

Data Science for Health and Biomedical Sciences  
Programming Assessment brief

In the Data Science assessment, you will be looking at the Public Health Scotland prescribing data, available here: <https://www.opendata.nhs.scot/dataset/prescriptions-in-the-community>

We will look at this data set in the Week 7 labs, and so you will be broadly familiar with it.

Your task is to explore the data further around a research question that you yourself set.

Here are some examples to inspire you:

- Trends in the prescribing of antihistamines across the four seasons – for instance, is there a spike in the hay fever season?
- Trends in the prescribing of dementia medication – for instance, can we see more prescribing in areas with an older demographic?
- Trends in the prescribing of methadone – for instance, can we see more prescribing in more deprived areas?

Here are examples of actual submission titles from previous years:

- How do local environmental factors affect prescription of short-acting bronchodilators by GP practices in Scotland?
- How do SSRI antidepressant medications vary across socioeconomically diverse regions in Scotland?
- Impact of the COVID-19 pandemic on antidepressant prescription patterns in Scotland
- Did COVID-19 pandemic influence changes in inhaler prescription and potential respiratory illness trends?
- What is the trend of prescribing cold and flu medicines during Fresher's Flu season (September-October) in various university areas across Scotland?
- The relationship between antihypertensive drug prescription trends, seasonality, COVID-19 vaccination, and stroke incidence in Scotland.

In working out the answer to your question, you may want to join the data set with other open data – for example, the register of all GP practices (available on the Open Data website), or the census. You may also explore different geographical units (health board is the obvious one, but you could also look at council areas or postcode).

A key part of your submission will be a Github page. The content of this page should not be longer than 10 pages, when printed. Please check the page numbers of your Github Page by following these steps:

1. In your browser go to "File -> Print"
2. In the Destination drop-down select "Save as PDF".

3. Check the number of pages displayed in that window.

Your report should include no more than 4 figures, out of which at least one should be a table, and one should be a plot. A faceted plot counts as one plot.

You may want to include a few references to justify some of your decisions (e.g. to contextualise your research question or support a particular way to deal with outliers). If you do, you can use any referencing style you like, as long as it is consistent. And don't forget to include a reference list at the end.

NOTE: I would not expect more than 3-5 references in a report.

You will need to submit a document via Turnitin on Learn. Your submission will be a PDF or Word document containing two links:

1. A link to your Github repo.
2. A link to the associated Github Page.

Your Github repo must include your original R Markdown file. We will clone your repo and run your file to check reproducibility.

Don't put your data on Github. If you can link directly to a data file (for example prescribing data), provide that link in your code or text in your markdown file. If you cannot link directly to a data file (e.g census data), describe where we can download it (provide census table code, etc.)

NOTE: As part of your submission, please include a note of if/how you used generative AI (ChatGPT etc.) to support you.

You will be assessed in five general areas:

1. Data Storytelling (are you using the data to tell a coherent and interesting story?)
2. Data wrangling (have you wrangled your data effectively, with a clear documentation of the process?)
3. Plots and tables (have you produced effective code for tables and plots, going beyond the defaults?)
4. Reporting (have you used some of the advanced features of R Markdown? Is your report well-structured and easy to follow?)
5. Reproducibility (is your code easy for a human to read and follow? Have you included comments where appropriate? Are we able to get hold of the data you have used?)

For the full marking criteria, see [here](#)