
R Tutorial 1

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, 16 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.
 - (a) Calculate the following numerical results using R:
 - i. $(23.4 + 9) \cdot 1.6$;
 - ii. $\frac{74+5}{(2 \times 3)^4}$;
 - iii. $3 - e^{\ln 3}$;
 - iv. $5 + \frac{\frac{2}{18}-1}{3}$;
 - v. $\cos(3\pi + 1)$;
 - vi. $12 - (\sqrt{12})^2$.
 - (b) What would you have expected in (iii) and (vi)? Explain what R is doing.
 - (c) Round the results obtained in (iv) to four decimal places using R.
2. Create a vector named `countto15` that contains all numbers from 1 to 15.
3. Create a vector named `symbols` with character entries “+” appearing 14 times, “!” appearing 8 times, “&” appearing 12 times, and “%” appearing 4 times.
4. Create a vector named `countby51` that is a sequence of 51 to 510 in steps of 51.
5. Evaluate the formula

$$x = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$$

for the values

-
- (a) $p = 2$, $q = \frac{1}{4}$;
(b) $p = q = 1$. Interpret the warning message for this question.
6. You recorded a car's mileage at its last eight fill-ups as
- 65311 65624 65908 66219 66499 66821 67145 67447
- (a) Enter these numbers into a variable named `gas`.
(b) Use the functions `head()` and `tail()` on the data. What do these functions return as their output? Based on these outputs, explain what operation is performed by each function.
7. Create a vector named `months` containing the numbers 29, 63, 7, 23, 84, 10 and 9. Use this vector to create a new vector named `years` by dividing the months by 12.
8. Provide the missing values in `rep(x=seq(from=__, to=__, by=__), n= __)` to create the sequence 20, 15, 15, 10, 10, 10, 5, 5, 5, 5.
9. For understanding data in a vector `v`, the command `diff()` can be useful. Using the help-function in R, briefly describe how this function is used and what output it produces. Give an example of using this function on a vector of your choosing called `v`. This vector should contain at least 10 elements.
10. You track your commute times for two weeks (ten days), recording the following times in minutes:

17 16 20 24 22 15 21 15 17 22

Enter these data into R and store it in a variable named `commute`.

- (a) Use the function `min()` to find the longest commute time, the function `mean()` to find the average, and the function `min()` to find the minimum.
(b) Use the `which()` function and appropriate indexing to list the commute times that are less than 20 minutes.
(c) Looking at the data, you realize that the 15 was a mistake. It should be 18. Fix this using the appropriate indexing.
(d) You track your commute for another 5 days and record the following times in minutes:
- 18 21 16 14 21
- Add these new values to the `commute` vector using the appropriate indexing.
- (e) Now use the function `max()` to find the longest commute time, the function `mean()` to find the average, and the function `min()` to find the minimum.
11. Which number is the largest: 2.2222e12, 10000000, 4.6e1, 3.1e6. Based on these results, interpret what is meant by the 'e' symbol in R.
12. Assess whether the string "R rules!" is equal to the string "r rules!" in R. Explain why R reports this result.
13. A friend of yours is learning how to use R and has written the following code:

```
> library(dplyr)
> mtcars %>%
  group_by(cyl) %>%
  summarize(mean_mpg = mean(mpg))
```

When they run this code it produces an error message:

```
Error in library(dplyr) : there is no package called 'dplyr'
```

You know that R does in fact have a package called `dplyr`. Briefly explain to your friend what they have done wrong and how they should fix this error in R. Ensure that you write the necessary code as part of your answer.