

Organising Yo

Data Science Wor

Dr Zak Varty

Week 1: What are we trying

- Data science is a **collaborative discipline**
- Be a good collaborator, to others and to yourself
- This week will show one framework to help you
- Like flossing not difficult but requires discipline
- We will take an opinionated and R focused approach to the settings.

One Project = One Director

- Sounds easy, but in practice it is not.
- Requires prospective project scoping.
- Entropy is ever increasing.

Properties of a well organised project

Ideally, we would like to organise our projects so that they have the following properties:

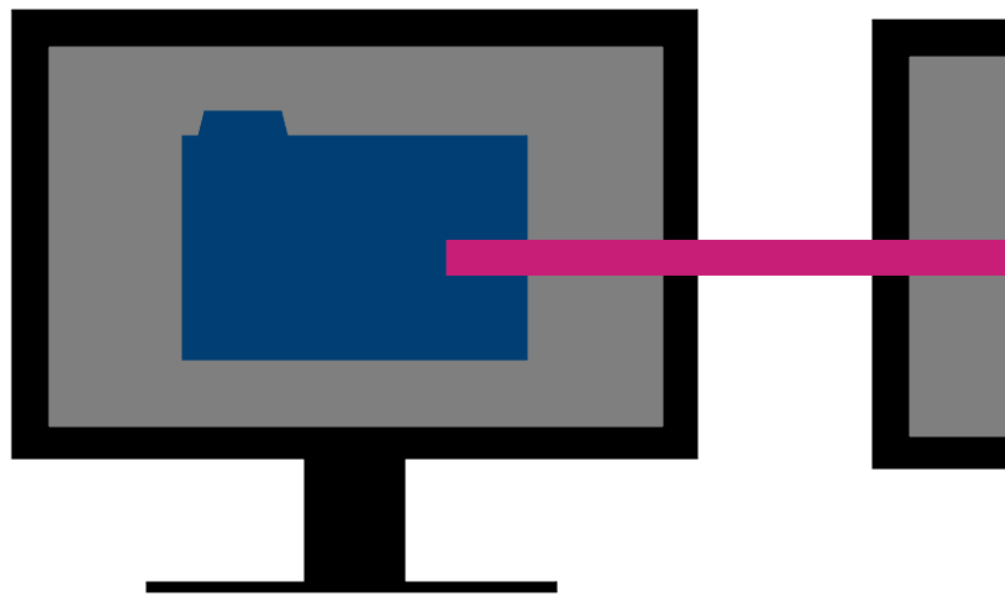
- Portable
- Version control friendly
- Reproducible
- Integrated Development Environment friendly

