

Custom Visual Exploration of the InfoVis Dataset

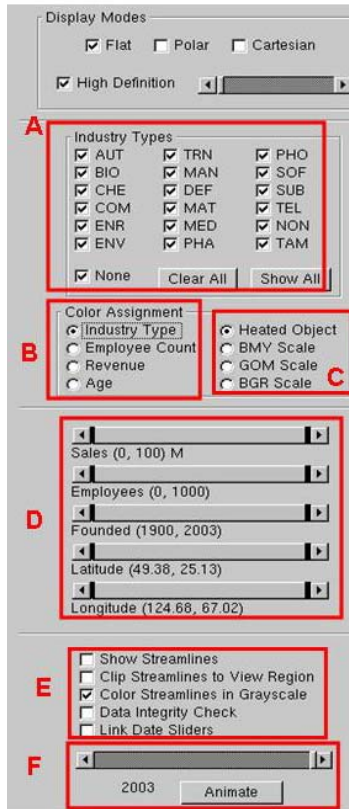
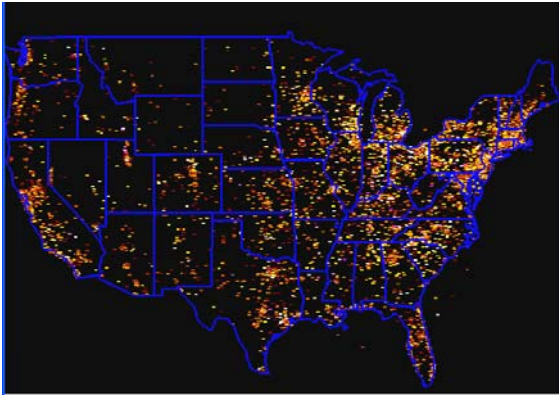
Kyle Hatcher Mike Bailey
Oregon State University

ABSTRACT

Our approach to the InfoVis Data Visualization Contest was to create a custom OpenGL interactive program to explore the data. In the course of using that program, we were able to see a number of patterns in the data that were not originally obvious. We also uncovered some serious inconsistencies in the dataset.

GENERAL DISPLAY

Our general display is shown below. The underlying USA map is drawn using a Universal Transverse Mercator (UTM) projection, to give it a more recognizable shape. Companies are drawn as colored dots which can be filtered based on numerous criteria set with the GUI.



Our GUI is shown here. The GUI is drawn using the GLUI package with our own range slider extensions. In Section A, the user can turn various industries on and off. In Section B, the user can set which variable controls the color of the dots. In Section C, the user can control what color mapping is used to the display the variable selected in Section B. We experimented with a variety of color scales and found we liked the Heated Object Scale (KRYW, used in the USA map above), BMY, GOM, and the rainbow scale (BGR).

Section D consists of five range sliders to filter which companies are displayed. A range slider is special because

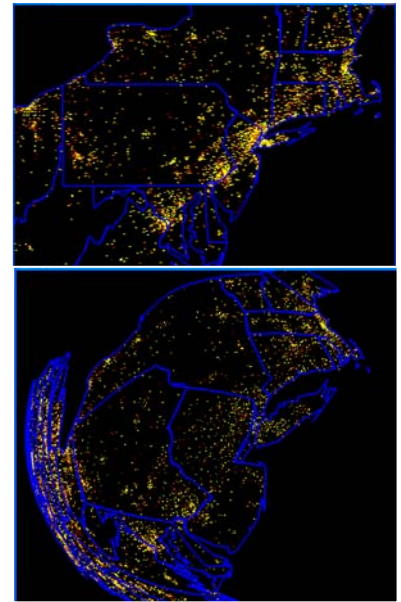
the user can adjust its minimum value, its maximum value, or both together. This is a custom extension to GLUI. A range slider is useful in this sort of visualization because it is often the set of data values within a particular range of minimum-maximum ranges that give the most insight. All five of these range sliders come into play when filtering the company dots.

Section E consists of five checkboxes to enable certain conditions. These will be discussed later. Section F controls which year's data is displayed. The year can be set manually via the slider or can be animated.

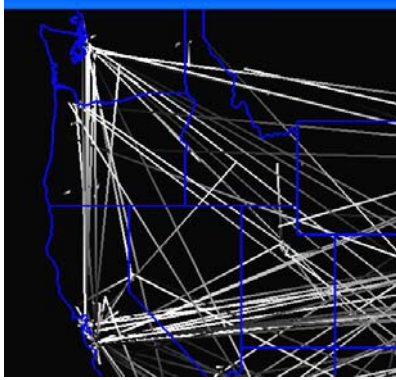
ANIMATION BETWEEN YEARS

When the year is controlled via the slider, the dots are instantly redrawn in their new location with new color values depending on how the data has changed. But, when the year is animated, the dots dynamically travel to their new location. This is very insightful because, when it shows a change in dot density in a certain region, it gives the viewer a sense of how much of that density change was due to companies being created or destroyed, and how much was due to migration.

It was at this point that we realized just how geographically dense the dataset really was. When zoomed completely out, company-rich areas of the country became solid chunks of color. When zoomed in, the individual dots reappeared, but so much of the country was clipped away that we could not tell where the animating dots were coming from or heading to. Our solution to this was to allow the user to switch between Euclidean and Polar Hyperbolic displays of the scene, as shown here. This allowed us to zoom in on any area we wanted to see in more detail, but still be able to see the rest of the country, albeit in a distorted form.

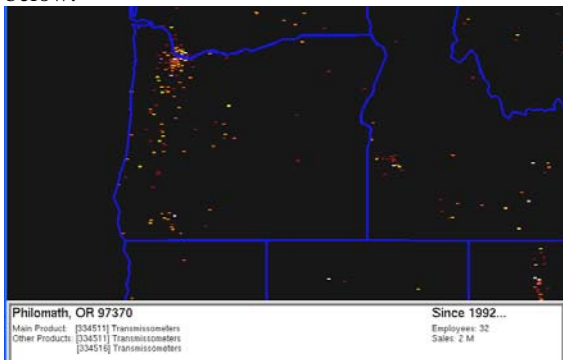


Because the year-animation cannot be shown statically in a paper, we also developed a "streaking" version of the animation, shown on the next page. Each streak shows a company move in the given year from the dark end of the streak to the light end.



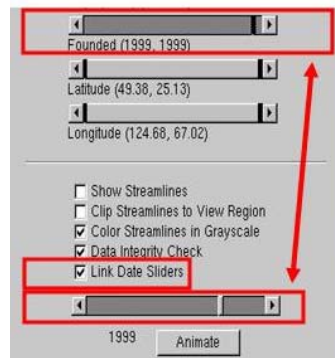
IDENTIFYING INDIVIDUAL COMPANIES

It also became apparent that we wanted to identify particular companies of interest based on their dot behavior. Our program used OpenGL picking to allow the user to select a company dot and see information on that company and its products in the information display area below.

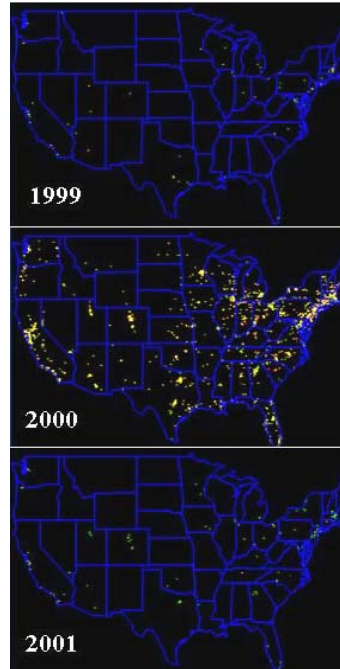


HOW MANY COMPANIES WERE FOUNDED IN A GIVEN YEAR?

We became curious about how many companies were founded in a given year, and seeing if the dot-com bubble became apparent. At first, we just set the year to the most recent (and presumably most complete) set of data and manipulated the “founded” range slider to just show one particular year. But, we realized that this had the effect of undercounting the new companies because many had merged or gone out of business before the most recent set of data was created. To be accurate, one needs to look at a given year’s data and then manipulate the founded range slider to show companies founded in just that year. To make this a one-step process instead of a two-step process, Section E in the GUI has a checkbox to link these two sliders, as shown here.



DATA INCONSISTENCY



With this addition to the program, we went looking for the dot-com bubble. These three images show companies founded in 1999, 2000, and 2001. Big boom or bad data? It does look very out of place, but then again, the dot-com boom was a *big boom*. Upon examining the data, we found that there were many 2000 companies in the year 2000 dataset whose founding data was listed as *after* the year 2000. We added a data integrity checkbox to optionally remove these. This eliminated some of the Year 2000 companies, but not enough to make the data appear consistent with the years 1999 and 2001.

CONCLUSIONS

- Creating a standalone program worked well, as we could add the features we wanted, most of which were beyond the scope of most visualization packages.
- We believe that there are some serious flaws in the data, especially concerning the year 2000. Some companies in the Year 2000 dataset are listed as being founded after 2000. Some companies in datasets prior to 2000 are listed as being founded in 2000. We’re guessing that some blank fields were misinterpreted as “00”.
- One flaw in this data mechanism is that the companies’ locations are their *headquarters*, not necessarily the locations they have the bulk of their employees. Some companies exhibit a strange oscillating back and forth across the country in successive years. Is this a data flaw? Is it due to tax incentive anomalies?
- One interesting application of the current-year slider is after clicking on a company, one can change the date and see the company’s statistics change. It makes it readily apparent how companies grow in size, and how after the boom era ended, there was serious downsizing without much loss in sales.
- We expected to see a predominantly westward migration among all high-tech companies, but were surprised how much eastward migration there was. This was especially obvious during the time animation.
- There is an interesting linkage between computer hardware (COM) and telecommunications/Internet (TEL) companies. Selecting just these two industries, setting color by industry type, and turning on the animation shows the westward migration both experienced during the time frame.