PSY6422 Data Management and Visualisation @ The University of Sheffield

Tom Stafford

2020-04-21

Contents

Chapter 1

Module Overview

PSY6422 Data Management and Visualisation is part of the MSc in Psychological Research Methods with Data Science taught at The University of Sheffield by Tom Stafford

Note this is a placeholder page. In 2020 most of this material was delivered offline. I am adding notes online as I can, so these pages in particular may evolve quickly

1.1 Motivation

You are aiming to produce reproducible workflows - scripts that automate all steps between raw data and the final data in your papers.

Point and click solutions (spreedsheets, SPSS) are inadequate.

As well as being *reproducible* (by you or other researchers) your work should be *legible* (to Future You, or other researchers) and *scalable* (it should work as well on 400,000 data points as on 40).

You will need help to do this. Therefore you will use Open Source solutions - these are analysis products which have a worldwide community of people using them, and the infrastructure which supports sharing advice and solutions.

In practice, this means you are going to start by using R (you could use Python, but this module is based on R)

1.2 Resources for current students

Google Drive (UoS login required to access):

Includes slides and workbooks, as well as these specific documents

- Timetable
- Useful information
- Assessment Criteria

And of course these pages (hosted on github, no login required)

You may particularly enjoy the Reading list

1.3 Course Outline

In 2020 we are covering a compressed curriculum. The topics we cover are:

1. Module overview - ethics and aesthetics of visualisaion, the importance of reproducible workflows
2. Project organisation - fundamentals of data storage, syncronisation and sharing
3. R / Rstudio - introduction to the statistical computing language
4. Making graphs
5. Data Management
6. Data Management 2
7. Coding principles - fundamental principles of coding, writing good code, asking for help
8. Rmarkdown
9. Git and github
10. Publishing to pages - like this one!
11. Advanced topics

Stretch goals:

Unfortunately we won't have time this year for a number of advanced topics which I would like to cover. Hopefully next year:

- Jupyter notebooks
- The terminal / ssh
- Interactive visualisation with Shiny apps
- SQL

1.4 Example Projects

The bulk of the assessment is to conduct and publish your own analysis project. Here is an example small project which gives an idea of what I mean

- SuperTues: Published, repo

Stretch goal is to build an interactive data visualisation using Shiny

- Here's one I built earlier Power analyser

1.5 More

- Russ Poldrack's Advice for learning to code from scratch