Portability

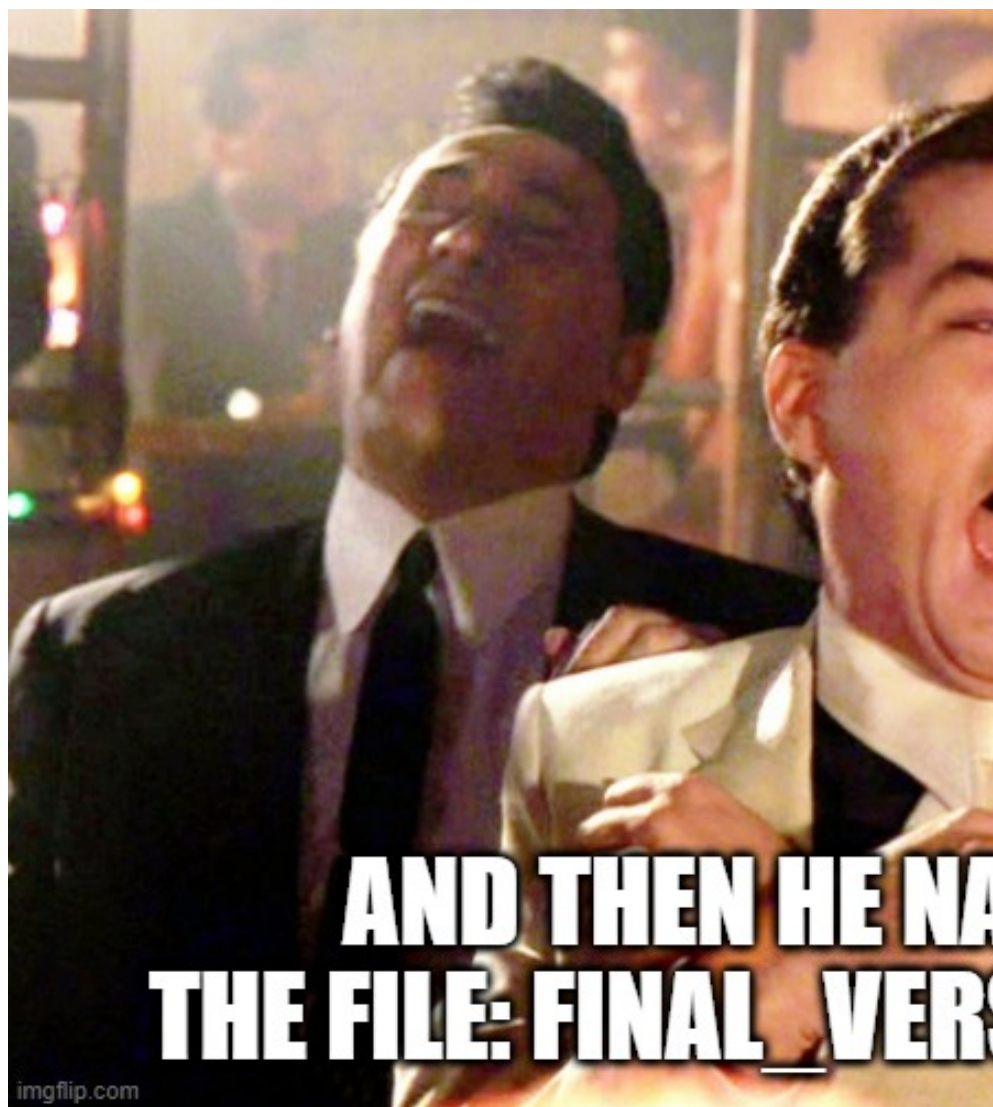
Is your work all in one place or scattered?

Can it be moved to a new location without breaking?

What do we mean by a new location, exactly?



Version Control Friendly



Reproducible

A study is **reproducible** if you can take the original code used to analyze the data and recreate all the results of the study.

Broman et al. (2017) “Recommendations to Foster Reproducible Research”

Integrated Development En

- Possible to code and manage projects entire line.
- Puts a lot of strain on you, both your fingers and your brain.
- Integrated development environments such as VisualStudio aim to reduce this burden
 - Autocomplete / shortcodes
 - Environment panes
 - Templating



exciting-



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

```
# M
Her
the
- A
- A
- A
The
as
##
Thi
der
car
[me
```

Inside the README

- Name
- Project Status
- Description
- Installation
- Usage / Examples
- Support
- Contributing



