STAT40800 Data Programming with Python (online)

Data Analysis Project

Autumn 2020/21

On Brightspace, under *Learning Materials → Project*, you will find a number of datasets related to the COVID-19 pandemic in Ireland. Your task is to formulate a research question based on this data and use the data to answer your question. You are not expected to use all of the datasets, just a sample. For example, you could choose to compare the cases, deaths and hospitalisations in different countries/regions, or perhaps focus on testing or the socio-economic impacts of the pandemic.

If you wish to use additional data, send me an email and I will decide if it is appropriate or not.

There is no strict report structure, but I suggest including the following sections:

1. **Abstract:** Outline the research question and summarise the contents of the report.

2. **Introduction:** Introduce the problem and review some background information.

3. **Analysis:** Describe your approach to answering the research question and present the results of your analysis.

4. **Conclusion:** Surmise your findings, interpret your results and suggest future lines of enquiry.

A rubric is also included under *Learning Materials → Project*. This will help you assess whether your project is of a high enough standard or not.

I strongly recommend using Jupyter so that text, code and figures can be interleaved. If using Jupyter, you **must** submit your report as both a PDF and *.ipynb*. If you are using a different Python IDE, you should use a text editor, such as Word or LaTeX, to write your report and submit a PDF of the report. As well as a PDF of the report, you **must** submit your Python code.

The final report is due at the end of Week 12 on Friday **December 11**.