Week Two - Assignment

Instructions

This assignment serves to evaluate your grasp of topics covered in week 2. It will assess your ability to proficiently load, assess data structures, subset datasets, apply functions (e.g., unique, head, min, max), merge multiple datasets, and saving data. You have been provided two datasets, each representing a 10% sample of the original 10% 2021 Population and Housing Census micro dataset (population.xls and education.xlsx). When addressing questions related to subsetting, employ various techniques to extract specific information from the dataset based on provided conditions but make sure to select the most suitable subsetting method for each question. **Ensure that your R code follows the best coding practices, including proper indentation and comments where necessary. Please make sure you are working from a project.**

PART 1 - INDIVIDUAL

1. Import the population dataset into your R environment.

2. Explore the data by using functions like head, glimpse () and str () and summary, etc to understand the structure and variables of the data. Provide brief comments on the structure of the population dataset.

3. What is the minimum and maximum age?

4. Subset the data for youth population aged 15 to 35 years and assign it to a new object. (**Verify if the "age" variable is of numeric class; if it isn't, convert it to numeric**)

5. Load the education dataset and print the unique categories of the educational level column. **What are your observations?**

6. Merge the education dataset with the population dataset using **cbind** and assign it to a new object (**Note: The number of rows for the two datasets is the same). What are your observations?**

7. Using an appropriate method and **the unique ids (nqid, pid) along with the common variable region in** both datasets, merge the datasets again. What is the difference between the two merged datasets?

8) Subset the merged dataset to include population 3 years and older.

9) Export the extracted data from question 8 in excel format

10) How many individuals are currently attending, have a master's degree, and are from the Greater Accra region using the data from question 8?

PART 2 - GROUP

1. In your groups, decide on which data set(s) of you will use for the final assignment, and request access to the data.
2. When you have access to the data, import the data into your R environment.
3. Explore the data by using the functions you learned in this week's lecture.
4. Write a short description of your data, using both the documentation on how the data was collected, as well as your insights from your R analysis.
5. Start brainstorming on what kind of analysis you would want to do with the data.