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile



exciting-new-pr

README.md

data/

raw

src/

dev

tests/

analyses/

outputs/

reports/

makefile



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

raw

dev



exciting-new-pr



README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

data

headers



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

data

headers



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

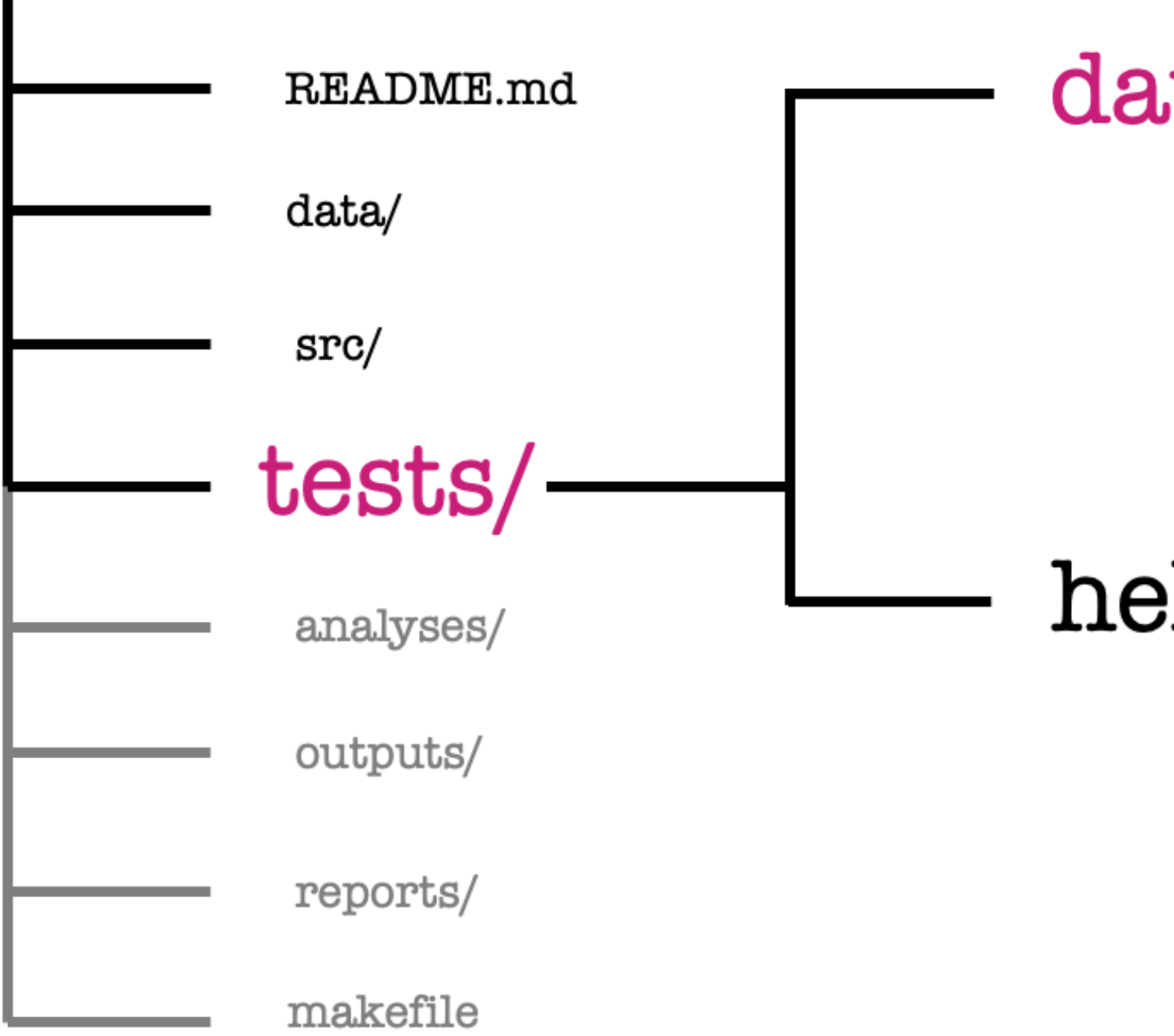
makefile

data

headers



exciting-new-pr





exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

data

headers



exciting-new-pr

README.md

data/

src/

tests/

analyses/

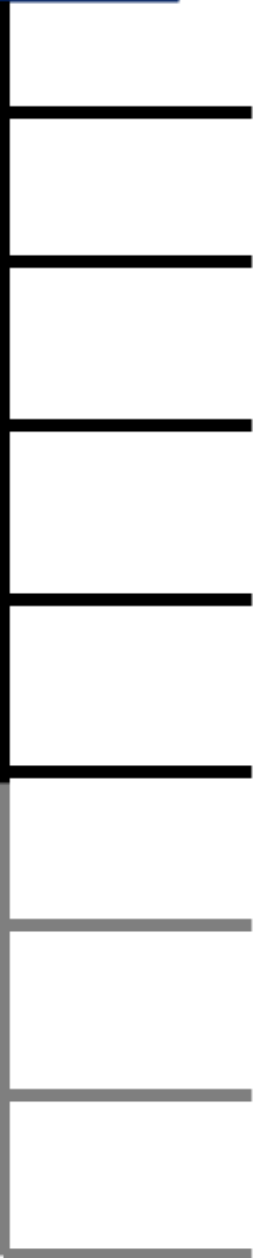
outputs/

reports/

makefile



exciting-new-pr



README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile



2

S



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

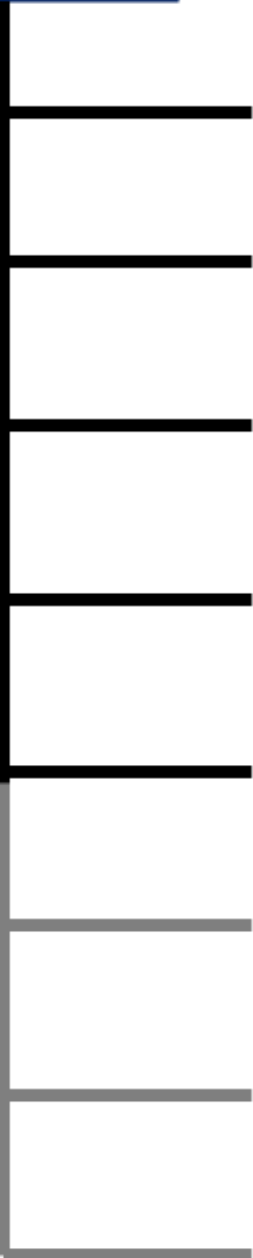
makefile

2

S



exciting-new-pr



README.md

data/

src/

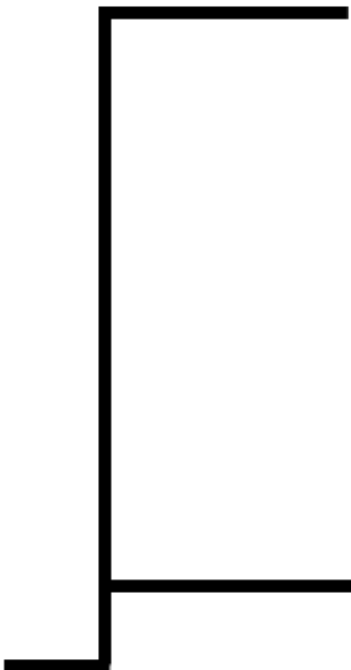
tests/

analyses/

outputs/

reports/

