Working with data types and structures in Python and R_B208593

Administrative Information

- CMVM - Usher Institute

Dr Mairead Bermingham, mairead.bermingham@ed.ac.uk

2022-06-06

2022-07-04

Data Collection

The data will be collected from NHS England accident and emergency attendances and admissions records from large type 1 emergency departments. The raw quantitative data will be stored in CSV format in NHS databases.

The inital raw data, the processed data, scripts (RScripts, Python scripts) and graphs will be stored in clearly labelled folders to allow replicability of the data process.

Documentation & Metadata

Dataset will be analysed using RStudio and Python notebook. Raw files will be processed in RStudio and Jupyter notebook wil be used to create a data capture tool. The RStudio will used to create a data dictionary for the processed data.

Ethics & Legal Compliance

We must be adhered to GDPR and obtain consent from the end-user to process and share the data collected with the data capture tool.

Storage and Back-Up

All information (dataset raw files, R scripts, Jupyter notebook scripts, data dictionary) and data will be stored in secure storage location such as DataStore.

Selection and Preservation

To ensure tha long-term preservation and storage of your data it is best to deposit it in a recognised research data repository. This could be either DataShare or DataVault (https://www.ed.ac.uk/information-services/research-support/research-data-service/after) at the UoE or a specialist disciplinary repository such as NERC datasentres, NCBI or EBI.

Processed data, R scripts, Jupyter notebook scripts, data dictionaries

Data Sharing

- Yes: go to 12

The anonymized data can be accessed via open data repository such as DataShare.

Responsibilities & Resources

I will be responsible for collecting or generating the raw, for the analysis of the data and preparation of the data for archiving at the end of your project only.

I will require further training as I am a first year MSc student in Data Science.