CPSC 447 23Sep: Design Proposal

Submit a single design proposal (M2) for the whole team as a *PDF* on Gradescope by Wed Oct 25, 6pm.

Submit individual peer evaluation forms for each team member on Qualtrics.

Project Design Report Requirements

When submitting your project proposal as a single PDF, please include the following sections in this order:

1. Basic Info
2. Overview
3. Data and Data Preprocessing
4. Tasks
5. Visualizations
6. Usage Scenarios
7. Reflections (previously: Team Communication Plan)
8. Work breakdown and schedule
9. Credits (new!)

Your design proposal should be somewhere between 5-10 pages in total, including text, images, and tables. It should include 1-5 sketches of your application. You may not need to update sections 1-4 depending on the feedback you received in M1, but you should still read the sections carefully to ensure they are fully up-to-date with the current version of your project. You will be assessed on the quality and clarity of your writing, the feasibility of what you propose, and your initial project description and sketch.

Sections

1. Basic Info
    - Project title
    - A list of all team members and student numbers
2. Overview
    - A few sentences that describe what problem your visualization is tackling and how. You don't necessarily need to create a project that is purely focused on exploratory data analysis. You may choose to tell an interactive data-driven story intended to be consumed by the general public. The

theme of your visualization can draw from any topic, including current affairs, history, natural disasters, and research findings from the sciences and humanities. State the intended audience. Be brief and clear.

3. Data and Data Preprocessing

    - You are free to choose any publicly-available dataset(s). You may find one of these resources useful:

        - Data is Plural (or as Google Sheet)
        - Kaggle Datasets
        - Google Dataset Search

    - In your report, briefly describe the dataset and the attributes that you will visualize. Provide a link to your data sources. Include the cardinality of the entire dataset (total number of items), the types of the attributes, and the cardinality for categorical attributes or the range for quantitative/ordered attributes.

    - When these numbers are large, you can be approximate with terms like "dozens", "hundreds", "thousands": for example, 200K instead of 196,873; hundreds instead of 376. You should differentiate attributes that come with the dataset from new attributes that you might derive for your visualizations. You may find it helpful to include a table, rather than simply relying on text.

    - If your dataset has many more than 20 attributes and you plan to visualize them all, then you may provide a high level descriptor of groups of attributes, for example say the dataset contains *demographic attributes* instead of describing every single variable.

    - Do you need to preprocess or clean up the original data? Do need to join multiple datasets? How do you plan to implement the processing pipeline? (e.g. in a separate tool, such as OpenRefine or Trifacta Data Wrangler; R scripts that are executed once; on the fly in JavaScript).

4. Tasks

    - Provide two concise lists of tasks your future system is intended to address. First create a list of domain-language tasks where each task is closely tied to the description of data above. Then abstract those tasks into abstract-language tasks, as discussed in class. You must articulate a minimum of three abstract tasks pertaining to at least five distinct attributes.

5. Visualizations

    - Building from your usage scenario, give a high-level description of the visualization interface you will build in terms of which components it will include. What are must-have components, without which you would consider your project a failure? Include why you chose your set of components, calling

back to your data and tasks. Remember to be realistic since you are actually required to implement this interactive visualization.

- Your sketch can be hand-drawn or mocked up using Powerpoint, a graphics editor, or wireframe tools such as Balsamiq. Most likely, you will need more than one image to show your proposed visual design and interaction sequences. Don't use more than five sketches. The intent of these sketches is a very basic illustration to communicate your ideas to the marker (so that we can provide feedback about scope and realism) and to help guide your team during the implementation of the first prototype. They should not limit what you end up doing as the final project!

6. Usage Scenarios

   - Provide a usage scenario. Usage scenarios are typically written in a narrative style and include the specific context of usage, tasks associated with that usage context, and a hypothetical walkthrough of how the user would accomplish those tasks with your visualization. The purpose of the usage scenario is to get you to think about how someone else might use the web application you're going to design, and to think about those needs before you start programming. It should also help you assess the plausibility of your identified tasks and the suitability of your chosen datasets to address those tasks. If you are using a Kaggle dataset, you may use their "Overview (inspiration)" to create your usage scenario, or you may come up with your own inspiration.

7. Reflection

   - Describe how your project has changed so far from your abstraction proposal.

8. Work Breakdown and Schedule

   - All team members must be roughly equally engaged in each part of the project. Everyone should be involved in choosing a dataset, creating the abstractions, designing the system, implementing the system, and writing the reports. You **cannot** have one team member do, for example, all the writeups and none of the programming.

   - For this milestone, include a detailed table of work completed so far **and** work that is yet to be completed, split into components such as choosing a dataset, writing the various sections of the report, deciding on domain tasks, abstracting your tasks, or proof-reading the document. For the design and implementation sections, you should also break the components down. For example, you may have one component be the creation of a simple static version of one of your views, the linking of two views together, or dedicated time for bug fixing and refinement for a view. You do not have to use these components exactly, but you should break your work down into similar-sized components.

   - Have each piece of the work be a different row in your table. Indicate which team member(s) worked/will work on it, the date you estimated it to be complete, how many hours you estimated it to

take, the date it was actually completed, and how many hours it took. These last two columns can be blank until the components are actually completed. You must include all of these columns in your table.

