DSA 557 1.0 Scientific Writing: Assignment 2

Total marks: 40 marks

Assignment release date: **28 January 2023**

Due date: **19 February 2023**

Late submission: Submitting late work will be penalized by 1 mark per day. I do not accept any submissions after 28 February 2023.

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

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 5.

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: 40 (8 marks for each plot).

- Suitability of the plot: 5 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: 3 marks

Additional Resources

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

visit: https://www.tidytuesday.com/

Requesting a deadline extension

You are allowed to apply for a deadline extension for the assignment under the following conditions only. Category A

An illness that requires the student to miss all of his or her classes for 2 days or more. Category B

A period of grief for the loss of a family member or friend. Category C

A family emergency. Valid documentation of proof is required prior to the assignment deadline.

The philosophy driving our late policy has three goals Keep things fair.

Encourage good time management.

Encouragement and appreciation of on-time submission students.