Introduction to Data Visualization (DS4073)

Group Project

1. Topics

There are six topics can be chosen:

Topic 1: Health data visualization

Topic 2: Climate and environment data visualization

Topic 3: Education data visualization

Topic 4: Financial data visualization

Topic 5: Classical Chinese poems data visualization

Topic 6: Other cool data visualization (You can choose other interesting topics)

2. Workflow Stages

- **Problem definition:** Each group should choose one topic and design your project title (a **concrete** title, e.g. COVID-19 situation in Shanghai from march, 2022).
- **Data collection and preprocessing**: Each group needs to collect the data (either downloading some open source datasets or scraping the data from web pages). The collected data may contain missing and noisy entries or in consistency schemas. Therefore, you need to perform data preprocessing and data cleaning.
- **Data and task abstraction:** Analyze the datasets and do the data abstraction and task abstraction. You need find 4-8 tasks/goals (depending on the number of members in a group, e. g. if there are 3 students in a group, you need find at least 4 tasks).
- **Design data visualization**: Use the knowledge learned from the lectures to design your visualization and explain why (e.g. goals/tasks, marks, visual channels, color mapping type, effectiveness principle, expressiveness principle, other design principles, etc.).
- **Implementation:** You need to use **R** to implement your visualizations. A dashboard may need to be created to display various types of visual data in one place.

3. Assessment of Group Project (Total: 100% + 10%bonus)

- Correctness and completeness (in data preprocessing, data abstraction, task

abstraction, visualization design, code implementation) 50%

- Documentations (justifications) 20%
- Presentation (Organize ideas into clearly identifiable sections with all information presented in logical sequence) 20%
- Demonstration (clarity and smoothness of the demo) 10%
- Bonus (outstanding design of visualization, extract an interesting task from an innovative perspective, etc.) maximum 10%

4. Requirements (*Please read the following requirements carefully*)

- Each group consists of **at most 4 students**. Each member in one group should be ready to speak for **3-4 mins** during the presentation.
- Prepare a set of **at most 25 PowerPoint slides** and a demo (.rmd and .html). The **PPT** should contain the background of the topic (why you study this topic), contribution of each member, the process of data preprocessing, tasks/goals, the design thought and its justification, the sketch of R implementation (What kind of packages, functions will you use? What are their functionalities?).
- Divide up the workload. Each student should be responsible for only one part of the presentation. Do not divide the presentation into many different small parts.
- Please submit your **group project title** to iSpace due to **November 4, 2023.**
- Please submit the **PPT, project report, demo files** (including the annotated R codes, html, datasets, and other related documents), and a **README file** (to introduce the information for your code and explain how to execute your code) to iSpace due to **December 4, 2023**. Compress all the documents in one **.zip file** and use your group number as the file name, e.g. *Group_1.zip*.
- **Project report** should be well organized and include background, data description and processing, task abstraction, visualization design and the corresponding theories/principles, visualization results, and conclusion.
- Don't plagiarize, otherwise, everyone in the group obtains zero grade.

The following websites may help you to find the dataset that you may be interested in, of course, you can find the dataset elsewhere or scraping from web pages.

http://archive.ics.uci.edu/ml/datasets.php

https://www.tableau.com/learn/articles/free-public-data-sets