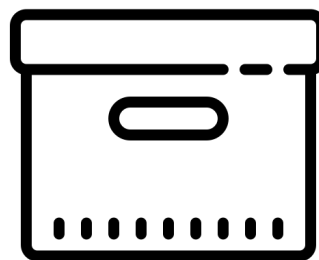
Name	Version	Status	Author	Requester
<input type="text" value="Type to filter..."/>	<input type="text" value="Mar 1, to Feb 6,"/> <input type="text" value="Type to filter by id..."/>	All ▾	<input type="text" value="Type to filter..."/>	<input type="text" value="Type to filter..."/>
▼ native-diagnostics-burden-report-drafts 5 versions: view latest			Science Team	VIMC
	Tue Feb 05 2019 latest (20190205-151702-1ba5e47a)	published	Science Team	VIMC
	Thu Jan 31 2019 out-dated (20190131-162935-86a83d30)	published	Science Team	VIMC
	Thu Jan 31 2019 out-dated (20190131-123847-53fe189e)	published	Science Team	VIMC
	Mon Jan 28 2019 out-dated (20190128-151914-a2c1edce)	published	Science Team	VIMC
	Mon Jan 28 2019 out-dated (20190128-125432-31d88275)	internal	Science Team	VIMC

Interface for Technical team



Interface for Science team

```
install.packages("orderly")  
orderly::orderly_new("myreport")
```



Interface for Other groups

Lessons learnt

you can blackbox too much

code reuse is really hard

reproducibility can be easy

problems are as social as technical

Our team



Work the way people
want to work

