Final Project Description

Purpose

The purpose of the final project is to provide hands-on experience designing, implementing, and evaluating a new visualization method, algorithm or tool. Projects will be carried out by a team of up to three people. Your project should address a concrete visualization problem and should propose a novel, creative solution. The final deliverable will be an implementation of the proposed solution and a 8-12 page paper written in the format of a conference paper submission.

In addition, each group will be responsible for presenting the project to the rest of the class for design discussions. This presentation should describe the visualization problem that the project will address, the relevant related work, and the approach the group plans to take to solve the problem. At the end of the class there will be a final project presentation session so that groups can show their work to others.

Schedule and Deliverables

[Project Proposal](https://canvas.uw.edu/courses/1037158/assignments/3072975) (Due February 8)

As a first step you should create a project proposal that includes the names of the members of your group and a short (1 to 2 paragraph) description of the visualization problem you plan to address.

[In Class Project Presentation](https://canvas.uw.edu/courses/1037158/assignments/3072974) (Due March 7)

In the last week of class, each group will summarize their project for the class. These presentations will occur approximately a week before the final project is due, allowing time for groups to adjust their project as a result of the feedback. The presentation should summarize the project and include the following material:

* Description of the problem and motivation explaining why it is worth addressing.
* A background survey of related work and a list of references.
* A description/storyboard and **demo** of the approach you use to address the problem.
* A description of the results.
* A list of the main tasks comprising the project and a description of what each person in the group has worked on.

**You will have 7 minutes for your presentation. Practice your presentation before the class! Also make sure you bring a laptop to class with your presentation ready to go. To keep things on time, I will have to use a stop-watch to time you presentation and cut you off at the 7 min mark. You will be evaluated on how well you are able to keep on time!**

[Final Paper/Implementation](https://canvas.uw.edu/courses/1037158/assignments/3072973) (Due March 14)

The final deliverables will include:

* **Code:** an implementation of your system (source code and executable).
* **Paper:** an 8-12 page paper written in the form of a conference paper submission. The paper should present related work, a detailed description of your system and a discussion of your design.
* **Presentation Slides**

Implementation

Your implementation should be able to handle typical data sets for the problem at hand, and run at speed compatible with the intended use (for example interactive visualization should run at interactive frame rates). Developing algorithms that scale to large data sets is particularly challenging and interesting. However, the project is not a programming contest and mega-lines of code is seldom associated with a good project.

I am very flexible about the underlying implementation of your projects. You can start from scratch using OpenGL or any other graphics and windowing toolkit, or use an available visualization toolkit. However, the project must include some new code written by your group. You should not simply use existing software such as Excel, Tableau, Photoshop, etc. to create the visualizations for your final project.

Your project does not have to contain an implementation that requires programming. Successful projects may also include human-subject experiments on perception/cognition of visualizations or other kinds of analyses. Please talk to me if you have any questions or concerns about alternative types of projects.

Paper

The final paper should be in the style of a conference paper submission. The paper should include content that is typical of papers that appear at [IEEE Visualization (Links to an external site.)](http://ieeevis.org/),[SIGGRAPH (Links to an external site.)](http://www.siggraph.org/s2014/" \t "_blank), or [CHI (Links to an external site.)](http://chi2014.acm.org/). In particular it should contain:

* **Introduction** - An explanation of the problem and the motivation for solving it.
* **Related Work** - A description of previous papers related to your project.
* **Methods** - A detailed explanation of the techniques and algorithms you used to solve the problem.
* **Results** - The visualizations your system produces and data to help evaluate your approach. For example you may include running times, or the time users typically spend generating a visualization using your system.
* **Discussion** - What has the audience learned about visualization from your work?
* **Future Work** - A description of how your system could be extended.

We have read papers from a number of conferences throughout the course, but if you are having trouble figuring out how to write your paper, take a look at representative papers from the conferences listed above.

Your final paper should be formatted using the 2 column formatting of papers that appear at IEEE Visualization, SIGGRAPH or CHI. Although there are some differences in format between these conferences, you are free to pick from any of these three. If you need help finding a formatting template, talk to me.

Grading

The final project will count for 40% of your final grade in the course. I will consider strongly the novelty of the idea (if it's never been done before, you get lots of credit), how it address the problem at hand, the methodology you employ in doing the research, and your technical skill in implementing the idea.

In small group projects, each person will be graded individually. A good group project is a system consisting of a collection of well defined subsystems. Each subsystem should be the responsibility of one person and be clearly identified as their project. A good criteria for whether you should work in a group is whether the system as a whole is greater than the sum of its parts!

