

in4086 Data Visualization InfoVis Project

(Read this document fully and in detail)

Visualization is much about designing creative solutions to the problem of displaying complex data in order to facilitate the user's understanding of it. Most in visualization is learned by practice and applying principles and concepts. We have already introduced several of these principles and concepts during the lectures.

The focus of this project is on the design of visual encodings, interaction and exploratory visual analysis tools for information-visualization (InfoVis) problems (see first lecture). It is VERY important that you have solid justifications of your choices based on what is presented during the lectures (e.g., perception) and to evaluate your work. During the lectures we also present analysis tools that aid to design effective visualizations (What, Why, and How by T. Munzner)

You have a lot of freedom in this InfoVis project and choosing a good topic is part of this project and will be evaluated. Please note that not all analysis questions are adequate for a visualization project. For example, if you want to know the mean of a value, you can just calculate it. This task is not adequate for a visualization solution.

The project can be a combination of the following flavors:

- 1) **A design study**, applying visualization to a domain or domains of your choice using existing tools. This modality requires little programming and corresponds to the type of work that some data analyst do in practice. Spend most time on the design of the visualization and its solid justification. While data acquisition, augmenting from other resources, cleaning data, and aggregating data will be important, these aspects should not be the only focus.
- 2) **A new visualization design or technique**, which involves programming. It corresponds to the type of work visualization experts do in practice. Here, you spend most of the time on building an interactive prototype. This includes the possibility of suggesting modifications to existing visualization technique, if it enables a novel view on a data set.

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 emphasize, 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.

Steps

- Choose a data set and task that you want to solve through visualization (see the end of this document for starting points) . Choosing the right data and task for a visualization solution will be considered in the evaluation. Not all questions about data are adequate for a visualization solution.
 - The data set should be complex enough to justify the visualization you propose and the methods that you use for the visualization.
 - The data has to have a certain level of complexity if you decide to not do any implementation (1), for example, it needs aggregation/clustering. Look for additional sources of data/information to enrich the dataset (use at least one, more is better).
 - Originality is relevant. Preprocessing will probably be unavoidable, so be prepared to do data cleansing, curation, filtering, etc. Notice that this is a visualization course and that is where the focus should be, so be careful not to focus 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 (as 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.
 - Choose an existing information visualization tool, for example a scientific prototype like iVisDesigner (<https://donghaoren.org/ivisdesigner/>), or request a demo license from a commercial package such as Tableau (<http://www.tableausoftware.com>) or IBM Analytics (<http://www.ibm.com/analytics/>). Your choice should be based on the considerations that you made in the previous step.
 - Build an interactive application in any language that you prefer to visualize the data with the chosen techniques. Using [D3](#) is a good option but there are more (e.g., [Bokeh](#)). You should have several techniques in your approach that are linked interactively and are shown simultaneously. One single image and view in the data with no interaction will not be considered positively.
- 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 (or not) to your discoveries. Explain why you designed your approach as you did and report on the effectiveness of your choices.

Work is done in triplets (Create your groups in Brightspace)

Deliverables (submission via Brightspace)

Deadline **January 8th 23:59 UTC/GMT**

Attention: The VolVis project timeframe will have overlap

- **Report** a maximum of **8 pages A4 size** including imaging material where you document: your project problem, and solution. Definition of your problem and justification of your choices is a very important part of your report. For example, justification from a perceptual point of view, and how it takes design principles into account. If needed, provide some support by citing relevant literature. You should also present results and evaluation of your work. You can use the steps mentioned above as a basis to guide the report. The document should use Arial Font 12 pt. The text has to be to the point. So 4 well written clear pages are better than 8 pages with extra redundant content. Do not forget to include images that demonstrate your approach.
- **Source code**, implementation or final outcome (e.g., D3 code, working website, tableau files), depending on your project. If you submit code make clear what you implemented yourself, and what relies on external code, as own code will be judged positively. Make it easy for us to evaluate your work, we cannot evaluate what we do not understand or do not have. Comment the code clearly.
- **Individual reports**: there should be a maximum of 100 words document for each of the group members describing what you did individually in the project. It is important that we know that your partner also agrees with it. It should be in big lines and pointing to the report to explain what was done if needed. 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.
- **Screencast**
Since describing an interactive process is not easy on paper, you should also make a screencast of 3–5 minutes (no longer!) that shows how interaction helps to do some of the tasks. There are a number of tools available to capture either your whole screen or a part of it:
 - Linux: Simple Screen Recorder, <http://www.maartenbaert.be/simplescreenrecorder/>. It does unfortunately not provide video editing facilities.
 - Windows 10: Hit WINDOWS + G to open the Xbox game bar, which has a built-in screen recorder. Click start recording and your current screen will be recorded. The video file can be found in the User/Videos/Captures folder.

