**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 three datasets, each representing a 10% sample of the original 10% 2021 Population and Housing Census micro dataset (population.xls, education.xlsx and housing.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.**

**Questions**

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. a) What is the minimum and maximum age?

b) 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**)

c) How many rows are there in the subset data, and what does it mean?

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

b) 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?**

5 a) 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?

b) Are there any missing values in the educational level column of the second merged dataset?

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

d) Subset the dataset from question 4e to include individuals that have never attended school and assign to a new object.

6. a) How many individuals are currently attending, have a master’s degree, and are from the Greater Accra region using the data from question 4e?

b) What is the lowest age among individuals with a PhD?

c) How many individuals are older than 15 years old and currently attending a primary school education level?

**Assuming you want to investigate the number of individuals who have access to improved water and sanitation.**

7. a) Merge the population datasets with the housing data and comment on the merged datasets.

b) What are the categories for main source of water and how many individuals source water from unimproved sources?

c) How many individuals are living in rented dwellings?

8. a) For the second part of the household analysis, narrow down the dataset to include individuals designated as heads of households, based on their relationship status.

b) Extract households where the head of the household is aged 30 or younger

c) Subset the housing data to include only the H07a and H07b columns

d) Print out the minimum and maximum number of rooms that is occupied by a household in Ghana.

9. a) Create a logical condition to identify households where the number of rooms used for sleeping is greater than the number of rooms occupied. Provide a brief explanation of the significance of this condition.

b) Assume that a large household is defined as a household with more than 5 rooms occupied. Create a logical condition to identify households that are considered large.

c) Create a logical condition to identify households where all rooms are used for sleeping.

You might have to download the 2021 PHC Questionnaire online to help with variable names using this link

**https://census2021.statsghana.gov.gh/gssmain/fileUpload/pressrelease/2021%20PHC%20Questionnaires.pdf**

10. Export the extracted data from question 8a in excel format.