IN40867 - Data Visualization - InfoVis Project

Read this document carefully from beginning to end, before you start!

Overview

The goal of this project is that you carry out a Visualization Design Study. Sedlmair et al. [1] define a design study as *"a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines"*.

In particular, your task is to

- define the data analysis problem that you want to solve and and find corresponding data. This will also include corresponding domain tasks and goals (i.e. you will define the domain from Munzners 4-level analysis framework (Figure 1) yourself)
- design and implement and evaluate an interactive visual analysis system using abstraction, choosing idioms, and implementing the solution.

You have quite some freedom in what you choose for this project. However, remember that not all analysis questions are adequate for a visualization project. For example, if you want to know the mean of a value, the best way to get it is by simply calculating it. This task is not adequate for being achieved through visualization. You can refer to Sedlmair et al. [1] for a set of guidelines for good Design Studies. Another helpful read to get started, with many code examples is *Making data visual : a practical guide to using visualization for insight* by Fisher and Meyer, which is also available as ebook in the library[1].

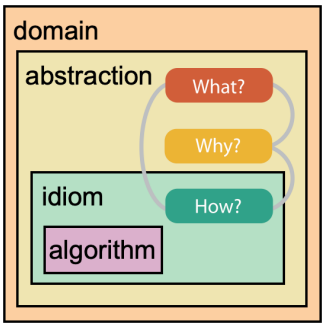

Figure 1: Analysis framework

Outline

- You have until the Christmas break (**deadline Dec. 18 2020, 23:59**) to submit the project. Take note that there will be overlap with the second project in the labs.
- You will work in **groups of up to three students** on the project.
- Deliverables: Group and Individual **Reports**, **Code** and/or Tableau project, **Video** showcasing your project
- We have yet to discuss much the theoretical background for this project during the lectures, but you can start today **thinking about interesting problems and data**. Read on for some tips.

[1] https://tudelft.on.worldcat.org/oclc/1018480478

Implementation

There are two flavors for this project that have some influence on the grading.

1. You can implement your design using an existing visualization tool such as Tableau[2] that allows you to create a complex visualization using small building blocks.
2. You can implement the project with custom code. We recommend Javascript and the D3[3] framework. There will be a brief introduction during the labs.

You are free to use other tools/frameworks but TAs will only be prepared to help with technical issues in Tableau and D3.

Depending on the complexity of the implementation and the visualization methods used, more weight will be placed on the implementation or on the design. The evaluation of your project will follow where you put the emphasis, and we will not strongly penalize your result if there is no implementation, but a good and well justified complex-data visualization design. Be aware that if you employ only simple visualization techniques like Bar Charts in D3, we will not consider your project to have a strong programming component and expect a stronger analysis and justification.

Deliverables

The following describes the deliverables for the final project. Failing to provide **all** listed deliverables will result in a fail!

1. **Report (Group)**: A **concise** report of your project following a typical research paper format, no longer than **4-5 pages** (including references and core images, you can add supplemental images in an appendix, but it should be possible to understand your paper without). We will use the **EuroGraphics format**. A LaTeX(you can for example use Overleaf as a collaborative writing environment) template is provided alongside this assignment in Brightspace. The report should follow the typical structure of an visualization research paper:
    - **Introduction**: Motivation and context of your work (problem description, what you are trying to solve) and contribution (what did you actually do).
    - **Related Work**: literature that is related to what you did.
    - Explanation of your proposed **method/visualization design**: The sections in this part are dependent on your project. Separate theory and the design of your methods from the implementation details. Justify your choices and connect with the related work and initial problem description.
    - **Results and Evaluation**: Present the results that you obtained. This can be images with different data sets, performance tables, etc. Critically discuss and evaluate the results. How good are they? How can they be improved? What are the negative aspects?
    - Conclusion and open/future work.

[2] Tableau: `https://www.tableau.com`
[3] D3 `https://www.d3js.org`

Have a look at published design studies (e.g. [2, 3, 4, 5, 6]) for good examples. [4] and [5] were published in the short papers track in the EuroVis and Vis conferences and are good examples of a condensed 4-5 page paper

2. **Individual Report**: Each person should submit a brief individual report of no more than 200 words, indicating their contribution to the final project. If you did everything together and there is no clear division of tasks then just say so. You can divide the tasks but all of you should be able to explain all aspects of the project, especially the visualization design and justification.
3. **Code, Tableau project** and any results that are not in the report.
4. A short **video**/**screencast** of no more than 5 minutes in which you present your project and what you have achieved. The format can be similar to a regular presentation with slides and explanation, and should include a demonstration of your final visualization and user-interaction. Make your project shine! Narrate your presentation and explain your work such that an audience that has not seen your project or the project details before-hand can follow along.

Assessment

We will evaluate your work based on the following points:

- Problem definition and originality: We evaluate how adequate the problem is to be solved with visualization. How well-defined the questions are that should be answered via the visualization, and how complex they are. Also the originality of the problem and solution will be considered.
- Reasoning/justification of visualization design: The data/task abstraction and how well the made choices are justified based on the principles and methodology seen in class will be evaluated. This is a very important aspect for all projects. For example, justify your choices for the chosen plots, used colors, layout, lines, etc.
- Effectiveness of the visualization design (how well it works): we evaluate the results of your solution. How are the results documented? How well does the application work and how effective is it in answering the original questions. Are there nontrivial findings due to this visualization?
- Technical ingenuity and contribution on the visualization (implementation): How well were technical challenges solved? How ”difficult” was the actual implementation of the ideas? How good is the code? (This is not relevant if you chose to use an existing visualization tool like Tableau and there will be more weight on the other aspects).
- The quality of (all of) your deliverables. How well is the report written? How clean and well commented is the code? How well is the screencast done?

