# PMIM102J Scientific Computing and Healthcare

**Assignment – January 2022**

## Assessment – Project

You are given access to a database containing the following information (note that this is all real, publicly available data—see information and documentation at the links below):

- **General Practice Prescribing Data Extract for Wales**
  A database of monthly information on what medications GP practices in Wales prescribe. The database includes all prescribing results up to 2015, combined into tables. The GP practice lookup data is just a single table from 2017.
  http://www.primarycareservices.wales.nhs.uk/general-practice-prescribing-data-extrac).

- **QOF Results for Wales, 2015**
  A database of how each GP practice performed according to the Quality and Outcomes framework, a method of assessing GP performance against targets, which affects payment. The results includes counts of how many people at each practice were diagnosed with certain diseases, as well as some performance measures.

  This is a subset of the QOF results, including just two tables: the performance of practices, and a lookup table of what all the indicators are.
  https://www.gpcontract.co.uk/browse/WAL/15

Design and implement an R program (using the attached style guide – see notes below) that reports information from this data:

### Part 1: Specific Questions/Tasks (50%)

1. Allow the user to select a practice. For that practice, report:
    - Whether the practice has medication information available
    - Whether the practice has QOF data available
    - If there is both medication and QOF information available for the practice, show:
        - The number of patients at the practice
        - The average spend per month on medication
        - Create a visualisation showing the spend on medication per patient, compared to other practices within the postcode area (i.e. first part of postcode).
        - Report the rate of diabetes at the practice, and visualize how this compares to other practices in Wales.
2. Finally, perform an all-Wales analysis comparing the rate of diabetes and the rate of insulin prescribing at a practice level. Is there a statistically significant relationship between the two?

Repeat using metformin instead of insulin. If you find statistically significant relationships, which is stronger?

## Part 2: Open-ended analysis (50%)

- Perform some analysis of your choice, of something that might be of interest to health organizations using the data.
- Report numerical results and visualisation.
- It may be customised based on user input if you choose.
- This part carries 50% of the marks. As a rough guide, this section should take at least as long, be as least as complex or represent at least as much work or effort as Part 1.

## Design requirements/Expectations

1. The R script should be easy to use for a non-technical user (the user should not have to modify code or step through individual lines of code to make it work).
2. Use the attached style-guide to help format your code. This is a **guide** not a set of rules, but if you are going to use a different style in your code, please make a note in a comment in the code or modify and include the style-guide in your submission.
3. The database on the test system will be exactly the same as given to students, and the naming and method of access will be the same as shown in the installation instructions/course lab work.
4. It's fine to use example code given in class (e.g. code that connects to the database) as part of your submission. It's fine to use publicly available R packages, as well. This is good practice. (It's not fine to actually copy other code into your work though, beyond what was given in class).
5. Modifying data in the database is not necessary, but permissible if you find it helpful. Any modified data should be in separate tables to avoid altering the original data. The assignment that you submit would need to create these tables, since they won't exist on the test system.
6. For the purposes of this project, each plot can simply be output to the screen, and any numerical or text results can be printed to the console.
7. Think about the software development process covered in the course…especially testing!
8. Working with imperfect, real world data is part of the assignment. You may have to do some research to learn what the data contains, as well as explore the data and test your assumptions.
9. Working from a somewhat ambiguous description of the task is part of the assignment. If a question is vague, the first step is to define a clear specification of the problem.
10. Did I mention you should test? ☺

## Deliverables:

1. **User Documentation –** A brief document (1-2 pages is sufficient) written for a hypothetical user, who doesn't know about this project and doesn't know how to write code. The documentation should cover: what the software does, how to install and use it, and the meaning of the outputs.

2. **R code –** An **RStudio project** containing one or more R files, that accepts user input and generates results as described above. The script should meet the requirements for part 1; for part 2 it should generate at least one visualisation and two numerical results from the data.

## Assignment requirements/submission

- Submit the assignment as a zip file on Turnitin.
- A standard assignment cover sheet is required—include it either as part of the user documentation or a separate file.

### Assessment Criteria

| Criterion | Value | Description |
|---|---|---|
| Design | 30% | What does the script do? Is that useful and well thought out? Is the method of obtaining user input logical? Is the output clear? |
| Implementation | 40% | Does the script work correctly? Is it robust? (For example, how well does it handle unexpected input?) Does the code implement the design in a logical, efficient way? |
| Code Style | 15% | Is the code neat, well organized, and easy to understand? Is the code sufficiently documented so that someone else could work on it and make changes? |
| User Documentation | 15% | Is the user documentation sufficient for a hypothetical user of the script to know what it does and how to use it? |