R Tutorial 3

Instructions:

- Answer all questions.
- Ensure that your findings and results are clearly stated and thoroughly discussed. Please support your arguments using suitable R code with the relevant outputs, interpretations, plots and graphs whenever possible. You should support your argument using appropriate theory that is appropriately referenced.
- The R commands that you use in obtaining your results for all questions must be documented in a R script file. These scripts must be clearly commented. Ensure that any output is clearly stated and interpreted separately from the code as additional comments.
- Include the task name, your name and surname, and your student number in your R script file.
- You MUST label each answer by question number and, where a question has multiple parts, label each part of the question CLEARLY.
- On completion of your assignment, please submit onto RUconnected. If there are any issues uploading onto RUconnected, you may email your submission to: a.langston@ru.ac.za. Please submit your R script file and any other saved data files and plots mentioned in the questions below. Your student number should be included in the name of each file that you submit.
- Each student must complete an individual assignment. You will be assessed based on the quality and/or correctness of the R code, its outputs, and your explanations and interpretations. Acknowledge any help you may have received. Feel free to note any help you may have given to other students in the course.
- This assignment must be submitted by Tuesday, 30 July 2024 by 17:00. Late submissions will be penalized.
- Please note the Rhodes University and the Rhodes University Department of Statistics plagiarism policies.

Questions:

1. There are a number of autonomous communities in Spain. The Wheat Table that follows gives the wheat harvested surfaces in 2004 by autonomous communities in Spain measured in hectares.

Wheat Table	
Community	**Wheat Surface**
Galicia	18817
Asturias	65
Cantabria	440
Pais Vasco	25143
Navarra	66326
La Rioja	34214
Aragon	311479
Cataluna	74206
Islas Baleares	7203
Castilla y Leon	619858
Madrid	13118
Castilla-La Mancha	263424
C. Valenciana	6111
Region de Murcia	9500
Extremadura	143250
Andalucia	558292
Islas Canarias	100

   Provide R code to answer the following questions:

(a) Create the variables `community` and `wheat.surface` from the Wheat Table in this problem. Store both variables in a data.frame named `wheatspain`.

(b) Find the maximum using the `max()` function, the minimum using the `min()` function, and the range using the `range()` function for the variable `wheat.surface`.

(c) Which community has the largest harvested wheat surface?

(d) Sort the autonomous communities by harvested surface in ascending order. The `order()` function may be useful here.

(e) Sort the autonomous communities by harvested surface in descending order. The `order()` function may be useful here.

(f) Create a new data frame called `wheat.c` where data for Asturias has been removed.

(g) Add the Asturias data back into the data frame `wheat.c` and call this data frame `wheat.d`.

(h) Add a new variable into `wheat.d` called `acre` indicating the harvested surface in acres rounded to two decimal places. Call this data frame `wheat.e`. Note that 1 acre $\approx$ 0.4047 hectares.

(i) What is the total harvested surface in hectares and in acres in Spain in 2004?

(j) Define in `wheat.e` the `row.names()` using the names of the communities then remove the `community` variable from `wheat.e`. Call this data frame `wheat.f`.

(k) Determine the communities with less than 40,000 acres of harvested surface and find their total harvested surface in hectares and acres.

(l) Create a text file called *wheatdata.txt* from the `wheat.f` data frame using the command `write.table()`. Ensure that you submit this file as part of your assignment.

2.

(a) Create a list named `some_info` that contains:

- the number 7;
- a vector with numbers from 1 to 10;
- the matrix $\begin{bmatrix} 4 & 8 & 12 \\ 16 & 20 & 24 \end{bmatrix}$;
- the text string "Hello".

(b) What is the value of `some_info[3]`? Explain how R is indexing the list in this case.

(c) Use appropriate indexing to replace the value 24 in `some_info` with 17.

(d) Name the elements of `some_info` as "number", "vector", "matrix", and "text", respectively.

(e) Add another element to `some_info` named "dataframe" containing the `wheat.f` data frame defined in Question 1(j).