- You must include a summary somewhere that indicates the total number of hours used overall as well as the total number of hours per team member. Separately include both the total number of hours used so far per team member/overall and the projected total number of hours to be required per team member/overall.
- The scope of the work should be roughly 50 hours per person (200 hours total across the team of four), for the entire project -- across all four milestones.
- Think about these questions: What are must-have components? What are optional components that can be implemented after the third milestone, if there is time?

9. Credits

    - Indicate any sources of inspiration, including any specific D3 code blocks that you consulted or built upon. Explain what changes you made and their magnitude (e.g. unchanged vs. minor tweaks vs. major functionality additions) for any code that you built upon.

Peer Review

Peer Review: You will be asked to assess the ratio of work done by each member of the team (including yourself). The common case is that everybody is doing roughly equal shares of the work. Sometimes some people are doing a little more or a little less. We particularly need to know if some people are doing much more work or much less work. Please provide additional comments in the free-form text field if you indicate any situation other than the work being equally split. Even if you have mutually agreed to a non-equal work breakdown as of an intermediate milestone (for example, a teammate is ill and plans to do more later), you should still report the actual work breakdown and explain the situation in the comments. We also welcome any comments regarding your experience with the course, overall.

The peer review will take place through a Qualtrics form that you fill out personally. At each milestone, report on the cumulative ratio across the entire term, not just the work since the previous milestone.

The link to the M2 peer review form is: https://ubc.ca1.qualtrics.com/jfe/form/SV_3mxJu981mH6J0A6

Examples from Previous Hall of Fame

Several example usage scenarios:

Best Streaming Service

Usage Scenario

Alice, a mother of two young children aged 5 and 7: Alice is trying to decide which streaming service to subscribe to that would provide the greatest selection of high quality movies for her children. To make such a decision, she needs a way to filter each platform's movies by MPA rating (to find the suitable G-rated movies) and compare the selection by platform and genre. This will help Alice identify the best platform to subscribe to for her children's entertainment. When Alice opens the application, she is able to easily filter the visualizations by selecting an MPA rating of G. She can quickly compare the proportion (as well as the total number) of such movies available by platform. At this point, she may already identify the specific platforms to subscribe to, if there is an overwhelmingly greater number of G-rated movies on one platform over the others. She can also compare the platforms by the number and proportion of high 'quality' or highly 'successful' G-rated movies.

Bored? Games!

Usage Scenario

Bob is a board game enthusiast and loves spending his free time learning about and playing games. He wants an application that will allow him to explore and discover new board games. When he opens the application he wants to see an overview of games available, so he focuses on the graph visualization of all the games in the dataset. To help guide his exploration, he wants to be able to quickly identify which games and themes are the most popular. For example, looking at the bar chart of themes, he might be surprised to see how many games have a Train theme. He decides he wants to further explore that theme, by selecting it as a filter and browsing the remaining games in the graph. As he hovers over a game mark, he's able to see its description, year published, designer, and more. He can also see which 5 other games are most similar to it. While exploring, he is reminded of a game he really likes named Catan. So he searches the dataset in order to locate the game. After finding it, he is able to see where it falls on the complexity scale, if it is considered a highly reviewed game or not, and what games are similar. Additionally, since Bob plays games with different groups of friends and family, he wants to be able to find games that suit a specific occasion. Therefore, he wants to filter the games in the dataset based on the number of players, minimum age, average playtime, theme, and complexity. For example, Bob has a Halloween party to attend. He sees that Horror is a theme that contains quite a few games so he applies that filter. Since the event is a party, he wants a highly reviewed game that is not too complex: he uses the hexbin plot to discover the distribution of games for these attributes and applies more filters to see if any games look good. He enjoys using this tool and decides he'll return when he's in search of a new game to play.

Visualizing Gun Violence in America

Usage Scenario:

American policymakers may want to understand/discover reasons that show correlation to shooting incidents involving gun violence, in order to prevent future incidents from occurring. Policymakers need to explore this dataset to compare which states and regions within a state have more shooting incidents than others while identifying any correlation between the geographical information of each incident and the characteristics of each incident (e.g. police-involved shooting, homicide, etc.). When a policymaker visits the web application, they will see all the views together. They can begin filtering the views by choosing a state through the map view and a particular period of time from the calendar. On the histogram view, they can then select the incident characteristic they wish to explore and compare the values with those from other states. For instance, they may click on “Mass Shootings” and may notice that more mass shootings take place in states where gun laws are looser and if they click on “Officer-Involved Shooting” then they may discover that more police-related shootings happen in neighbourhoods dominated by a specific race. Policymakers can use these insights to create new bills that have stricter firearm laws, institute better mental health outreach programs, or train the police better to resort to the crime of murder less frequently.

Additionally, policymakers may wish to compare distributions of the number of victims across different age groups and compare and identify any correlations on how different incident characteristics may result in different distributions of victims being unharmed, injured, killed, or arrested. A policymaker performs these tasks to see how the nation’s law enforcement is doing in terms of handling gang violence and/or treating police-induced violence situations. They may select “Officer-Involved Shooting” and notice that many of the participant_status values indicate “killed”, which could raise a red flag since officers would likely be treating citizens too rashly. They may also select “Drug Involvement” as the incident characteristic and then select the age group from 10-19 years of age to see how many teens are participating in both drug activity and gun-involved shootings related to such activities. The policymaker may notice that in California, there’s an alarming amount of drug-related shootings involving victims of 10-19 years of age. Upon realizing this, the policymaker may choose to increase support for law enforcement in those areas and increase border security measures to mitigate illegal drug imports and undocumented immigrants.