- Windows/OS X: Camtasia, <http://www.techsmith.com/camtasia.html> . A fully functional trial version can be used for 30 days.
- OS X has built in screen casting capabilities using Quicktime, via file->new screen recording.

There are multiple video editing tools you should be able to use to generate your screencast. Make sure that in the screencast it is clear what you are trying to show and imagine presenting to a person who has not done the project with you. Just showing the interaction without any explanation is not really helpful so use voice over or annotations.

Report short guidelines (based on Michel Westenberg guidelines)

- Do not underestimate the difficulty of technical writing, so reserve enough time for writing the report.
- 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.
- The report should not read like a manual the justification of what you did is very important so do not forget to document it. This is mentioned clearly in the steps shown above.

How we evaluate your work

We will evaluate your work based on:

- 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 what has been chosen and of the solutions 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 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?
- The quality of your deliverables including all deliverables (report, screencast, source code, etc) is a factor. How well is the report written? How clean and well commented is the code? How well is the screencast done?

Overall: the final mark will be a mix of the points above and the relevance of each point will depend on whether the project focuses on a complex design problem or rather on a technical aspect. The overall effort will be considered too. The grades can be individual, if the work of the team members differs significantly in quality and/or effort.

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

D3 and Assistance

The JavaScript library [D3.js](#) is a powerful library to generate your own visualizations. It has a wealth of possibilities. Therefore, we consider that it is important that people who are working on Data Visualization have knowledge of its existence and are aware of its potential.

During the first two practical sessions you will get an introduction to D3 including exercises. These are meant to give you a good start in the use of D3. The exercises are not part of this project and will not be evaluated. We still recommend that you do the exercises. Even if you decide not to use D3 in your project, you will have learned about a specific tool for Data Visualization that might even be useful for you in the future. It is NOT mandatory to do your InfoVis project in D3.

The **practical sessions** are meant for you to work on the projects. Assistance will be present most of the time.

If you have questions related to the project that you did not manage to get answered during the practical sessions, send an email to: datavis-ewi@lists.tudelft.nl

Example datasets

The web offers many interesting datasets. Here are some examples to get you started but please do not hesitate to look for others. Feel free to pick something you are passionate about!

- <https://toolbox.google.com/datasetsearch>
- The [Guardian Data Blog](#) is chock full of examples and datasets in easily editable formats.
- [Datasets listing](#) of Berkeley's post-graduate visualization course.
- [Google Public Data Explorer](#) has a growing list of data sets.
- <https://github.com/jdorfman/awesome-json-datasets> datasets in json format.
- <https://perso.telecom-paristech.fr/eagan/class/igr204/datasets> collection of data sets in .csv format.
- <http://www.vispubdata.org/site/vispubdata/> publication information in most important publication venues in visualization.
- <http://hci12.cs.umd.edu/newvarepository/benchmarks.php> benchmark data sets in InfoVis.
- <http://opendatanederland.org> Dutch open data.
- <https://data.overheid.nl> open data from the Dutch government.
- 30 places to find Open Data on the web, <http://blog.visual.ly/data-sources/>.
- <http://datacatalog.worldbank.org> worldbank data indicators.
- [Kaggle.com](#) is a website offering a lot of datasets and data science challenges. You might get some inspiration looking at it.
- The subreddit [dataisbeautiful](#) features very nice data projects.
- Since the EU requires large companies to let users download their data, you can also use your own data for your data visualization project. Such as [this project](#) using OV Chipcard data, or [this one](#) using WhatsApp message data.