STAT40620
Data Programming with R

Assignment 1

Instructions

- This assignment is due on Wednesday 10th October 2018 at 11:59pm.

- You should submit your assignment to the 'Assignment 1' assignment object in Blackboard.

- You should submit two files only:

  1. `Rmd` file or a single `R` script file detailing the commented code you used to obtain your answers.

  2. final document in either `pdf` or `Word` which should contain answers to the questions below.

     - If you created an HTML file, please convert it to pdf. You can use Google Chrome: `File > Print > Destination [Change...] >` select `Save as PDF`.

- To get one bonus mark you must submit the `Rmd` file and the resulting document which shows all your code.

- The assignment grades are capped at a maximum of 10 marks, bonus included.

- The marks available for each question are shown in brackets.

- Assignment 1 is broken up into 3 tasks: data manipulation, analysis, and creativity.

- You may have to learn and discover some new functions. Use `help()` and `help.search()` to find what you need.

Task 1: Manipulation

The dataset `EurostatCrime2015.csv` records offences (values per hundred thousand inhabitants) by offence category in 41 European Countries in 2015.

1. Load the dataset `EurostatCrime2015.csv`. Notice that the first column of the `csv` file contains the names of the countries that must be read as row names [Hint: Load in the file using the function `read.csv`]. [0.5]

2. What is the size and the structure of this dataset? [0.3]

3. Produce appropriate commands to answer the following questions:

   (i) Add a new column called `Sex.crime` which contains the sum of all the crimes that have a sexual component: `Rape`, `Sexual.assault` and `Sexual.violence` [1]

   (ii) remove the columns `Rape`, `Sexual.assault` and `Sexual.violence`. [1]

4. Work with the dataset you created in question (3ii), and list the countries that contain any missing data. [1]

5. Remove the countries with missing data from the dataframe. [1]

6. What is the size of this new dataframe? [0.2]

Task 2: Analysis

Produce appropriate commands to answer the following questions:

1. According to these data what was the most common crime in Ireland in 2015? [1]

2. And the 3 least common crimes in Ireland in 2015? [1]

3. Which country have the highest record of offences (per hundred thousand inhabitants)? [1]

Task 3: Creativity

Do something interesting with these data! Create a table or a plot which shows something we have not discovered above already and outline your findings. [2]