DATA SCIENTIST 22

Inlämning
(Deadline: 14 Oct 2022)

1. Basic algorithmic thinking in R:
Get input temperature of 10 cities for a month from the user.
Create 3 functions to compute the minimum, maximum and average temperature of each city in that month:

a) By using manual calculations
b) By using the in-built functions in R

Be sure to output and show the results of these functions.

2. Data Pre-Processing:
Download the chronic kidney disease dataset from kaggle: https://www.kaggle.com/datasets/mansoordaku/ckdisease

Pre-process this data to fix proper formatting, corrupted data, missing data etc

You are also allowed to download a dataset of your choice instead. This can be done from one of the many sources of freely downloadable data.
Here are some places to start:
https://www.kaggle.com/datasets
https://www.data.gov/
https://github.com/awesomedata/awesome-public-datasets
You need to download the datasets in a readable and editable format.
The data should contain an acceptable number of data corruptions.
The size of the data does not matter, it can be as small or as large as you like.

In your document, you should first describe your data – what it is about, what are the different variables and what they mean etc.

Explain and document all your steps in detail.
This documentation should have good explanations for all the steps

3. Data Visualization:
Use R to input your data and visualize it with the help of at least 3 different plots. You may choose any library of your choice.
Edit the visual parameters of the data to make the information you want to express clearly visible. Your final visualization should have variables clearly labelled and units equally proportioned.

Ideas you might want to consider:

- Does my data have a spatial or geographic component?
- Does my data have a temporal component, showing change over time?
- How many variables am I trying to represent?

4. (For grade VG only) Attempt *either one* of these questions:
a. Create a list of functions mimicking all the functions of a cash register. For example:
recordSaleValue()
giveOutChange()
checkBalance()
..and others

The program should go like this:

- Ask the user if they want to buy something out of a list of items
- Ask them how many of that item they want to buy
- Ask them if they want to add anything else from the list
- Tell them their final payment
- The user enters an amount
- Tell them how much change they have received

OR

b. Download a larger-sized dataset from the internet. Clean up this data. Create complex, beautiful looking and themed plots in R using the `ggplot2` package.

- The plots should all have different themes
- There should be legends in all these plots
- The legends should also be styled well

- You may also add text labels and annotations

Explain and document all your steps in detail.
This documentation should have good explanations for all the steps.

Your final submission should have:

1. A pdf containing ***all*** the steps of the process described in detail.
    - Add screenshots at every step
    - Add code snippets at every step
    - It should contain all of your plots and their captions
2. A zip file containing R scripts that documents all of your exploration and experimentation
    - This should contain comments to help understand the code
    - They should be named properly enough to understand which file is doing what

Good luck!