

DSI Data Science: Introduction to R Assessment 2

Introduction and Goals

During the duration of the course, we had the opportunity to explore the RStudio environment, where we could navigate through its features. We also had the chance to download the tidyverse package (Wickham, H. et al, 2019) and delve into its functions, which enabled us to manipulate, wrangle, and visualize data effectively. Moreover, we gained knowledge about crafting personalized functions, as well as exploring tools and techniques for functional programming and data simulation. Additionally, we touched upon significant topics such as reproducibility, bias, diversity, inclusion, ethical considerations, equity concepts, data security, and best coding practices.

Grading Rubric

This assignment will be graded as Pass/Fail. The purpose of this assessment is to ensure that learners can acquire the following skills:

Component	Yes	No
1. Learner demonstrates proficiently in manipulating tabular data using <i>dplyr</i> : A Grammar of Data Manipulation.		
2. Learner can apply manipulation and wrangling techniques to reshape data effectively.		
3. Learner is capable of creating meaningful visualizations suitable for different types of data using <i>ggplot2</i> : A Grammar of Graphics.		
4. Learner is able to develop custom functions capable of analyzing input data and generating output.		
5. Learner understands the importance of equity, diversity, and inclusion (EDI) practices in the field of data sciences.		

Tasks

Please respond to all the questions sequentially. You have the option to utilize a PDF editor, the provided R markdown template, or any other platform/software of your preference to generate the PDF document containing both the questions and answers. Alternatively, you can open your R script using a text editor and export it as a PDF document. **When submitting your answers, make sure to include both the question number and the question itself, along with your responses.** This requirement is in place to ensure that no questions are skipped and that all sub-questions within each main question are addressed. When copying and pasting the questions, you may disregard any formatting such as italics.

For questions 1 and 2, we will use the **built-in dataset *mtcars* from R base package *datasets***. To access this dataset, type the following in your script:

```
data(mtcars)
?mtcars #to view data format and understand column names
```

1.

- (a) Please create a custom function called “mpgFilter” using both tidyverse functions and other base R functions if needed. This function should have two arguments: “dataset” and “minMpg”. Within the function, utilize `dplyr::filter` to filter the user-provided dataset based on the specified **minimum** mpg (miles/gallon) in the “minMpg” argument. Finally, return the resulting output to the user within the function. Please include all the necessary code.

#CODE HERE

- (b) Run the mpgFilter function you wrote using the built-in dataset `mtcars`. Set the dataset argument to `mtcars` and minMpg argument to 20. Save the resulting output from function mpgFilter into an object called `output_mpgFilter`. How many rows and columns are in the resulting data frame, i.e., the dimensions of `output_mpgFilter`? Please include all the necessary code.

#CODE HERE

- (c) Using the “output_mpgFilter” object obtained previously, mutate it by adding a new column utilizing a dplyr function. This new column should display the ratio of mpg (miles/gallon) to hp (gross horsepower) from the mpg dataset. You can achieve this by dividing the mpg by hp. Name for the new column “Ratio.MPGHP”. Please include all the necessary code.

#CODE HERE

2. Generate a plot for the `mtcars` dataset using the `ggplot2::ggplot` function. Customize the plot by incorporating layers from the `ggplot2` package. This customization should include an informative title, clear labels, axes, color and/or size, and an overall appealing appearance. Please include the code for generating the plot. Finally, provide the created plot along with a detailed description, consisting of 3 to 4 sentences, explaining what the plot represents.

#CODE HERE

3. To ensure equity, diversity, and inclusion (EDI) practices in your data science work, what steps can you take? Provide examples. Additionally, what challenges might you encounter or circumstances that could be beyond your control? Please elaborate on this topic using approximately half to three-quarters of the page.

Submission Instructions

Remember: Submissions should only be in PDF format. When submitting your assignment, name PDF using format: `LASTNAME_FirstInitial_A2.PDF`.

E.g., `GALLUCCI_J_A2.PDF`.

In your submission to the instructor, you must provide both the question number AND the question, in addition to your answers. Answer all the questions, in order.

How to Use R Markdown To Create A PDF With Answers

R Markdown is a formatting system that enables the creation of reproducible and dynamic reports using R. These reports allow for the inclusion of R code and its corresponding results in various formats such as slideshows, PDFs, HTML files, Word documents, and more. If you opt to use R Markdown to generate PDF files with solutions, it will require some time to familiarize yourself with R Markdown’s syntax and capabilities. To simplify this process, a template named ‘**Assessment2_template.Rmd**’ has been provided, which you can utilize. To locate the provided R Markdown template, please navigate to the Assessments folder on GitHub: https://github.com/UofT-DSI/04-intro_r/tree/main/Assessments

Open this template file in RStudio

In order to generate PDF documents from R Markdown, two essential components are required:

- The “rmarkdown” R package
- The LaTeX distribution.

Note: There are various LaTeX options available, such as MiKTeX, MacTeX, TeX Live, and TinyTeX. You have the freedom to choose any of these LaTeX distributions that suits your needs. In this instance, I will demonstrate the use of TinyTeX. To install TinyTeX, you can utilize the “tinytex” R package.

```
# install the rmarkdown package

#install.packages("rmarkdown")

#library("rmarkdown")

# install the tinytex package

#install.packages("tinytex")

#library("tinytex")

# to install TinyTeX

#tinytex::install_tinytex()
```

The provided template already includes all the question numbers and the corresponding questions. You will only need to insert your answers. Before you begin entering your answers, please follow these steps:

1. Open the file ‘Assessment2_template.Rmd’ in RStudio.
2. Locate the ‘Knit’ icon in RStudio.
3. Click on the ‘Knit’ icon and select ‘Knit to PDF’ to generate a PDF output.

You are recommended to add small chunks of code at a time and ‘Knit’ the document. For more information including basics, see <https://rmarkdown.rstudio.com/lesson-1.html> or seek help during tutorial early.