The final mark will be a mix of the points above. The relevance of each point will depend on the focus of your project. The overall effort will be considered too. The grades can be individual, if the work of the team members differ significantly in quality and/or effort.

If there are issues with the collaboration in the group. Please, report this on time.

Get Started

- Choose a dataset and problem that excites you. Think about your interests and make it personal. This will not only make for a fun and unique project it will also increase your motivation. Tip; thanks to GDPR you can request your data from many online services, maybe one of those might be interesting to explore for you. If you read this and the your first idea is to build a COVID19 dashboard; I understand that this is a timely and relevant topic, but please consider the immense amount of work already out there. Think if you can really add something novel. And finally consider your mental health. Do you really want to occupy your mind with the topic even more? In any case, your data set should
  - be complex enough to justify the visualization you propose and the methods that you use for the visualization,
  - and have a certain level of complexity, especially if you decide to not do any implementation (project flavour 1). Use multi-faceted data or enrich with data from multiple sources, present the data on multiple levels of detail, e.g. through aggregation/clustering, etc.
  - Preprocessing will probably be unavoidable, so be prepared to do data cleansing, curation, filtering, etc. Remember that this is a visualization course, though. The focus should be on the visualization, not on the preprocessing alone.
- Derive and document important aspects of the data that could be of interest to an analyst. Formulate a set of tasks that an analyst might want to perform with the data, and some specific questions where visualization can play a major role. For example, develop the analysis based on T. Munzner's method shown in class. Make sure that at least some of the tasks and questions require interaction and/or multiple linked views in order to be performed or solved.
- Consider/design various interactive visualization techniques and combinations thereof (e.g. using linked views) that support these tasks, and that are suitable to analyze this data. Justify your choices and discuss marks and visual-channel encodings (as seen in class). Use the lecture material and other sources if needed to justify your choices. Discuss pros and cons of your design choices.
- Results/evaluation: Go back to the tasks and questions you formulated and use the application you build to make interesting observations about the data. We would expect some none-trivial goals. Document how you came to these observations and how your design or visualization technique was beneficial to your discoveries (or not). Be open and critical here, if you find out at the end that your solution was not optimal stating that and pointing to possibly better solutions is a useful result and will score higher than uncritically stating that you achieved all your goals.

Report guidelines

Note, these guidelines are not exhaustive.

- Do not underestimate the difficulty of technical writing, so reserve enough time for writing the report.
- The paper/report is **not** a manual. The justification of what you did is very important. Do not forget to document it.
- Be precise. It is not sufficient if only you understand what you mean. If the reader cannot understand it, it is usually your fault and not the reader's.
- Use illustrations and screenshots to clarify methods and results.
- Each figure and table should be numbered and accompanied by a caption explaining what the reader sees in the figure or table.
- Refer to figures and tables in the text by using their numbers, for example, "Figure 1 shows...", do NOT use text like " The figure below shows..." . Furthermore, each figure and table must be referenced in the text somewhere.
- Use proper expressions, for example, "don't" should be written as "do not", "it's" as "it is", and so on. The pronoun that goes with "it" is "its" without an apostrophe.
- Spell check, grammar check, and proof read the document before handing it in. Most readers, in particular examiners, will be irritated by poor spelling and poor grammar.
- Do not use material that you did not write yourself. Copy-and-paste without citation, quotation, or reference, is considered plagiarism. Copied material will not be considered as valuable as self-developed content.

Example Data Sources

Please look below for some exemplary online data sources. Take these as a starting point, you are not limited to these sources. Feel free to pick something you are passionate about!

- As mentioned before, consider to request your data from any of the online platforms you use, maybe your WhatsApp history[4] or OV travel history[5].
- Google data search: `https://toolbox.google.com/datasetsearch`
- Kaggle list data and challenges: `https://www.kaggle.com`
- Dutch government open data: `https://data.overheid.nl/`
- Dutch open data repository: `https://opendatanederland.org`
- The Guardian Data Blog `http://www.guardian.co.uk/news/datablog` has a lot of examples and data.

[4] https://www.reddit.com/r/dataisbeautiful/comments/9v5okd/a_healthy_breakup_seen_through_whatsapp_history_oc

[5] https://www.reddit.com/r/dataisbeautiful/comments/9uni2t/oc_my_train_journeys_in_the_netherlands_since_2015/

- Data on publications in the biggest visualization venues: `https://sites.google.com/site/vispubdata/home`
- Worldbank data indicators: `https://datacatalog.worldbank.org`
- A large list of data sets in json format: `https://github.com/jdorfman/awesome-json-datasets`
- A large list of data sets in csv format: `https://perso.telecom-paristech.fr/eagan/class/igr204/datasets`
- r/dataisbeautiful `https://www.reddit.com/r/dataisbeautiful/` can be a nice inspiration, be aware that you will also find many bad examples there

References

[1] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

[2] M. Meyer, T. Munzner, and H. Pfister, "Mizbee: A multiscale synteny browser,," *IEEE Transactions on Visualization and Computer Graphics (InfoVis)*, vol. 15, pp. 897–904, nov 2009.

[3] M. Brehmer, S. Ingram, J. Stray, and T. Munzner, "Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2271–2280, 2014.

[4] C. Nobre and A. Lex, "Oceanpaths: Visualizing multivariate oceanography data," in *Proceedings of the Eurographics Conference on Visualization (EuroVis '15) - Short Papers*, The Eurographics Association, 2015.

[5] N. McCurdy and M. Meyer, "Galstamps: Analyzing real and simulated galaxy observations," in *IEEE Visualization Short Papers*, pp. 276–280, IEEE, 2019.

[6] Y. Zhang, K. Chanana, and C. Dunne, "Idmvis: Temporal event sequence visualization for type 1 diabetes treatment decision support," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 512–522, 2019.