Stat 604
Assignment 04 R

Basic Homework Guidelines:

All of your program code must be written and saved as an R script with the file extension **.R** so that it can be easily opened and executed in R at any time. An easy way to do this is to save the script from the editor that is included with the R software. Be sure to add the .R extension to the end of the file name so your system will identify it as an R script.

Proper documentation is a "best practice" in programming. Whether explicitly stated or not, the header and comment lines will be worth a total of 10 points in every assignment. Each script must begin with 6 comment sections: (1) The name of the script file, (2) The path showing where the script is saved, (3) Created by, which includes your name, (4) Creation date, with the date the script was initially written, (5) Purpose, with a brief explanation of what the script is written to do, and (6) Last executed, containing the date of the last time you ran the script. Just below the "Last executed" comment, include a function that will return the system date and time so that the actual execution time will appear in the console "log" in the pattern: "2021-08-25 16:43:18 CDT". (This function was used in Assignment 02.)

When you have finished writing all your code and saving the script, shut down R or at least clear the console (under the R edit menu). Run your script from this clean environment so the console will contain only one instance of output from the script execution. Convert or print the contents of the console and your output to a PDF document.

After answering the questions at the end of the assignment, convert the script document to PDF for submission to Canvas. Use your First Initial and Last name in your file names according to the following pattern: FKincheloe_HW04_console.pdf, FKincheloe_HW04_output.pdf, and FKincheloe_HW04_script.pdf. Make sure each file is properly uploaded and submitted to the assignment on Canvas.

Scope:

This assignment reinforces the concepts covered in Lessons R01 through R04.

Specific Instructions for this Assignment:

There is a file named states.RData included with these instructions on Canvas. This is an R workspace containing information about the state of Texas. Download the states.RData file from Canvas to an easily accessible folder on your computer.

Perform in R each of the exercises listed below. Include a comment line in your script above the section for each step so that each is clearly identified.

1. Perform housekeeping steps to ensure you start with a clean workspace. The first housekeeping function should display the contents of the workspace. The second housekeeping function should clear the workspace but it is to be commented out so it will not be run automatically should you execute the entire script. Add a step to show which libraries are loaded in your session.
2. Use a function to set up your R session so that everything written to the console will also be directed to a separate text file while still appearing in the console. Include the full path to show where the text file will be written.

3. Invoke R help to research the seq function in the available documentation. This command is not to be part of your program script but will be referenced as the answer to one of the questions at the end of the assignment.
4. Unless you are specifically instructed to give an object a certain name, you are expected to use a name of your own choosing. Write a single line of code to create in the workspace and display a vector of numeric values from 5 to 80 with an increment of 5. Show the type of data in the vector. Show the length.
5. Create in the workspace and display a vector of numeric values from 0.4 to 20 with an increment of 0.4. Show the type of data in the vector. Show the length.
6. Use the first vector to create and display a matrix by columns that is 4 columns wide.
7. Combine the two vectors as rows to create and display a new matrix.
8. Combine the two vectors as columns to create and display a new matrix.
9. Create a vector that contains the nine numeric values 67, 72, 75, 95, 58, 82, 88, 93 and 100. Execute a command that will display only the second, fourth, fifth and sixth members of the vector.
10. Create another vector that contains character strings with values of Dasher, Dancer, Prancer, Donder, Blitzen, Vixen, Comet, Cupid, and Rudolph. Execute a command that will display only the first four members of the vector.
11. Combine the character vector with the numeric vector to create and display a data frame. Execute a function to show the data storage type of the new data frame. Show the contents of the workspace.
12. Load the **states** workspace that you downloaded from Canvas. You may use the R menu to load the workspace initially, but your script must contain a line of code that will load the workspace the next time you run the script. Some versions of R will make an entry in the console log showing the command that loaded the workspace. If you get this line, you may copy it into your script. Otherwise, you will need to find the command syntax in the course slides or R documentation and write the command yourself. Show the contents of the workspace with the newly loaded object(s).
13. Display the object type and the type of data in Texas.
14. Display the object type and type of data in column 1 from Texas.
15. Display the structure of Texas.
16. Display a summary of Texas.
17. Display the first 20 rows and all but column 3 from Texas. Use a negative index value.
18. Create and display a new object from Texas using the first 15 rows, the first column and third column.
19. Add a command that closes the text file and stops sending output to it.
20. After you have run your script for the final time, answer the following questions in a series of comments at the bottom of the script.
    a. What command did you use to invoke help on seq?
    b. How many packages are loaded in your R Session? (Count only those listed as "package:").
    c. What type of data is in the vector created in step 4?
    d. Explain how the values from the first vector are used in the creation of the matrix in step 7.
    e. What is the type of data in the data frame created in step 11?
    f. What is the class and data type of column 1 from Texas?
    g. How many observations and variables are in the Texas data frame?
    h. Explain the relationship between the median and mean of the Pop column?

21. Convert your text file from step 2, script, and console to PDF format then upload and submit the three files to the assignment on Canvas.