

STAT40730

Data Programming with R

Assignment 1

Instructions

- This assignment is due on Wednesday 9th October 2019 at 11:59pm.
- You should submit it to the ‘Assignment 1’ assignment object in Brightspace.
- You should submit two files only:
 1. Rmd 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.**
- You may submit it multiple times before the deadline, but only the last version will be marked.
- There is a maximum of 10 marks for this assignment. This assignment is worth 10% of your final grade.
- The marks available for each question are shown in brackets.
- Late submissions will score 0, unless a “Late Submission of Coursework” form is submitted.
- Assignment 1 is broken up into 3 tasks: data manipulation, analysis, and creativity.
- You may have to discover and learn some new functions. Use `help()` and `help.search()` to find what you need.
- Some tips on using R Markdown are given at the end of this document.

The dataset `EurostatCrime2017.csv` records offences (values per hundred thousand inhabitants) by offence category in 41 European Countries in 2017. Full information on the dataset is available here: https://ec.europa.eu/eurostat/cache/metadata/en/crim_off_cat_esms.htm.

- Complete your assignment using R Markdown, check that all the output and code are correctly shown in your final document. Knit your document frequently to fix errors. Once completed, submit the `Rmd` file and the resulting `pdf` or `word` document which shows all your code. [1]

Task 1: Manipulation

1. Load the dataset `EurostatCrime2017.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 (number of rows and columns) and the structure of this dataset? [0.3]
3. Produce appropriate commands to answer the following questions:
 - (i) For some countries `Theft` includes also burglary, and theft of motorised land vehicle, in others they are recorded separately. Add a new column called `All_Theft` which contains the sum of all the crimes that have a theft component:
 - `Theft`,
 - `Theft_of_a_motorized_land_vehicle`,
 - `Burglary`,
 - `Burglary_of_private_residential_premises`Please consider `NA` values as 0 in this case. [Hint: you may want to use the function `apply` or `rowSums`, check their help files to see how to deal with missing values.] [1]
 - (ii) Remove the columns: `Theft`, `Theft_of_a_motorized_land_vehicle`, `Burglary`, and `Burglary_of_private_residential_premises`. [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. How many observations and variables are in this new dataframe? [0.2]

Task 2: Analysis

Work with the dataset produced in question 6. Produce appropriate commands to answer the following questions:

1. According to these data what were the 3 most common crimes in Ireland in 2017? [1]
2. Which country has the highest overall 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]

Tips for R Markdown

- Be aware that a common error is to give the same label to two different code chunks!

```
```{r cars}
summary(cars)
```
```

```
```{r cars}
plot(cars)
```
```

You can fix this by changing the label to one of them:

```
```{r cars2}
plot(cars)
```
```

- If you want to improve the appearance of your plot in your knitted document you can set up the dimension of your figure:

```
```{r, fig.height = 10, fig.width = 7, fig.align = "center"}
plot(Nile)
```
```

- In case of an error in your code, add the option `error = TRUE` into the R chunk to run the code, show the error message on the knitted file. For example:

```
```{r, error = TRUE}
x <- "a"
sum(a)
```
```

- For all the available options for the R chunk, you can see here: <https://yihui.name/knitr/options/>
- R Markdown website: <https://rmarkdown.rstudio.com/>
- R Markdown cheatsheet is available here: <https://www.rstudio.com/resources/cheatsheets/#rmarkdown>