Final Report Assignment

Reproducible Data Analysis 2023 T4

The goal of the final project is to **analyze a dataset with R** and **write a report about it using Quarto**. Futhermore, you will **present** your analysis during the final class meeting on 2023-11-21.

Dataset

The dataset comes from the Tidy Tuesday Project, which releases datasets for practicing data analysis skills once per week. There are five datasets to choose from. Each group will select one dataset during class on 2023-11-07:

- World's Oldest People
- Olympics
- Chocolate ratings
- Transit cost
- Student to teacher ratios

Report Content

You should use the R skills we have learned in class to **analyze the data**.

Your report should include the following components:

- **Introduction**
  - What is the dataset **about**?
  - Where does the dataset **come from**?
  - What **question(s)** will you answer by analyzing the dataset?
- **Visualization**
  - Please generate **at least two graphs** that show relationships between variables in the data

- **Discussion**
  - What is the **message** in your data visualization?
  - What **insights** can the dataset provide (in other words, **what conclusions** can you draw from the data)?

If you cite any references, please include a reference list at the end.

Example Report

I have made an example report that you can refer to when working on your project: https://github.com/data-analysis-chiba-2024/gapminder-report

Group Work

This is a **group project** (four members per group). Group members will be responsible for dividing the work amongst themselves. For example, two members could work on one graph, and the other two members could work on the other graph. You will be assigned to groups in class on 2023-11-07.

Please **use git and github to share your work** by pushing and pulling as we learned in class. I recommend that each person creates a copy of the report (you could name the file with your initials, such as `report-jn.qmd`) when you are working on it. Then, when you are ready to submit the final version **one person** should edit and push `report.qmd`. This is to avoid **merge conflicts**, that is, incompatible changes made to the same file by different people. There is a way to fix those, but it is an advanced topic that we weren't able to learn in this class.

Presentation

You will present your analysis in class on 2023-11-21. You can use whatever software you want to create the presentation. (We didn't have time to cover it in class, but **Quarto** can be used to make slides too! Ask the instructor if you're interested).

Your presentation should include the same basic information as the paper:

- Introduction
- Visualization
- Discussion

Each group will have **10 minutes** to present, and up to **5 minutes** to answer questions.

Evaluation

The instructor will evaluate each report and presentation according to the rubric.

You will also evaluate **each other groups' presentation** (**Form A**; hard copy, turn in at the end of class on 2023-11-21), as well as **the contribution of each other member of your group** (**Form B**; on Moodle, due by 2023-11-21 11:59 PM).

Please see the rubric at the end of this description for evaluating other presentations and for your own reference.

Language

Both the report and presentation must be **in English**

Submission

The **report** (`report.qmd`) is due on 2023-11-20 11:59 PM. It should be submitted the same way as the homework assignments: by pushing to GitHub. Please also submit the assignment on Moodle as usual by entering the URL of your repo. Note that you should submit (commit) the `report.qmd` file, not the `.docx` file that it generates.

The **presentation** will be given in class on 2023-11-21.

Rubric

	4 (excellent)	3 (good)	2 (needs improvement)	1 (poor)
Introduction	The dataset is described succinctly and clearly, including where it comes from.	The overall purpose of the dataset is clear, but some details are lacking.	Only the title of the dataset is given; no information about where the data come from.	Introduction missing or no useful information provided
Visualization	All graphs are clearly labeled and easy to read. The message conveyed by each graph is clear.	Some labels may not be easy to understand (not changed from the default in the data). There is a pattern in the graph, but it is not well explained.	Cannot understand what the labels mean. No clear pattern in the graph or explanation given.	One or more of the graphs is missing.
Discussion	The message of the visualization is clearly explained. A conclusion is provided that draws insight from the data.	Some interpretation of the visualization is provided, but the insights to be gained from the data are not clear.	A general discussion of the dataset is provided, but there are no direct references to the visualization or concrete conclusions drawn.	Discussion missing, cannot be understood, or mostly unrelated to the dataset.

	4 (excellent)	3 (good)	2 (needs improvement)	1 (poor)
Style and grammar (some aspects only apply to the presentation)	All English grammar is correct. Style is consistent. Slides are easy to read.	Some grammar may be incorrect, but meaning is still clear. Slight deviations in style (such as different formatting of references) may occur. Slides are generally easy to read, but may be a bit cluttered.	There are grammatical errors that obscure the meaning of what the authors are trying to say. Style is not consistent. Slides are poorly designed, at times hard to understand.	Even basic English is frequently incorrect. The message is very hard to understand because of poor English. Slides have no consistent style and are very difficult to read.

In addition, your report will be scored on the following criteria:

- The code should run without errors
- The `.docx` file should be generated correctly from the `.qmd` file
- The `.docx` file should not include any code chunks (code chunks should be hidden with `|# echo: false`)

Scoring

Points will be assigned for the final project as follows:

- Final report: 45%
- Final presentation: 45%
- Participation (filling out Form A and B): 10%