Assignment for PMIM102J

Swansea University
Prifysgol Abertawe
Medical School
Ysgol Feddygaeth

Module number:	PMIM102J
Module name:	Scientific Computing and Health Care
Title of assignment:	*Database Project January*
Student ID number:	*445348*
Word count:	
Declaration:	I understand the following conditions which apply throughout this course: 1. I confirm that I am the sole author of this work. 2. I understand that proof reading by a third party is discouraged, but if used, records should be available as per guidelines. 3. I understand the need for academic integrity and that all my submitted work will adhere to its principles. 4. I understand that the teaching team will take measures to deter, detect and report any academic misconduct. 5. I agree to my work being submitted to the TurnItIn academic database. 6. I understand the importance of assignment deadlines and the need to seek help in good time where personal circumstances interrupt my work.
Please copy and paste this declaration onto the front of the submission.	

PMIM102J: Scientific Computing and Health Care – Database Project
Student ID: 445348

Introduction:

This guide aims to give you a general overview of that R script presented to you. The program will consist of three r scrips, Main.R, Logic.R and DBS_Connection.R.

The scripts designed to take full advantage of the R Studio IDE and require a PostgreSQL connection with PGAdmin. If these are currently not on your machine, please download them here:

R Studio: Click here
Postgres: Click here

This documentation takes an assumption that you have access to the gp_data_up_to_2015 available for your required Postgres database version and that the user knows the required port numbers for their database connection (the port number needs to be updated in the script before running).

Code Overview

This program can be run either directly from the source run or by executing the script line by line within the IDE. However, depending on what method you use will depend on when the plots and tables will show in the plot and view areas. The console prints outs will happen as they are happening within the script sequence. Note, the general table exploring at the beginning of the script is better when the user runs the code manually as they are then also able to enter in tables that they would like. Pre-populated details have been added for instances where the source run method gets selected.

Additional packages required are:

- RPostgreSQL
- GetoptLong
- Magrittr
- Dplyr
- Lubridate
- Gt
- Glue
- Plotly

If these packages are not installed on your machine, then please go to the Logic.R file and the DBS_Connection.R file and uncomment out (remove the #) in front of the required install.packages(X).

The Main.R file is where the user will need to be in order to execute the scripts. This main file will then pull in the required other scripts. The Logic file is where the application's main logic will be carried out, and the DBS_Connection file is where the necessary functions are to handle the database connection.

All outputs will appear in either the console window or within the R Studio plots or view tabs.

Part 1: Set Analysis

In part 1, precise tasks have been set out. These require the user first to enter a practice id into the console window. If the user inputs an incorrect practice ID, they will be asked to repeat the process until they do. This console input is not case sensitive. The selected GP practice will then return to the main script and allow the program to find its region of the located GP practice. This region information is due to additional commands within the file requiring this information.

The practice ID select is the only time that the user will need to interact with the system. The rest of the actions have all the required information they need. However, if a suitable practice id is selected but has missing records, the user will need to choose another practice id. But if the file has been executed with source run, the program will keep running, meaning the user will have to terminate the running of the file or wait till all the required actions have finished.

Outputs

The first output is a combination of a console output of the top 5 drugs and a table output in the viewer. Both the console window and view table display the same information, just in whatever method the user feels more comfortable using. The second output is the GP's cancer patient details. This output shows the number of patients diagnosed with cancer with the indicator CAN001 and the GP's percentage in the console window. Additionally, the viewer will display a table with the details from the database. The third output displays similar information about cancer diagnosis but this time about the GP, region and Wales. However, this output will have a graph showing the % of the three different areas (GP, region, Wales). The fourth output displays the practice's total spend on drugs and the cost per patient in the console. If Inf appears, this sign shows that the practice has information in the GP data table but no matching details within the QoF Achievement table. Also, a scatter plot gets displayed with all the values. The fifth and final output is the GP correlation check against cancer, dementia, hypertension and diabetes. The results get revealed in a plot.

Part 2: Open Analysis

Part 2 follows a similar process to part 1. However, this section focuses on the region as a whole. The user will select a region from the database, and if the user selects an area that doesn't exist, they will get asked to re-enter the region. This input is also not cased sensitive, allowing the program to handle any style of input from the user as long as that input matches a value within the database.

Again the selected region is the only time that the user will need to interact with the system. The rest of the actions have all the required information they need. However, if a suitable region id is selected but has missing records in some of its practices, the user will need to choose another region. But if the file has been executed with source run, the program will keep running, meaning the user will have to terminate the running of the file or wait till all the required actions have finished.

Outputs

The first output is a combination of a console output of the top 5 drugs for the selected region and a table output in the viewer. Both the console window and view table display the same information again, allowing them to choose what method they would like to use. The second output is the region's smoking patient details. This output shows the number of patients diagnosed with suitable smoking screening with the indicator SMK SCR, along with the region's percentage in the console window. Additionally, the viewer will display a table with the details from the database. The third output displays similar information about the smoking diagnosis but also with Wales too. However, this output will have a graph showing the % of the two different areas (region, Wales) and a tooltip. The fourth output displays the region's total spend on drugs and the cost per patient in the console. If Inf appears, this sign shows that the selected region has information in the GP data table but no matching details within the QoF Achievement table. Also, a scatter plot gets displayed with all the values. The fifth and final output is the region correlation check against smoking, dementia, hypertension and coronary heart disease. The results get revealed in a plot.