Tsinghua Shenzhen International Graduate School
Data Visualization
Spring 2025

Assignment 2

Issued: March 15, 2025 **Due:** March 29, 2025

In this assignment, you will have the opportunity to apply the techniques you have learned to create a visualization diagram about any topic that interests you.

Requirements

1.1. To get started with this visualization assignment, you need to choose a topic first. The topic could be related to your current research work, your personal interest, or a news article that you care about. If you're having trouble finding a topic, we have provided two additional datasets for you to choose from. Please refer to the Section "Additional provided datasets" for more information.

1.2. You are required to create **at least two visualization graphs** for the dataset related to your chosen topic.

The purpose of these graphs is to effectively convey **different aspects** of the important information related to your chosen topic. Take the additional provided dataset – "National Football League (NFL) Play-by-Play Data" for example, you could create the first graph that depicts the relationship between the success of a team and the presence of star players. Then, you could create the second graph that reflects the rise and fall of each team over time, providing insights into which teams have been most successful historically.

Please note that this assignment places a strong emphasis on "Creativity". Therefore, it is advisable to avoid just using simple plot charts or scatter plots. Instead, please explore more advanced visualization techniques and tools to present your data in an engaging and innovative manner. Be sure to push the boundaries of your creativity while maintaining a clear and concise representation of your findings.

1.3. You are required to write a **report** with a description and interpretation of your graph. The report should contain, at minimum, the following information:

(a) An introduction of your chosen topic.

(b) A description of the dataset you are using, including attributes, types, and other relevant information.

(c) The graphs you created for this visualization task.

(d) A description of your graph that explains what it represents, the data it displays, and the insights it provides. This section should highlight any important features or trends that are visible in the graph.

(e) An explanation of the visual encoding techniques used in your graph. This section should detail how you achieved the visual representation of your data, including the choice of colors, sizes, shapes, and other design elements.

1.4. In addition to writing a report, you are also required to prepare for an **oral presentation** to showcase your work. Please note the following instructions:

(a) Your presentation should be no longer than **3 minutes**. Please practice your presentation to ensure that you are able to convey all of the necessary information within this time frame.

(b) You should prepare a **one-page slide** to accompany your presentation. This slide should include the graphs you have created as well as any other key information you would like to highlight.

(c) During your presentation, at least one question will be asked by the professor or your classmates. Please be prepared to provide a clear and complete answer to the question.

The exact presentation time has not been decided yet. We will inform you as soon as we have more information on the schedule.

1.5. It is important to be transparent about any external resources used in completing an assignment. If you have discussed the assignment with anyone or used any large language models, such as ChatGPT, Github Copilot, or BLOOM, please specify how you used the external resource(s) - for example, did you use them to brainstorm ideas, write code, or proofread your work? If you have completed the homework unaided, please state that explicitly in your report by saying, "I have completed this homework unaided."

Submission

To ensure that your submission is complete and easy to manage, please make sure to follow these guidelines:

1.1. Your submission should consist of two parts: **the programming part** and **the report part** (Both Chinese and English are acceptable). Please compress both parts into a single .zip format file, which should be named using the following format: **'[Your Name]_[Your Student Number].zip'**.

1.2. The report should be integrated into a single PDF file, named 'report.pdf'. Please make sure to put this file outside of the appendix part folder, so it's easy to locate.

1.3. The appendix should be placed in a folder named "Visualization Appendix." This folder should contain all of your code, slides, and other appendix files.

1.4. To ensure that your file is easy to upload and download, please limit the size of your uploaded file to no more than 50MB.

Grading

Important! Your homework will be graded based on the following criteria:

- Completeness (40%): Have you completed the visualization tasks using your chosen visualization tools? Have you provided a description and interpretation for your topic and graph in the report? Have you successfully presented your work in a clear and concise manner?

- Clarity(30%): Have you made your graphs easy to read and understand? Have you used appropriate colors, sizes, shapes, and legends to distinguish different elements on your graphs? Have you added informative titles and labels to your graphs? Have you provided a well-prepared presentation, or given a thoughtful answer to the presentation problem?

- Creativity(30%): Have you gone beyond the minimum requirements of the assignment to create a particularly engaging design? Have you used creative approaches to visualize your data and convey your message effectively?

Additional provided datasets

- National Football League (NFL) Play-by-Play Data

  NFL Play-by-Play Data is an opensource dataset that provides National Football League (NFL) play-by-play data. This dataset contains all the regular season plays from the 2009-2018 NFL seasons. Each play is broken down into great detail containing information on: game situation, players involved, results, and advanced metrics such as expected point and win probability values. It facilitates the researchers to carry out modern and reproducible research in football analytics. Here, you are encouraged to explore this dataset and try to discover some interesting phenomena or make some reliable predictions at player/game levels.

  [1] To download the dataset:

  `https://cloud.tsinghua.edu.cn/d/d933abea194547268110/`

  [2] A brief introduction:

  `https://github.com/ryurko/nflscrapR-data`

- High Spatial-Temporal Resolution of PM2.5 Measurement

  S&M-HSTPM2d5 is an open-source dataset for anyone interested in researching air pollution in Chinese cities Foshan, Cangzhou, and Tianjin. This open-source dataset offers high spatial-temporal resolution PM2.5 measurements, complete with timestamps and GPS locations of both mobile and static devices. Researchers can use this data to create fine-grained city-scale air pollution maps, providing critical environmental insights for city managers and residents alike. By leveraging this information, city managers can make informed decisions when it comes to planning, managing, and developing their cities. Meanwhile, residents can use this data to steer clear of polluted areas and reduce their exposure to harmful pollutants.

  [1] To download the dataset:

  `https://cloud.tsinghua.edu.cn/d/af81a3e78a1e4fefb5d9/`

  [2] A brief introduction:

  `https://zenodo.org/record/4028130#.Y_tUUGRBwuU`

  Important! Please note that if you decide to use this dataset – the same dataset as you did in Assignment 1, please avoid using the same graphs or visualizations that you used in Assignment 1. We encourage you to explore new ways to visualize the data and discover new insights that may not have been uncovered in your previous assignment.