Programming Assignment

Deadline: submit assignment by Monday August 10th 2020, 11pm BST

Background

Unintentional injury is one of the leading causes of death in children and young adults. The elderly are also more vulnerable, and it is a common cause of emergency hospital admission in people of all ages. Many injuries may be preventable, and through the collection and analysis of data, patterns can be explored, and evidence gathered, helping to inform policy makers and the public so that useful evidence-based prevention strategies can be developed.

In this project you will be exploring the **Unintentional Injuries** datasets (admissions and deaths) from the Public Health Scotland open data platform. The datasets include information on admissions and deaths relating to injuries and assaults in Scotland.

Data

Three data files are provided for your analysis:

- `ui_admissions_2020.csv`
- `ui_deaths_2020.csv`
- `hb_lookup.csv`

These are in CSV format and can be downloaded from the following link:

- https://github.com/lisanorman7/hds_assignment

Here is a link with more information on the variables in the dataset:

- https://www.opendata.nhs.scot/dataset/unintentional-injuries

A few points to note about the dataset:

- Many of the rows contain aggregated data, due to each variable including an "All" category
- The variable "HBR" refers to Health Board Region codes which is the variable to match if joining with the Health Boards look-up.

Assignment

The purpose of this assignment is to explore the injures datasets and answer the questions below. To answer the questions you must use the datasets provided. The R code used in all aspects of loading, processing, analysis and reporting must be visible in your final submission. You are encouraged to include tables and figures to illustrate your report. You can use any R packages of your choosing to create this report; you are not restricted to those discussed in the course.

To help you get started we have included an example of a basic script which shows code for reading in the data plus some basic cleaning, plotting, and a table code too. Please feel free to write your own code if you prefer, or to change and build on the provided starting point. All files related to the assignment, including the R Markdown file (`assignment_helper.Rmd`) and the example output from this file (`assignment_helper.pdf`) can be accesed here.

Alternatively, right click on the link below and click **Save link as...** to download the example R Markdown file directly to your own computer:

- assignment_helper.Rmd

Report Questions

In your report, you must explore the following questions:

- What types of injures are most common and for which demographic (the data include age and sex)?
- What is the rate of death in admissions of these injuries?

Your report should be written as if it were to be read by government advisor who might be responsible for producing policies or guidance to help cut down on the number of injuries. You only need to provide the evidence, you do not need to include recommendations or guidance on this.

We are not looking for one specific analysis for 1. and 2. As you're working with the data, feel free to make decisions on whether you want to focus on a specific demographic or location(s) if that makes your message clearer and analysis more sound.

Requirements

To carry out this assignment, you will need:

- access to RStudio – this is where you will carry out your analysis and produce your report in R Markdown.
- to make sure the following packages are installed:
    - **rmarkdown** and **knitr** – for compiling the R Markdown document
    - **tidyverse** – for access to data wrangling and plotting functions you have learned in this course
    - Optional: **janitor** for cleaning up the variable names as they're read in from the CSV file.

If you have followed the course material you should already have most of these set up and ready to go.

Document Layout

You are required to submit a single Word or PDF document as your report. This document must be the direct result of "Knit to Word" or "Knit to PDF" from R Markdown in RStudio. You must not apply any edits or modifications to your report outwith RStudio.

Your report must include the following sections:

- **Title**: should be informative and about the analysis in the report
- **Overview**: located directly below the title, should include a brief summary of your analysis
- **Data Processing**: should include descriptive text and code to show how the dataset was loaded into R and processed. All processing must happen inside the document, starting with the raw CSV file. You must not edit the CSV directly/using Excel or similar.
- **Results**: this is where you present your findings

Other sections are allowed but not required.

In addition, you must ensure the following:

- **Figures**: there should be at least one plot and one table in your report and a maximum of 3 figures (one figure is enough and could include multiple plots)
- **Code**: All R code must be shown (`echo = TRUE` ensures this and is set as default)

When you come to writing reports on your own data, you will often choose not to show your code but just the final text and figures. For the purpose of this assessment all code is required, and it needs to appear in your Knitted report, not submitted separately.

Submission

One week before the submission deadline, a link to Atlas Pebblepad and further submission instructions will appear here:

- [Link to upload your Word or PDF report to Atlas Pebblepad]
- [Link to further submission instructions]

Help

If you have any issues or questions relating to the assignment, please post on the **Questions about Assessments** discussion board on the course website in Learn.

Marking Criteria

[25%] **Reporting**: Demonstrates features of R Markdown, e.g., YAML, Markdown, chunk options, adheres to structure, figure guidelines, answering of questions

[25%] **Demonstrating reproducibility**: Report rendering, commenting, code readability, object naming, use of raw data

[20%] **Intellectual contribution**: Demonstrates understanding of R and R Markdown, uses initiative through intellectual contribution and added value

[15%] **Wrangling**: Produces effective code to deal with data wrangling and tidying, e.g., dplyr, tidyr, clarity of process documentation

[15%] **Plotting**: Produces effective code to create plots, e.g., ggplot2, clarity of presented information