makefile

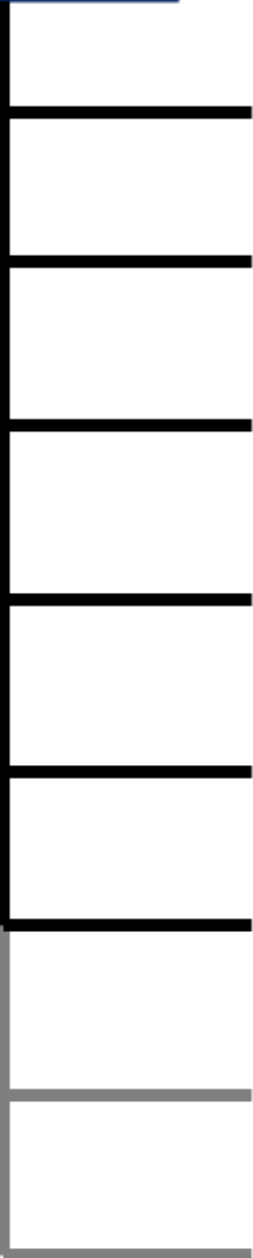


2

S



exciting-new-pr



README.md

data/

src/

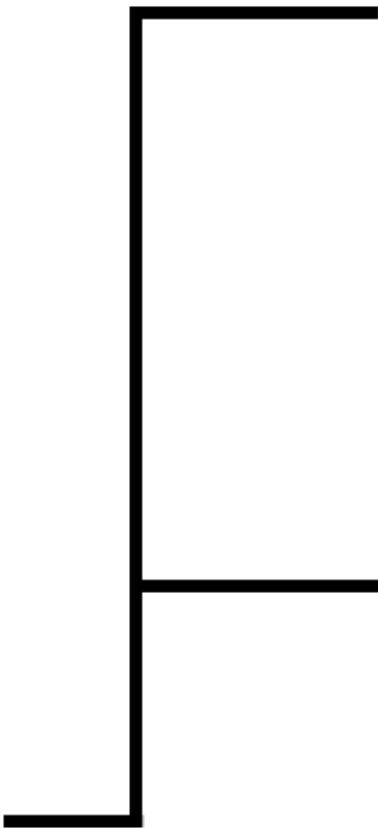
tests/

analyses/

outputs/

reports/

makefile

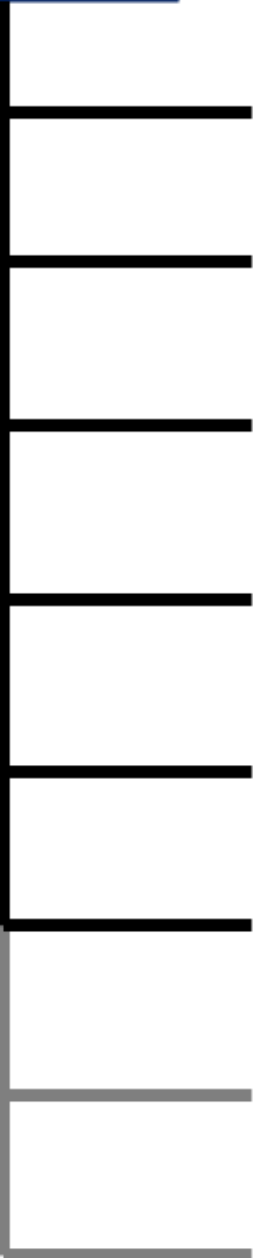


2

S



exciting-new-pr



README.md

data/

src/

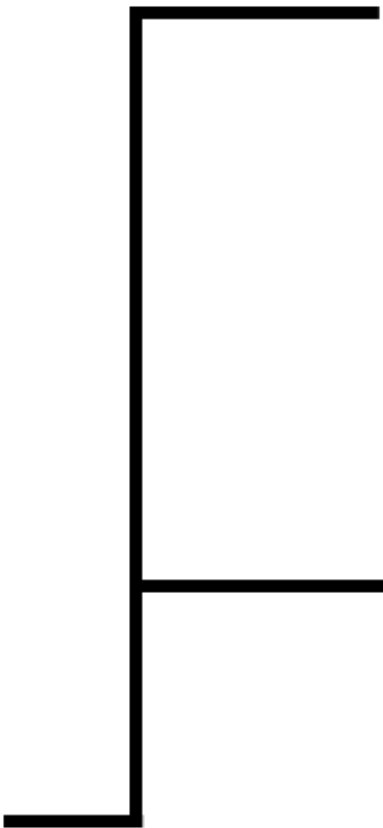
tests/

analyses/

outputs/

reports/

makefile



2

S



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

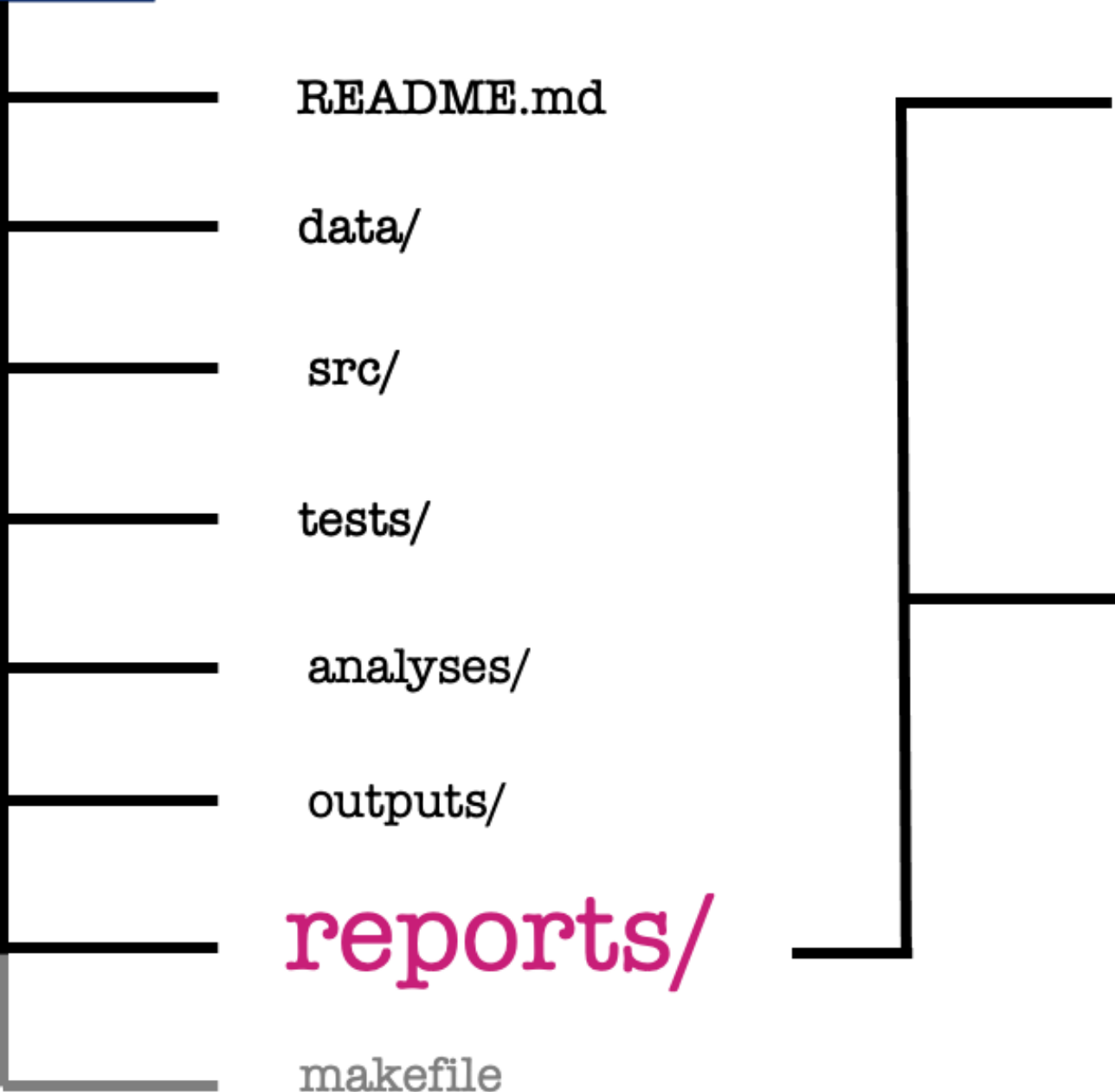
makefile

2

S



exciting-new-pr



2

S



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile

2

S



exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

makefile





exciting-new-pr

README.md

data/

src/

tests/

analyses/

outputs/

reports/

[makefile]

Summary

- Introduced a standardised project structure;
- Good starting point for most data science projects.
- Exceptions are **apps** and **packages**.

References

Broman, Karl, Mine Cetinkaya-Rundel, Amy Nussbaum, Christopher
Wickham. 2017. “Recommendations to Funding Agencies for S
Statistical Association, 2:1–4.