

Data Visualization

Final Project: Exploratory Visualization

For the final project, you will be building an exploratory visualization about a dataset of your choosing. Your visualization should expose multiple dimensions of the data to user inquiry in a way that allows them to see important relationships between them. This project will give you the opportunity to combine the various aspects that you have learned about principled visualization design, especially considering what to do with large amounts of multivariate data.

The Data

The dataset you choose to work with is up to you. The most important consideration is to find something that is important to you--that you want to learn more about, and that you think people should be interested in. Your data should also have the following characteristics:

- **Multivariate.** The dataset you choose should feature items that contain multiple (more than three) attributes, including ones of different types (e.g., categorical, ordinal, ratio, etc.).
- **Large-scale.** Your dataset should contain "too much stuff" in at least one direction--too many items, or too many attributes to be able to display all of them at once.
- **Accessible.** The data you use must be free to access for anyone at Carleton.
- **Interesting!** Choose something that you care about, and you think other people will or should care about!

It is worth noting that you do not have to make use of *all* of the data that you find in a particular dataset--i.e., every single attribute and item. Given the scope of many of the datasets you may be considering, some amount of selection/filtering/pruning will be necessary to allow you to focus on a cohesive set of questions. However, the subset of data that you *do* select to visualize should still have all of the above properties.

As you are looking for data to use, there are a number of resources available to you. You may use a dataset from the Assignment 3 Datasets if you would like (though you may **not use the same dataset you used for Assignment 3**). Kristin Partlo in the library also put together an excellent guide for this class:

<http://gouldguides.carleton.edu/cs314>

If you have questions about datasets, don't hesitate to ask. Librarians Kristin Partlo and Ann Zawistoski are both excellent resources for finding data that interests you. Paula Lackie and her Data Squad are also useful resources to help get your data into the shape you need for your visualization.

The Goal

For the last project, you told multiple stories about a single dataset using (mostly) static images. For this project, the goal is to create an **interactive** visualization that offers insight into multiple dimensions of the data.

The design for this visualization is intentionally open-ended. We have discussed many different ways of displaying information in this course, and this is your chance to apply them. However, you must do so in a **task-driven** way. This means that the design decisions you make should all be in service of answering *specific questions* about the data, specifically how different attributes are related to one another. These questions should be complex enough to justify the use of an interactive system--they should not be answerable with a simple query, for instance.

Interaction

With more items and/or attributes than you can reasonably fit into a single image, you will necessarily have to employ some of the reduction strategies that we have discussed--specifically filtering and/or aggregation. As part of your tool, you should include interaction that helps users find detail or explore relationships beyond what you are able to show in one image.

There are many different ways this can be accomplished through interaction. Some that we have seen include:

- Overview visualizations providing details-on-demand (e.g., details about which items are contained within an aggregate, or about attributes that aren't being shown)
- Filtering mechanisms that allow users to select what subset or aggregation they want to observe
- Focus + Context techniques that show a detailed view of a subset alongside a less detailed view of the context around it
- Brushing and selection that allows users to follow individual elements across coordinated views.

The interactions and visual encodings that you choose should be designed (and prioritized!) to afford the task(s) you choose to focus on. For example, if you identify a specific attribute as being the most important for the questions you want to ask, then it probably shouldn't be hidden behind multiple levels of drill-down.

Implementation

You should implement this visualization using D3. If you want to use a different system to create it, you may petition me by Friday (though I will retain the same expectations for interaction).

You may use Flask or other technologies to provide a back-end to your visualization if you would like to, but you are not obligated to do so. It is completely fine for this project to restrict yourself to D3's CSV capabilities.

Evaluation

You will be evaluated in the following categories:

- **Complexity** - How complex are the relationships that you are able to show within your data? Does your visualization allow users to answer different kinds of related questions?
- **Design** - How well have you tackled the challenge of affording insight into more data than you can fit on the screen? Do your choices of reduction techniques fit the data you are working with and the questions you want to ask of it? Do the non-interactive pieces of your visualization follow good good practice as we have discussed up to this point?
- **Implementation/presentation** - How well does your visualization *work*? Note that I am not expecting something perfect in the limited time you have, but your final submission should use D3 to accurately show your data and afford the interactions you design.
- **Justification** - Are you able to thoroughly justify the design decisions that you made in support of a set of questions about the data? How well does your final write-up illustrate this design, and the questions you asked of the dataset?

Deadlines

Monday, May 22: Choose dataset. Send me an email describing the dataset you plan to use and the questions that you want to be able to answer about the data.

Wednesday, May 24: Initial design. As with Assignment 3, you should turn in a PDF to Moodle (here) that shows your design and bring a print-out to class to critique. (Bring at least one print-out per group member, though if you want, you can bring more to share with your critique group.) Your PDF cannot show every aspect of your design, of course, as it is static and your design will feature interaction. Describe the interactions you plan (*briefly*) in words.

Saturday, June 3, 12:00 PM. Final turn-in. You will be turning in the final project on Moodle, in a .zip file. I will provide more detail about the things to be contained in this file closer to the deadline, but you should expect it to include:

- Any code and data needed to run your visualization
- A README describing how to run it, where the data is from, what each file does, etc.
- A PDF write-up describing your design process, the choices you made, how well you were able to answer the questions you wanted to ask, etc.

Partners

I encourage you to work with a partner for this project, as it will give you someone to brainstorm with, debate design decisions, and develop/debug. If you would rather work on your own, that is fine, too. If you want to work in a group of three, you can petition me to do so by Friday. *However*, I will expect more (roughly 50% more!) from a group of three than I will from a pair of two. Also, in your email describing your dataset, you will need to lay out a plan for the individual responsibilities of each member.

If you would like me to help you find a partner, please let me know AS SOON AS POSSIBLE, and definitely by Friday.

Submission status

Attempt number	This is attempt 1.
Submission status	No attempt
Grading status	Not graded