

Lab Assignment 3 – CSE 564, Spring 2011

Due: Monday, April 4, 11:59pm

This lab will introduce you to information visualization. To get you started quickly, we will use a well-established graphical toolkit, `protovis`, <http://vis.stanford.edu/protovis>. `Protovis` uses JavaScript and can be run directly in your web browser. Note, it does not seem to work on IE 9, but it works well on Firefox. First have a look at various examples at the website. Then download the toolkit and documentation.

Upon completion of the assignment please submit the following via blackboard:

- the complete software
- a comprehensive report that illustrates with screenshots, narrative text and code snippets of all aspects of your work
- a link to your webpage so the work can be tested for grading (the data should be there as well)

You will need some data. There are a number of data repositories on the web. Popular are:

- XmdvTool homepage: <http://davis.wpi.edu/xmdv/datasets.html>
- UCI machine learning repository: <http://archive.ics.uci.edu/ml/datasets.html>
- IBM Many Eyes project: <http://www-958.ibm.com/software/data/cognos/manyeyes>

We will explore visualizations for bivariate and multivariate data, hierarchical data, and relational (network) data. To demonstrate each task, you will use (at least) two (2) datasets: one will be provided in this assignment and one you will identify yourself from the repositories mentioned above (or elsewhere). For the latter, describe the data, the reason why you chose it, and give the location on the web.

1. User interface data

First, get the basic infrastructure in place. Perhaps use the bivariate data task (listed in #2 as a test case). Implement a file browser that you can use to select a dataset. Here is a link that has useful information: <http://www.html5rocks.com/tutorials/file/dndfiles>. If you want to see how file reading is done, please check out some sample code provided by Markus Lacay, a former student of this department. The code is here: <http://www.wave255.com/testjs/FormProtovisTest.html>

Now that you can read in a dataset, you want to give users an option on what visualization technique to use. As a general strategy, it is best to create distinct visualizations on separate webpages and have their data be referenced as they are in the `protovis` examples. So, create one master page with a layout table and create several iframes to load its children pages. You would have a list with radio-buttons stating all visualization techniques (scatterplot, treemap, etc.) you support. Then the user would hit a 'Go' button and the corresponding child page would be loaded visualizing the data you just read in. You could format the data into a JavaScript Array so all your JavaScript apps can read it easily.

2. Bivariate data (and user interface setup)

Now, from the UCI repository download the housing dataset. It has 14 dimensions and 506 data items. Your first visualization paradigm is the scatterplot. Use a similar button interface as mentioned above

interface to allow the user to pick two dimensions from the list and plot them in a Protovis scatterplot widget. Don't forget to locate one more dataset for demonstration purposes.

3. Multivariate data

Extend the single scatterplot to a scatterplot matrix where more dimension pairs can be visualized. The user should be able to pick the dimensions shown from the list. In addition, also implement parallel coordinates with brushing capabilities. The user should also be able to interactively determine the order of the dimension axes. Think of a good way in which this could be accomplished. Use again the UCI housing dataset and the one you picked above for demonstration.

4. Hierarchical data

Node-link trees and treemaps are two alternative modalities that can be used for this. The former shows hierarchies better, while the other gives a better sense for proportions. Use the cereal dataset from the XmdvTool page to show various proportions. This dataset contains categorical and numerical attributes. To transform the numerical into categories, you could subdivide them into a number of bins (say three: low, mid, high). Then, for example, using the treemap, it might be interesting to first subdivide by brand (there are 7) and then by calories to see what the proportion of each brand on the market is and what their calorie portfolio looks like. You could then subdivide calories further into, say ratings. Or you could build the tree in the opposite way. Modify the selector so this type of exploration is possible. Show the current tree also in a node-link diagram. Find one more interesting dataset. Also look

What does the cereal dataset look like in parallel coordinates? Choose different colors for the top dividing attributes (say brand).

5. Relational data

Here we shall explore matrix diagrams and pair them with a force directed layout (network diagram). Use the social network of Whitehouse visitor-visitee dataset posted at:

<http://www-958.ibm.com/software/data/cognos/manyeyes/datasets/white-house-visitor-visitee-network/versions/1.txt>

A visualization of this dataset is already available at:

<http://www-958.ibm.com/software/data/cognos/manyeyes/visualizations/white-house-visitor-visitee-network>