

Project 1

Topics and grading scheme

Current status: *Finalized*

Step 1: Choose from one of the topics below. You may modify it or create your own, in which case confirm with one of the instructors.

Step 2: Choose a topic, from the ones below (you may modify it as you wish), and comment it in the sign-up sheet: [link here](#) or on ORTUS

Step 3: Choose a presentation order (a number between 1-16), and comment it in the spreadsheet as well.

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

- *3 points:* presentation on **February 22**. Criteria:
 - Was the idea clear?
 - Were questions from students / instructors answered well?
- *5 points:* submitted material. Criteria:
 - Is there at least one visual? Is it labeled and titled?
 - Is there text explaining the visual(s) in full sentences? Are all parts of the visual(s) explained?
 - Is there analysis interpreting the results and the context?
 - Have sources been cited?
- *2 points:* Peer review.
 - Average of everyone's grading of your presentation

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, in full sentences, so that someone who did not see your presentation can still understand the main idea.
- Suggested length is 1-2 pages, suggested format is landscape, for better 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.
2. **Comparison:** Find data visualized online (similar to the discussion in the #labs channel in Slack) that you do not like. Create two of your own visualizations from the same data set that improve it. Discuss how the first impression changes and what is emphasized.
3. **Photo analysis:** Take several different photos in your home, on the street, etc. Plot their color distribution on the hue spectrum (see Lab 3). Take pictures of different objects that have a similar hue curve. Discuss how such pictures can be distinguished using some other color-based property.
4. **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.
5. **Animated infection:** Create an animated visualization of how infection spreads among individuals based on exposure time, distance and probability, as the individuals move through environment. *Technical suggestion: create several still plots in matplotlib and combine them into a moving image.*
6. **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>
7. **Open data analysis:** Using open data portal such as <https://data.gov.lv/lv> , choose a data set and create a visualisation for it. Discuss the purpose of your visualization, for example, are you making a point, or exploring the data space. If possible, take several data sets (with the same / similar labels) and compare them.
8. **Making choices:** Use the Forza Horizon cars game dataset to create a visualization that would help choosing a car for various conditions or parameters. Indicate which are "best" in some parameters and compare those separately. *Dataset:* <https://www.kaggle.com/deepcontractor/froza-horizon-5-cars-dataset>
9. **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>
10. **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>

11. **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>
12. **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>
13. **Travel analysis:** Create a visualization of the change in trends over time of taxi usage in New York City. *Dataset:* <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>
14. **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
15. **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>
16. **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>
17. **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/13-saeima/deputati1>