Suggested Project Topics

To get you started I posted a few suggestions for [project ideas](https://canvas.uw.edu/courses/1037158/pages/final-project-ideas). [Edward Tufte's (Links to an external site.)](http://edwardtufte.com/) site is also an excellent place to look for project ideas. His question/answer area is full of ideas that would make good class projects.

Collaboration

If you are interested in finding people to work with, please post on and use the[Final Project Collaboration discussion board.](https://canvas.uw.edu/courses/1037158/discussion_topics/3133006)

Submission Details

**You must join an**[**"Final Project" Group**](https://canvas.uw.edu/courses/1037158/groups#tab-27354)**in Canvas with your fellow group members. If you are working individually, you still must join a group, where you are the only member.**

Please upload your work to an external website, such as your own website or OneDrive. Please make sure you have permissions enabled for URL sharing.

**You must submit your work in two ways:**

1. One person per group must submit via the Canvas assignment submission tool the URLs for the project submissions.

* Final Project proposals should be submitted here: [Final Project Proposal](https://canvas.uw.edu/courses/1037158/assignments/3072975) (Due Feb 8)
* Final Project deliverables (paper, code, presentation slides) should be submitted here: [Final Project Deliverables](https://canvas.uw.edu/courses/1037158/assignments/3072973) (Due March 14)

2. Post the same links to [Final Project Discussion Board](https://canvas.uw.edu/courses/1037158/discussion_topics/3133019). If you are working in a group, please note the names of your group members; only one person needs to submit per group. **At the completion of the quarter, each group should have one discussion post with the following information:**

* Group member names
* Link to project proposal
* Links to presentation slides
* Links to final deliverables: paper and code or executable file

# Final Project Ideas

### Integrating User Comments

Many popular sites that present visualizations and other data (i.e., tables) enable users to comment on the presentation - for example, the New York Times or Guardian have dedicated readers who comment on interactive visualizations published on the site. Comments often refer to parts of the visualization, such as a particular point in a scatterplot or the variable shown along one axis. Design a method or tool for taking a data visualization and a set of comments about that visualization and integrating the comments into the visualization. Imagine, for example, that a user can click on a point in the visualization and see all the comments associated with that point.

### Evaluating Storytelling Visualizations

As we have discussed in class, good ways to evaluate a visualization include identifying whether the visual encodings are expressive and effective, and assessing whether the transformation to the data, such as scaling or filtering, are appropriate. However, an open question is how we can evaluate whether a visualization that is intended to convey a particular story or message has succeeded. Design and implement a tool that a designer could use to determine if their visualization has succeeded in conveying the story or message they want it to convey.

### Visualizing Processes

Many processes, such as sorting algorithms or statistical sampling processes, are easier to understand when visualized. For this project you could identify one or more processes that are difficult to understand (for example, choose a new machine learning technique) and develop a visualization to help people learn about the algorithm. An alternative idea related to visualizing processes would be to find existing implementations of visualizations of algorithms, and try to develop a set of design guidelines and a taxonomy to describe existing visualizations of processes that can help future designers decide how to visualize a new algorithm.

### Label Layout

It is often useful to add text annotations and labels to graphs and charts. The best labeling algorithms use an optimization approach to find a good, overlap free layout for all of the annotations. For this project you could develop a general-purpose labeling tool for optimally placing labels/annotations in a chart created using D3.

### Graphical Perception

We have seen a number of papers in the class that conduct graphical perception experiments to evaluate how quickly and accurately people can decode various types of graphs and charts (e.g. Cleveland and McGill's experiments comparing bar charts to pie charts). However there are many types of graphs and charts for which such experiments have not yet been performed. For this project you could run a graphical perception experiment for a kind of chart that hasn't been investigated yet. The challenges are to identify which aspects of perception to test, designing the experimental methodology and analyzing the resulting data.

### Optimally discriminable colors

When using colors to depict nominal (categorical) data, it is essential to pick colors that are perceptually discriminable. Moreover, it should be easy for users to name the colors so that they can talk about the data in terms of color categories. An effective set of colors will simultaneously optimize both of these constraints. Develop an algorithm for choosing a set of colors subject to these constraints. Color could be constrained in other ways as well. For example, their might be a constraint to enforce color harmony in the chosen set. Or perhaps the colors should conform to natural color palettes. Generalize your algorithm to include constraints like these.

