Project 1: Topics and grading scheme

Introduction to Visualization, Spring 2023

Step 1: Choose from one of the topics below. You may modify it or create your own, in which case confirm verbally or by email with one of the instructors.
Step 2: Indicate what you have chosen as a comment in the sign-up sheet: link here
Step 3: Choose a presentation order (a number between 1 and 21), and comment it in the spreadsheet. Presentations will start at 11:10 and go until the end of the lab (12:50).

Grading scheme: Your grade will be out of 10, with points assigned by the following scheme:

- *3 points:* Presentation on **February 23**. Criteria:
  - (2 points) Was the idea clear?
  - (1 point) Were questions from students / instructors answered well?
- *5 points:* Submitted material. Criteria:
  - (1 point) Is there at least one visual? Is it labeled and titled clearly?
  - (1 point) Is there text explaining the visual(s) in full sentences? Are all parts of the visual(s) explained?
  - (1 point) Has color, text size, layout been used in a visually appealing way?
  - (1 point) Is there analysis interpreting the results and the context?
  - (1 point) Have sources been cited?
- *2 points:* Peer review. Average of students' grading of your presentation:
  - *(scale of 1-10)* What did you think of the presentation?
  - *(scale of 1-10)* What did you think of the poster?

Keep in mind:

- Your submission should be in PDF format.
- Your project should be submitted to ORTUS before you present.
- You should have both image(s) and text, so that someone who did not see your presentation can still understand the main idea.
- Suggested format is a single page (does not have to be standard size), optimized for viewing on a computer screen.
- Your presentation should be 4 minutes max in length, 1 minute for questions.
- When others are presenting, think of questions to ask!

Topics: Suggested topics are given below. Please choose one. You may modify the details.

1. **Text analysis:** Choose a website with lots of text (such as news). Take several pages from this website, and identify the most common words in the website. Compare across different parts of the document, different dates. Make conclusions in the context of the type of website you chose.
    - *Suggested modifications: choice of website, restriction of analyzed text, …*
2. **Comparison:** Find a data visualization online that you do not like. Create your own visualization from the same data set that improves it. Discuss how the first impression changes and what is emphasized.
    - *Suggested modifications: Compare existing visualizations of the same data, compare several visualization with the same flaw, compare several visualization using a particular color / layout, …*
3. **Weather data:** Find weather data (such as, but not limited to: temperature, precipitation, humidity, …) of a given location for several years. Plot the data to find patterns in a year and to compare years. And / or, choose several locations on the map throughout a year.
    - *Suggested modifications: Frequency of data measurement, different type of regular measurement, comparison of different sources, …*
4. **Animation:** Create an animated visualization of how infection spreads among individuals based on exposure time, distance and probability, as the individuals move through an environment. *Technical suggestion: create several still plots in matplotlib and combine them into a moving image.*
    - *Suggested modifications: Animate collected data that increases over time, animate a line plot, …*
5. **Daily visual:** Take an aspect of your everyday life and log it for a week. For example. what time you get up, what time you eat, how many cars / people you see in a day, etc. Create a visualization of this aspect. *Inspiration:* http://www.dear-data.com/theproject
    - *Suggested modifications: Log several aspects and overlay them to observe any influences, use existing data online, log activities of several people, …*
6. **Open data analysis and accountability:** Using an open data portal such as https://data.gov.lv/lv , choose a data set and create a visualization for it. Discuss the purpose of your visualization, for example, holding governments accountable, or exploring the data space. If possible, take several data sets (with the same / similar labels) and compare them.

- *Suggested modifications: Start with a hot political topic and find the original data source, analyze voting data (participation, selection compared with social indicators), …*

7. **Visual journalism:** Create a story about companies that went public. Create visual(s) that motivate your story, plotting change over time, or comparison across prices. *Dataset:* https://www.kaggle.com/shivamb/company-ipos-2019-2021
    - *Suggested modifications: Compare several companies, analyze a company's stock price from creation to bankruptcy (indicating important events), analyze a major event's impact on many stock prices (such as a bubble bursting, stock market crash), …*
8. **Sales analysis:** Create a visualization (or multiple) using the data from the sale of video games. Suggestions: compare the success of the same game on different platforms, which game genres are more popular in certain countries/areas, is there any developer preference by region etc. *Dataset:* https://www.kaggle.com/sidtwr/videogames-sales-dataset
    - *Suggested modifications: Compare the sales to the overall profitability of the company, compare different video game sales relative to their launch date, …*
9. **Song popularity:** Create a visualization to explain which property (or more likely the sum of) can lead to producing a popular song.
   Dataset: https://www.kaggle.com/yasserh/song-popularity-dataset
    - *Suggested modifications: Use video / book / other media popularity, restrict to a particular geographic region, …*
10. **Data quality:** Take a dataset (or multiple if you want to do comparisons) from the Machine Learning Repository and create a visualization to describe the quality of the data. That is, is the data skewed in any direction, is any class over- or underrepresented, how often and to what degree are samples from the dataset insufficient for usage (if you are looking at multiple datasets). *Source:* http://archive.ics.uci.edu/ml/datasets.php
    - *Suggested modifications: Construct data that could lead to many opposite conclusions, …*
11. **Medical data:** Create a visualization (or multiple) describing how certain factors vary between different diabetes patients. For example, how BMI varies between genders, does the total cholesterol level seem to have an impact on blood pressure, do blood sugar levels change with age, etc. *Dataset:* https://www4.stat.ncsu.edu/%7Eboos/var.select/diabetes.tab.txt
    Technical suggestion: https://scikit-learn.org/stable/datasets/toy_dataset.html#diabetes-dataset

- *Suggested modifications: Use a map visual for geographic differences, ….*

12. **Pokemon:** Create a visualization or multiple, for example, identifying the most optimal Pokemon choice for the highest success rate in battles (you can use one or two generations to make it easier) or determining if there is any connection between the base stats and the physical attributes (size and weight) of the Pokemon. *Dataset:* https://www.kaggle.com/rounakbanik/pokemon
    - *Suggested modifications: Use other games, …*
13. **Ikea furniture:** Inspect the Ikea furniture dataset and determine if there is a relationship between the name of an item (such as length, how far down the alphabet the letters are) and the price of the item. Discuss how a name can imply high cost, and determine how / if that shows up in your visualization. *Dataset:* https://www.kaggle.com/ahmedkallam/ikea-sa-furniture-web-scraping/version/2
    - *Suggested modifications: Compare with country of origin, …*
14. **Political accountability:** Compare how active representatives of the Saeima (deputāti) are in terms of attending meetings, compared with how much compensation they request. Visualize their political party as well. Interpret any trends or outliers, and what that may mean. *Dataset:* https://www.saeima.lv/lv/14-saeima/deputati1
    - *Suggested modifications: Consider a different country, local assemblies, …*