

# DSA 557 1.0 Scientific Writing and Communication

## Rules

- There are two assignments listed in the document. Each is worth 50 marks.
- It is the student's responsibility to ensure that the submission is complete, correct, and accessible to the instructor.
- The instructor reserves the right to penalize any student who fails to adhere to the submission rules or who engages in academic dishonesty.
- Students are encouraged to contact the instructor if they have any questions or concerns about the submission rules or the assignment itself.
- It is the student's responsibility to manage their time effectively to ensure that they have enough time to complete the assignment within the given timeframe. This includes allocating enough time to research, plan, and write the assignment, as well as allowing for unexpected circumstances or delays. It is important for students to plan their workload and prioritize their assignments to ensure that they are able to complete them on time and to the best of their ability.
- Contact mode: emails only. Please do not make telephone calls.

## Assignment 1

**Total marks: 50 marks**

Assignment release date: **27 March 2024**

Due date: **7 April 2024**

Late submission: I will not accept any submissions after April 12, 2024. A penalty is applied for submissions after April 7, 2023 (-2 marks for each day).

### Assignment description:

In this assignment, you are going to work on a dataset from TidyTuesday project. TidyTuesday is a weekly social data project in R organized by the R for Data Science community. It is a great way of improving your Data wrangling and visualization techniques, sharing and learning from others.

## TidyTuesday Data Repository

Every week since 2018, TidyTuesday moderators have provided a weekly challenge focused on the creation of data visualizations based around interesting, timely, or socially-important topics, such as those with data about the Women's World Cup, incarceration trends, and school diversity in the United States. The official TidyTuesday data repository is available at <https://github.com/rfordatascience/tidytuesday> (Scroll down on the page until you see DataSets) section. Select any dataset from this repository for your assignment. Links to data repositories are given below.

[Link to 2018 datasets](https://github.com/rfordatascience/tidytuesday/tree/master/data/2018)

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2018>

[Link to 2019 datasets](#)

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2019>

[Link to 2020 datasets](#)

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2020>

[Link to 2021 datasets](#)

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2021>

[Link to 2022 datasets](#)

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2022>

[Link to 2023 datasets](#)

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2023>

## Task

Data is just a collection of numbers until you turn it into a story. Here you are trying to deliver the story through data visualization. Visualize data and create a report to tell an engaging story for your data analysis.

- i. Include suitable plots to uncover the stories, patterns in the data.
- ii. Maximum number of plots to be included in the report is 6.
- iii. Interpret the plots.
- iv. Your report should be fully reproducible. You can use Quarto or Jupyter notebook to prepare your report.
- v. Use Python to visualize data.

## Submission

If you use jupyter notebook submit .ipynb file.

If you use Quarto submit .Qmd and .html file

Submit your files via LMS.

## Marks

Total marks: 48 (6 marks for each plot).

- Suitability of the plot: 4 marks
  - Choose the right type of vizualization or chart for your data.
  - Use text and labels to clarify, not clutter.
  - Use color to highlight important information or to differentiate or compare.
  - Avoid misleading visualizations.
  - Keep your visualizations simple and informative.
- Interpretation of the chart: 2 marks

Overall structure of the report: 2 marks

## Additional Resources

Tidy Tuesday: A weekly podcast and community activity brought to you by the R4DS Online Learning Community

visit: <https://www.tidyuesday.com/>

## Assignment 2

**Total marks: 50 marks**

Assignment release date: **27 March 2024**

Due date: **20 April 2024**

Late submission: I will not accept any submissions after April 20, 2024.

### Assignment description

The goal in this assignment is to write a literature review paper (of approximately 1,500 words) on the topic of machine learning applications in medicine and health care.

Use the list below as a guide when writing your review paper

- An overview of the major applications in different fields in medicine and health care
- A review of the approaches used in different applications
- For each approach: discuss the problem(s) it addresses, methods used, theoretical foundation, the results of this approach, and any unresolved issues and study limitations.
- A description of the challenges for future research.

### Format

Papers should be approximately 1,500 words (author information, title, abstract, and the list of references are excluded from the 1,500 word limit). Citations and references can be in any common format (APA, Chicago, or ACM are acceptable, for instance).

### Submission

Submit PDF file.

Submit your files via LMS.

### Marks

1. Factual correctness, completeness of the answer, and accurate formatting and style of writing: 15
2. Synthesis and critical analysis of the topic: 15
3. Methodological explanation of the concepts: 20