### Perceptual metric for evaluating effectiveness of graphs and charts

We have read and discussed a number of papers on perception of graphs and charts (Cleveland's The Elements of Graphing Data describes summarizes the most comprehensive studies on this topic.) Use as much of the data space as possible to depict data, and clearly show scale breaks are two well known principles for improving perceptual effectiveness. The goal of this project is to develop a quantitative metric for the effectiveness of a given graph or chart. Given a graph or chart (either as an XML or json specification) compute how well it conforms to the perceptual principles outlined by Cleveland and others.

### Automated Table Design

Both Stephen Few and Edward Tufte have described a number of principles for designing more effective tables. Yet, the default table designs in Excel and Latex do not do a good job of highlighting the important information in the table. Develop a system for automatically designing more effective tables of numerical data. You may assume that metadata about the rows and columns is part of your input. For example the data type (N, O, Q), whether or not the data represents a date, financial data, etc. may be assumed as part of the input. The challenge is to operationalize the principles in Few and Tufte to automatically generate more effective tables.

### Family "Tree" Visualization

The visualization and exploration of genealogical graphs (commonly called "family trees") is an incredibly popular activity pursued by millions of people. In a research paper, McGuffin descibes the properties of such graphs and devises a new visualization ("dual-trees") for better facilitating exploration. The approach maintains the basic design approach of most family tree visualizations -- a subset of genealogical relations are depicted as a tree structure, with tree depth signifying one's "generation". While undoubtedly valuable, such diagrams omit other useful aspects of the data. Actual temporal (as opposed to generational) relations are lost, non-hierarchical patterns aren't well depicted, for instance, socially undesirable yet very real occurrences such as divorce, re-marriage, and incest are not well communicated, nor is the geographical distribution or migration of families made clear. One project idea is to design and implement one or more additional visualizations of genealogical data to unearth other important, yet often suppressed, aspects of the data. The visualization could be designed as a self-sufficient visualization (e.g., communicating both lineage and time) or as a multi-component system in which new visual components complement existing techniques.

Visualizing Seattle tech

 The Puget Sound Region is an innovation center.  Bloomberg ranked Washington State as the most innovative in the nation.  This project is meant to shine a light on that innovation and the central role that our key large companies and other entities play in fostering the ecosystem.  This is at its core, an update to a [static poster (Links to an external site.)](http://blog.seattletimes.nwsource.com/brierdudley/WTIAPoster_FINAL.jpg) from 2007 but one that lays the groundwork for an interactive site that offers a visual picture of the innovation in the area and could point to areas of deficit.  The MVP or Minimum Viable Product is a visualization of how the technology ecosystem in our area draws on major employers such as Microsoft, Amazon, F5, EMC, T-Mobile and the UW as the breeding ground for tomorrow’s big companies.  Examples include zuiily whose founder came from Amazon.  Extrahop whose founders came from F5.  Startup Glowforge whose founder sold his first company to Google.

 The broader vision is a wiki like updatable database and interactive site that lets entreprenuers and founders update their information about company sector, size, date of founding, date of going public, revenue, employees, source of technology (aka UW or organization other than the founders), funding by round, etc.  This database would then be available for anyone to use to create visualizations for their own purposes.

 Madrona is planning on hiring an intern to work full time on this project this summer but we want to lay out the data format and the visualization possibilities prior in order to organize the work more effectively.  We are planning on putting some media work behind this image and expect it will be used broadly by us, our companies and appear in the media locally.

 Given the length of time you have to work on this project for class – here are a couple suggestions for how to focus it in order to make it doable.

* Select one company – such as Microsoft, use Linked in data scraping (we can show you how to do this) to pull list of founders who worked at Microsoft, use Pitchbook for company data (a subscription we have and will talk to them about sharing with students for this project) and create a picture of companies that have been started in the last 10 years in our area by people who worked at Microsoft.  Create a visualization of this data.
* Use the Geekwire 200 and other startup lists to trace back the origin of those founders.  Use Mechanical Turk to create data sets (which do have to be cleaned) Create a visualization of this data
* Create a schema for a broad set of data types that you would envision interesting for this project and how a visualization would work.  Help fill in a core set of data as proof.

Our goal here is to create a structure we can build on for the final project which I would like to have completed in August.  Additionally, it is possible that we would hire a summer intern or two out of this project.

 Questions?  Erika Shaffer, [erika@madrona.com](mailto:erika@madrona.com); 206-972-5514 for more on us –[www.madrona.com (Links to an external site.)](http://www.madrona.com/)