



Introduction to **principles** of digital project management

Dr. Ian Hussey

Digitalisation of Psychology

Make projects & decisions understandable
for other people*

*including your future self

{find-the-data exercise}

Knit early, Knit often

The minimal test of whether your code works is to knit your R Markdown file.

If it doesn't work for you, it won't work for others.



Separate code and data

GitHub repo for project

```
├── method
│   ├── qualtrics implementation
│   ├── self-report scales
│   └── self-report scales.docx
├── code
│   ├── analysis.Rmd
│   ├── processing.Rmd
│   ├── models
│   │   └── model_fit.rds
│   └── plots
│       ├── figure_1.pdf
│       └── figure_2.pdf
├── data
│   ├── raw
│   │   ├── data_raw.csv
│   │   └── data_raw_codebook.xlsx
│   ├── processed
│   │   ├── data_processed.csv
│   │   └── data_processed_codebook.xlsx
│   └── results
│       ├── results_summary_statistics.csv
│       └── results_regression_table.csv
├── communication
│   ├── preprint
│   │   └── hussey et al. - 2024 - title.docx
│   └── presentations
│       └── hussey et al. - 2024 - conference presentation.pptx
├── README.md
└── LICENSE
```

Separate raw data from processed data

Alter the raw data only as much as you are legally and ethically obliged to

Eg deanonymise it

Other than this, these files should only be read.

GitHub repo for project

```
├── method
│   ├── qualtrics implementation
│   ├── self-report scales
│   └── self-report scales.docx
├── code
│   ├── analysis.Rmd
│   ├── processing.Rmd
│   ├── models
│   │   └── model_fit.rds
│   └── plots
│       ├── figure_1.pdf
│       └── figure_2.pdf
├── data
│   ├── raw
│   │   ├── data_raw.csv
│   │   └── data_raw_codebook.xlsx
│   ├── processed
│   │   ├── data_processed.csv
│   │   └── data_processed_codebook.xlsx
│   └── results
│       ├── results_summary_statistics.csv
│       └── results_regression_table.csv
├── communication
│   ├── preprint
│   │   └── hussey et al. - 2024 - title.docx
│   └── presentations
│       └── hussey et al. - 2024 - conference presentation.pptx
├── README.md
└── LICENSE
```


Include a data dictionary for each data file

You can create these manually or use various R packages.

They can greatly vary in their complexity.

But they should at minimum explain what variables represent.

GitHub repo for project

```
├── method
│   ├── qualtrics implementation
│   ├── self-report scales
│   └── self-report scales.docx
├── code
│   ├── analysis.Rmd
│   ├── processing.Rmd
│   ├── models
│   │   └── model_fit.rds
│   └── plots
│       ├── figure_1.pdf
│       └── figure_2.pdf
├── data
│   ├── raw
│   │   ├── data_raw.csv
│   │   └── data_raw_codebook.xlsx
│   ├── processed
│   │   ├── data_processed.csv
│   │   └── data_processed_codebook.xlsx
│   └── results
│       ├── results_summary_statistics.csv
│       └── results_regression_table.csv
├── communication
│   ├── preprint
│   │   └── hussey et al. - 2024 - title.docx
│   └── presentations
│       └── hussey et al. - 2024 - conference presentation.pptx
├── README.md
└── LICENSE
```


Separate processing code vs analysis code

Do not do analyses in processing
or processing in analyses

Eg create exclusion variables in processing,
but apply exclusions in the analysis

GitHub repo for project

```
├── method
│   ├── qualtrics implementation
│   ├── self-report scales
│   └── self-report scales.docx
├── code
│   ├── analysis.Rmd
│   ├── processing.Rmd
│   ├── models
│   │   └── model_fit.rds
│   └── plots
│       ├── figure_1.pdf
│       └── figure_2.pdf
├── data
│   ├── raw
│   │   ├── data_raw.csv
│   │   └── data_raw_codebook.xlsx
│   ├── processed
│   │   ├── data_processed.csv
│   │   └── data_processed_codebook.xlsx
│   └── results
│       ├── results_summary_statistics.csv
│       └── results_regression_table.csv
├── communication
│   ├── preprint
│   │   └── hussey et al. - 2024 - title.docx
│   └── presentations
│       └── hussey et al. - 2024 - conference presentation.pptx
├── README.md
└── LICENSE
```

Link your specific analyses to your manuscript

In analysis.Rmd, link each analysis, table, or plot to how it is used in the manuscript.

Eg

‘Figure 1 in manuscript’

‘Hypothesis 1 in manuscript’

GitHub repo for project

```
├── method
│   ├── qualtrics implementation
│   ├── self-report scales
│   └── self-report scales.docx
├── code
│   ├── analysis.Rmd
│   ├── processing.Rmd
│   ├── models
│   │   └── model_fit.rds
│   └── plots
│       ├── figure_1.pdf
│       └── figure_2.pdf
├── data
│   ├── raw
│   │   ├── data_raw.csv
│   │   └── data_raw_codebook.xlsx
│   ├── processed
│   │   ├── data_processed.csv
│   │   └── data_processed_codebook.xlsx
│   └── results
│       ├── results_summary_statistics.csv
│       └── results_regression_table.csv
├── communication
│   ├── preprint
│   │   └── hussey et al. - 2024 - title.docx
│   └── presentations
│       └── hussey et al. - 2024 - conference presentation.pptx
├── README.md
└── LICENSE
```

Include a README.md or README.txt

.md is just a text only markdown file, can't run code. Can be opened in RStudio or dedicated editors like Typora.

Include specific instructions on how to reproduce your results!

What files to run, in what order, and where to find the output.

GitHub repo for project

```
├── method
│   ├── qualtrics implementation
│   ├── self-report scales
│   └── self-report scales.docx
├── code
│   ├── analysis.Rmd
│   ├── processing.Rmd
│   ├── models
│   │   └── model_fit.rds
│   └── plots
│       ├── figure_1.pdf
│       └── figure_2.pdf
├── data
│   ├── raw
│   │   ├── data_raw.csv
│   │   └── data_raw_codebook.xlsx
│   ├── processed
│   │   ├── data_processed.csv
│   │   └── data_processed_codebook.xlsx
│   └── results
│       ├── results_summary_statistics.csv
│       └── results_regression_table.csv
├── communication
│   ├── preprint
│   │   └── hussey et al. - 2024 - title.docx
│   └── presentations
│       └── hussey et al. - 2024 - conference presentation.pptx
├── README.md
└── LICENSE
```

Include a license!

Don't make people guess if they can reuse your data or code

Don't leave your project in legal limbo

'CC-By 4.0' or 'MIT license' are both probably fine.

Licensing is a mess - but no license is a way bigger mess.

GitHub repo for project

```
├── method
│   ├── qualtrics implementation
│   ├── self-report scales
│   └── self-report scales.docx
├── code
│   ├── analysis.Rmd
│   ├── processing.Rmd
│   ├── models
│   │   └── model_fit.rds
│   └── plots
│       ├── figure_1.pdf
│       └── figure_2.pdf
├── data
│   ├── raw
│   │   ├── data_raw.csv
│   │   └── data_raw_codebook.xlsx
│   ├── processed
│   │   ├── data_processed.csv
│   │   └── data_processed_codebook.xlsx
│   └── results
│       ├── results_summary_statistics.csv
│       └── results_regression_table.csv
├── communication
│   ├── preprint
│   │   └── hussey et al. - 2024 - title.docx
│   └── presentations
│       └── hussey et al. - 2024 - conference presentation.pptx
├── README.md
└── LICENSE
```

See the
'example of a well-organised github repo'
directory for an example

Practice