School of Geography
University of Leeds
GEOG5995: Programming for Social Science: Core Skills

2023/24
Level 5 / 15 Credits
Fran Pontin

Final Project

1. **The assignment**

   The aim of this assignment is to give you the opportunity to showcase the skills you have learnt throughout this module. You should complete this assessment independently utilising the programming, mapping and data skills acquired through the practical elements of the course and through your own reading and independent work.

2. **This assignment contributes 70% to the final mark of this module.**

3. **Individual work and academic integrity**

   This is an individual assignment. Whilst it is natural to consult with your course mates, or to help each other with unfamiliar software or tasks, it is not permitted to share research, ideas, data or text that form any part of this report/essay.

   Students must also be careful to avoid plagiarism. Plagiarism means presenting the ideas of others as if they are your own. Examples of plagiarism include failing to cite the sources of ideas and information that are not your own, and copying sections of text from sources that you have read. Collusion and plagiarism are both serious types of academic malpractice. If we suspect that these or other breaches of academic integrity have occurred then they will be investigated according to university protocols.

   For LUU information and advice on 'Plagiarism, Fraudulent or Fabricated Coursework and Malpractice') see: https://leedsuniversityunion.knowledgeowl.com/help/plagiarism

   For Skills@Library information on academic integrity and how to avoid plagiarism see: https://library.leeds.ac.uk/info/1401/academic_skills/46/academic_integrity_and_plagiarism

4. **Submission of assignment**

   The deadline for submission of this assignment is 14:00 (2 pm) on 15th December 2023.

5. **Word limit**

   The word limit for this assignment is 2500-word **equivalent.**

6. **Marking criteria**

   The School's standard coding marking criteria apply. See:
   http://www.geog.leeds.ac.uk/courses/other/programming/info/MarkingScheme.docx

7. **Feedback**

   Feedback on your coursework will be returned to you 15 working days after the submission date. In cases where this is not possible we will inform you.

8. **Further details**

Please refer to the module handbook for full details, including referencing, submission, late penalties, and word count policy.

Final Project

Youshould create a Jupyter notebook documenting the investigation and interpretation of open data to gain insight which can be used for public good. The project should use openly available data (ensure these are attributed properly) from at least two sources. These data should be suitably joined to provide insight and two polished data visualisations produced to communicate the results of the analysis.

Examples of potential projects include:

- Do health promoting neighbourhoods have lower neighbourhood churn in Sheffield?
- Are neighbourhoods with better Fast-Food Access at greater Food Insecurity risk in Manchester?
- Are there more potholes reported in more deprived areas in Glasgow?
- Is there an association between household age and price in Leeds?
- Are supermarkets less likely to deliver to areas with a high student population in Liverpool?
- Is public transport reliance associated with covid-19 rates in London?

The Jupyter notebook should follow the data science process outlined in the module. You should use markdown cells to explain and justify the steps taken in the code. Whilst you can use visualisations throughout for data exploration, please present two final visualisations to communicate your 'public good' insight, one spatial and one non-spatial. For both these visualisations indicate your intended audience (e.g. academic, the public, policymakers, business stakeholders) and the subsequent data visualisation choices that have been made to ensure these outputs are suitable for this audience.

You will be assessed on the following:

- Readability of the Jupyter notebook including markdown cells and commented code **(5%)**
- Logical Flow of the Jupyter notebook **(5%)**
- Data cleaning **(30%)**, including:
    - Exploratory data description (with citations where appropriate)
    - Justification of data cleaning choices (with citations where appropriate)
    - Data visualisation to aid data cleaning
- Statistical modelling of the association between your chosen variables (**10%**)
    - Effective interpretation of the model
    - Justification of model chosen
- Data Visualisation of end results: **(30%)**
    - This should include one spatial and one non-spatial visualisation

- o Evidence of modifying visualisations to make them:
  - Accessible (colour blind audience etc.)
  - Easy to interpret
  - Readable (e.g. change variable names from how they have been presented in original table to improve understanding)
  - Explanation of data visualisation choices (with citations where appropriate)
  - Evidence of critical selection of chosen visualisations to present (with citations where appropriate)
- Reproducibility **(20%)**
  - o You should ensure all the data and steps to produce the analysis undertaken is reproducible
  - o You should create a GitHub repository with suitable documentation This should include:
    - an introductory markdown file explaining:
      - The background/context of the project
      - What data the GitHub repository contains
      - What the code aims to do
      - Any further information someone might need to be able to run your code and reproduce your analysis
    - All files needed to run the analysis
    - Your .ipynb notebook

The Jupyter notebook will be marked against the SoG generic marking criteria and specifically the following additional criteria:

- Understanding of task
- Structure and flow of the report
- Breadth of content
- Reading and referencing
- Aesthetics and design
- Writing / presentation
- Data

The Jupyter notebook markdown word count is limited to 1500 words.

It is up to you to source data for your final project and you should remember to fully cite the sources of all data used. You will find some ideas for datasets throughout the course materials. Websites such as the Consumer Data Research Centre (https://www.cdrc.ac.uk/) , Data Mill North (https://datamillnorth.org/) and Kaggle (www.kaggle.com/datasets) have lots of geospatial datasets.

Remember you may be penalised for missing files/images, so carefully check your zipped file contains all the necessary files before you submit.

Submission procedure:

All coursework must be submitted electronically via Minerva. Each module and each assessment has an electronic submission area which can be located by following the 'submit my work' link on the left hand side of the module homepage.

Please upload the Jupyter Notebook and all files need to run the notebook analysis as a zip file using the 'Zip file upload' option of Turnitin. Contained within the zip folder please also submit a .txt file with a link to your GitHub repository
e.g. https://github.com/MyGithubUserName/GEOG5995_Final_Project

When submitting your project, ensure you submit ALL relevant files (csv, xlsx, shp, geojson, files and additional data files). Marks will be deducted if your work is incomplete (missing shp files etc.), you can also include links to URLs in a separate .txt file if you wish. Place all files into a single zipped file labelled with your Student ID number and upload this using the 'zip file upload' feature of Turnitin. If you have problems submitting your files, e-mail them to Fran Pontin (F.L.Pontin@leeds.ac.uk) BEFORE the deadline.