

# 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.

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”. (This function was used in Assignment 02.) 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.

For this assignment, include the housekeeping steps at the top of your script. The first function should display the contents of the workspace and the second function should clear the workspace.

When you have finished writing all your code and saving the script, shut down R or at least clear the console. Run your script from this clean environment so the console will contain only one instance of output from the script execution. Copy and paste the contents of the console to a document that can be converted to PDF or use **File, Save to File** (Windows) from the R GUI menu to save the contents of the console to a text file. Sections of the console from steps that produce a listing have been pasted into a sample output file and posted on eCampus so you can see the desired results of these steps. You will need to turn in your entire console and, for these sections, your output should match the samples.

NOTE: Depending on the size of your console window when you run the script, your line breaks may be different. This difference is acceptable.

Convert the console, output and script documents to PDF for submission to WebAssign. Use your Net ID in your file names according to the following pattern: FKincheloe\_HW04\_console.pdf, FKincheloe\_HW04\_output.pdf, and FKincheloe\_HW04\_script.pdf. There are three “questions” on WebAssign; one for each of your homework files. Make sure each file is properly uploaded to the appropriate question. If the upload is successful, you should see the name of the file listed as a hyperlink below the browse box. After you have submitted your homework, make sure that the links are visible and will open up as a PDF document from within WebAssign.

### Scope:

This assignment reinforces the concepts covered in Lessons R01 through R03. NOTE: You will need to nest several functions together in a single command to accomplish some of the objectives of this assignment.

### Specific Instructions for this Assignment:

There is a file named HW04.RData in the **Assignment Data Files** section on eCampus. This is an R workspace containing data on Oklahoma public schools. Download the HW04.RData file from eCampus 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
2. Use a function to set up your R session so that output is directed to a separate file while still appearing in the console. Include the full path to show where the output file will be written.
3. Research the sequence function in the available documentation. Write a single line of code to create and display a vector of numeric values from 4 to 100 with an increment of 4. Show the type of data contained in the vector.
4. Create and display a vector of numeric values from 0.8 to 40 with an increment of 0.8. Show the type of data in the vector.
5. Use the second vector to create and display a matrix by columns that is 5 columns wide.
6. Combine the two vectors as columns to create and display a new matrix. Include a comment on this step that explains how the values from the first vector are used in the creation of this matrix.
7. Combine the two vectors as rows to create and display a new matrix.
8.
  - a. Show the contents of the workspace.
  - b. Load the HW04 workspace that you downloaded from eCampus. 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. This code must contain the full path to the workspace file. 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 R documentation or lecture notes and write the command yourself.
  - c. Show the contents of the R workspace after including the newly loaded workspace.
9. Display the structure of the object loaded from the HW04 workspace.
10. Display a summary of the object loaded from the HW04 workspace.
11. Compute the average of the HSTotal column.
12. Perform a logical test to show which HSTotal values are not missing and are smaller than average. In programming, the term “hard coding” refers to putting an actual value like 350 in your code instead of using another expression to automatically determine that value. Do not hard code the average in your logical test. Instead, use the expression developed in step 11 as a part of the logical test in this step.
13. Display school, city and HSTotal of records that meet criteria in previous step. Use the expression from step 12 in the row portion of your subscript so it will serve as a mask to return only the values you want. Use column numbers in the column portion of your subscript.
14. Use the apply function to compute the average class size for grades 7 through 12.
15. Use the apply function to create a new column called AvgClassSize by computing the average class size of grades 7 through 12 for each school.
16. Display the first 25 rows of the modified data frame.
17. Add a command that closes the output file and stops sending output to it.
18. Submit your script, output, and console to WebAssign.