

---

## R Tutorial 4

---

### 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, 06 August 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. Consider the data frame `wheatUSA2004` in the `PASWR` package, which contains the USA wheat harvested crop surfaces in 2004 by states. Provide R code to answer the following questions:
  - (a) Import the data set into R.
  - (b) Attach the data frame `wheatUSA2004` and use the function `row.names()` to define the states as the row names.
  - (c) Define a new variable called `HA` for the surface area given in hectares (1 acre  $\approx$  0.4047 hectares) and add this to the `wheatUSA2004` data frame.
  - (d) Remove the `STATES` variable from `wheatUSA2004`.
  - (e) Sort the data frame according to the harvested surface area in acres in descending order. The `order()` function may be useful here.
  - (f) Which states fall in the top 10% of states for harvested surface area?
  - (g) Save the contents of `wheatUSA2004` in a new file called `wheatUSA.txt` using `write.table()`. Ensure that you submit this file as part of your assignment.
  - (h) Find the total harvested surface area in acres for the bottom 10% of the states.
2. Consider the data stored in the `pressure.csv` file on RUconnected, which contains data on the relationship between temperature in degrees Celsius and vapour pressure of mercury in millimetres (of mercury).
  - (a) Import this data set into R using `read.csv()` and store it in a variable named `pressure`.
  - (b) Plot the pressure data. Ensure that an appropriate title and the relevant axis labels are included on the plot. All other arguments should be left as their default values. Ensure that you submit this plot as part of your assignment.
  - (c) Plot the pressure data using solid orange squares. Ensure that you submit this plot as part of your assignment.

- 
- (d) Plot the pressure data with a dashed line. Ensure that you submit this plot as part of your assignment.
- (e) Complete the below steps. Ensure that you submit this plot as part of your assignment.
- Plot an empty plotting region using the pressure data.
  - Add the pressure data to this empty plot using the `lines()` function. Use a blue dotted line.
  - Add the pressure data to the current plot using the `points()` function. Use green stars.
  - Add a legend to this plot using the `legend()` function. The final plot should look similar to the plot below:

