Final Project: Visualizing a Dataset of Your Choice (25%)

**Duration:** 21 days  
**Assigned:** February 28  
**Proposal due:**March 8 (electronic submissions)  
**Proposal presentation:**March 9 in class  
**Final project and write-up due:** 11:55pm March 20 (electronic submission)

Description

In this assignment, you will apply the data visualization and interaction ideas discussed in lectures, and the skills learned in the previous projects, to a dataset of your choice. You will present your ideas to your peers in a clear and concise manner, and you will demo your project for the instructors. Note that due to the size of the class, you are required to partner with other students to do this project. The team size should be between 3-4. Your final project will take the form of a fully fleshed-out system similar to Project #2, though since this is your second go-around in interactive system design and a team effort, we expect you collectively to do substantially better than what you did in your prior work. See below for a specific list of tasks your interface should accomplish.

Datasets

You are responsible for finding your own dataset this time. Pick something that you're curious about, and that you believe will produce interesting visualizations. The dataset should be rich, containing at least 4 related attributes. You are allowed to use a website from the Opportunity Project website, but may not select any of the prior suggested datasets.

Here are some ideas (although we encourage you to find your own!), along with possibly interesting attributes of each dataset:

* Download your Facebook data (timeline posts, friends list, events, messages): [link](http://www.facebook.com/help/131112897028467/)
* Download a dump of Wikipedia (articles, images, links between articles, edit history, article discussion, user activity): [link](https://dumps.wikimedia.org/)
* Download one or more monthly dumps from Stack Exchange, a question-and-answer website (# of questions in each category over time, # of upvotes or downvotes per question or category, # of answers or comments per question or category, tag frequency, keyword frequency in questions and answers, user activity): [link](http://www.clearbits.net/creators/146-stack-exchange-data-dump)
* Check out the various [multivariate datasets](http://archive.ics.uci.edu/ml/datasets.html) available from UC Irvine's Machine Learning Repository. In particular, the [cars](http://archive.ics.uci.edu/ml/datasets/Auto+MPG) dataset has been used in many high-dimensional visualizations, such as parallel coordinates.
* Download a Twitter dataset. [Here's a dump](http://vis.cs.ucdavis.edu/~iliao/tweets2009-06.txt.gz.gz) of tweets from 6/2009. It contains over 18 million tweets, about 990 MB compressed. The accompanying paper and the network of followers are available [here](http://an.kaist.ac.kr/traces/WWW2010.html).
* A list of datasets mostly from visualization contests: [link](http://web.cs.ucdavis.edu/~ma/ECS289/datasets.html)

Note: previous visualization work has been done on these datasets, and on similar types of data. You should strive to create a new, unique visualization, instead of merely reimplementing what others have already done. This is why it is important to describe your ideas well in the project proposal; we will be able to tell you if your ideas are too similar to previous work.

Proposal

The initial steps in this project is forming a team, selecting a dataset, and formulating a proposal. You will submit the proposal (Word document, ~2 pages, including images) describing the dataset you've chosen, and how you plan to visualize it. What interesting questions can you ask about the dataset? Based on these, you should list three specific tasks you think your system can perform for your dataset. For example, you might want to compare individual data values between each other, or efficiently filter a dataset based on some user input. How will you use visualization to answer them? Specifically, describe how you will use shape, color, size, connections, position, movement, and other visual channels to render different aspects of the data.

Also consider how interaction will play into your visual design. What visualization parameters will be under user control? How can you let your users easily navigate and sift though this dataset? What interaction is important and why?

The proposal will prepare you for the final project, in which you will implement your ideas. That being said, it is expected that your design will change somewhat during the course of implementation.

You must state how tasks are divided among project members.

Proposal Presentation (in-class)

You will give a short presentation (HTML or Powerpoint, ~2 minutes) summarizing the dataset and detailing your visual design and expected interaction. Your presentation should show pictures (hand drawn, computer-generated, etc.) which clearly convey your visual idea. Further, you must discuss your intended interaction and how your complete package will answer "interesting" questions.

You will present your design to the class. Your overall goal is to convince Dr. Ma, Chris, and the rest of the class, that your visualization is good for the dataset and the work allocations are reasonable.

Evaluation Step

Similar to Project #2, your team will perform an evaluation on your developed system. This time however, you will actually conduct a more formal, task-based evaluation by conducting a short user study. For your built system, take the three tasks from your proposal that you designed your system can accomplish. You will coordinate with another team and have its members (at least 3) separately try to perform these tasks using your system. Afterwards, each subject in your study (i.e., the other team members) will fill out a [NASA-TLX](https://en.wikipedia.org/wiki/NASA-TLX) for each task. Plot these results in your final write-up, and include any other qualitative commentary or feedback about your system from the study that you feel is pertinent.

Final Write-up

To describe and document the final design of your project, you will submit a final write-up (Word document, ~4-6 pages, including images). This write-up will be an extension of your initial project proposal; it can include the original proposal, to show what you initially set out to do, or it can further develop the proposal and add more detail. You will also describe your system evaluation here.

The final project write-up should include the following:

* A description of the dataset and its relevance. That is, what you are trying to visualize, and why you want to visualize it.
* How you initially proposed to visualize the data, and why you want to visualize it that way.
* How the visualization system changed throughout the design / implementation process, such as things you tried that just didn't work.
* Justifications and explanations for changes and your team's design decisions.
* A description of the final visualization system, including implementation details, visual encodings, and interactions.
* Your evaluation section.
* How tasks were divided among team members.

Tasks and Grading Breakdown (100 pts)

You will implement the design that you proposed to me and presented to the class. Again, this system should enable us to answer questions that would require exhaustive effort if done with conventional spreadsheet browsing or database queries. This project has much room for creativity. Your objective is to impress me by making a highly interactive and visually appealing system that is useful and provides new insight in the data. The grading for this project is based on combination of proposal, your created interface, your final presentation (demo), and final write-up.

* **10 pts:**Initial Proposal Write-up
* **10 pts:**In-class Proposal Presentation
* **40 pts:**Interactive System
  + Includes both overview and detail view(s).
  + Must include at least 3 visualization techniques used in the system.
  + System visualizations and UI flow are connected (no independent system components).
  + Should display good design choices for user interaction and visualization desgin based on what you've learned in class.
  + Must include at least three of the following features:
    - Temporal, spatial, or data value filtering.
    - Data selection, either single point clicking or lassoing.
    - Include a view that incorporates data clustering, dimensionality reduction, splatting, PCA, etc.
    - Animation
    - Include a visual-based querying mechanism
    - Pop-up widget / tooltip visualization
    - Leverage an online API for data retrieval
    - Interactive re-mapping of color, shape, and/or size for selected attributes
    - Be able to save the state of the visualization and a sequence of user interactions, and be able to go back in time to a saved state (i.e., system provenance)
    - Introduce a novel visualization technique.
* **10 pts:**Final System Demo
* **30 pts:**Final Report