PRMDS Project (PSY6009) requirements

Version: 2021-08-03, Summer 2021. Feedback welcome

This is the final MSc project for students on the PRMDS (“Data science”). These projects are marked using the same criteria as other MSc projects, but, analogously to the Advanced Stats stream project, must meet additional criteria reflecting the training provided on the Data Science stream. The checklist below is designed to provide clarity for students and supervisors that they have met these requirements.

Checklist

Project data and materials archived in a repository, linked to from the project write up.

If the repository is private (e.g. a google drive folder) a statement is included as to why sharing publicly is not possible (e.g. commercial, confidentiality or non-consent constraints)

Project data and materials to be accompanied by documentation explaining what is archived, providing details to aid their use/reuse. (ideally full documentation should be in a properly archived repository, which is linked to from the methods section)

Materials labelled so as their relation the project is clear

Data should be accompanied by a codebook describing data provenance, format, variable names, descriptions and possible values. (for details on this consult the relevant materials taught on PSY6422).

Analysis produced by script (e.g. in R) which is commented and runs to produce all project outputs (i.e graphs and statistical analyses) with no or minimal need for human intervention (e.g. no steps between raw data and final analysis which require hand coding, etc)

Any non-scripted stages are fully documented (in the archived repository)

Analysis script shared in a repository

All supplementary materials clearly organised in a systematic manner (e.g. using subfolders for different components)

Motivation

Our aim is to produce research which is open, reproducible and scalable. This means sharing as early as possible the methods, materials and outputs using freely accessible tools and platforms. “Reproducible” means curating outputs so that they are well documented as well as open, and archived in a form which allows ourselves or others to audit what we have done, and ideally, do it themselves. For analysis purposes this means sharing data processing scripts which generate our outputs (statistical analyses, graphs). “Scalable” means that, as far as possible, we design research processes which can be adjusted to new and/or larger projects (e.g. by formalising/automating outlier identification, rather than doing it “by eye” and editing an excel sheet).

This aim is increasingly formally recognised in funder and journal requirements. For example, many journals (e.g. Psychological Science) now offer badges to indicate articles which have been published with open data, open materials or preregistration, as well as requiring a statement in the manuscript about open research practice (e.g.).

These requirements for Data Science students are therefore designed to be no more additional work than best practice for submission to a leading academic journal, as well as supporting the long-term value of the project work, allowing it to be reproduced, built on or otherwise integrated into the supervising academics ongoing work.

Supervisors are not required to have any special knowledge of data science. Supervisors are not required to audit the analysis code. Students will be taught all they need to meet these requirements outside of project supervision.

Notes

Project analysis does not *have* to be in R, but that is what PRMDS students are taught. Any programmatic, reproducible, method which produces the analysis is acceptable.

PRMDS students take the advanced stats module but are not *required* to use an advanced statistical method in their project (although they are allowed/encouraged).

Note that PRMDS MSc students may do projects on any topics, using any methods, just as PRM and PRMAS students. They may be particularly strong on projects which involve data preparation, exploration and visualisation without involving primary data collection (“data projects)”. The project *requirements* for PRMDS students (above) are not the same as the additional *guidelines* for writing up data projects; for these see:

> Guidance for Data Projects

More information on the specific topics taught DS students can be found at these PSY6422 course pages:

> https://tomstafford.github.io/psy6422/

PSY6009 Research Project in Psychological Research Methods with Data Science (60 credits). From the module E1 form:

The module allows students to conduct, analyse, and write up a research project under the guidance of an academic supervisor. Students gain first-hand practical experience of managing the research process, from the formulation of a specific research question on the basis of a review of relevant literature and guidance from the supervisor, to the design, execution, analysis, and report of a study. Projects typically involve the management, analysis, and visualisation of scientific datasets, through the use of current techniques in data science. Projects are written up in the standard format for submission to an appropriate academic journal.