

RECENT DEVELOPMENTS FOR INTERACTIVE STATISTICAL GRAPHICS ON THE WEB USING “NVIZN”

Jürgen Symanzik*, Utah State University, Jon Hurst, Utah State University,
Lacey Gunter, Michigan State University.

*Department of Mathematics and Statistics, Logan, UT 84322–3900,
e-mail: symanzik@math.usu.edu

Key Words: Digital Government, Federal Data, Geographic Display, Micromaps, Tables, WWW.

Abstract

In this paper we present recent developments for interactive statistical graphics on the Web, derived from the Java-based software development kit “nViZn” (read “envision”). We demonstrate how to access and arrange data from Federal Agencies using interactive tables and micromaps. For example, the data can be sorted according to several criteria, different types of displays can be used at the same time, multiple variables can be displayed simultaneously, and dynamic queries can be conducted.

1. Introduction

During the last five years, Federal Agencies and other institutions began intensively investigating the distribution of large, geographically referenced, statistical data sets through the Internet. Such Web-based distribution is aimed at replacing printed tabular displays and at providing access to current data quickly. Several approaches have been developed that provide a user-friendly Web-based interface to tabular and graphical displays of Federal data. The user can interactively and dynamically query the data, compare different geographic regions, and look at the data at different spatial resolutions, e.g., at the state or the county level.

Shneiderman (1999), originally published in 1994, indicates the following advantages and disadvantages of dynamic queries (of data bases): “The dynamic query approach lets users rapidly, safely, and even playfully explore a database. [...] The dynamic query approach is poorly matched with current hardware and software systems.”

In Section 2 of this paper, we summarize how we can display Federal data via interactive tables and micromaps. In Section 3, we describe how the software development kit (SDK) “nViZn” has been used to implement an interactive, hierarchical system for the access of Federal data. We finish with a discus-

sion in Section 4, in particular whether the disadvantages of dynamic queries indicated by Shneiderman (1999) still hold today.

2. Interactive Tables and Micromaps

The distribution of statistical data through printed tables has never been an effective approach for understanding large data sets. Over the last decade, several attempts were made to convert tables of Federal data into more meaningful graphs and maps (Carr, 1994; Carr and Nusser, 1995; Carr and Olsen, 1996). The *Table Lens* (Rao and Card, 1999) is an early interactive tool (first described in 1994) for exploring multivariate numerical and categorical data in tabular form. Current Internet technology allows a user to present interactive tables, graphs, and maps, and to conduct dynamic queries through Web-based applications. These tables, maps, and graphs can be easily rearranged, and users can display subsets and summaries of the data at different geographic levels. This flexibility is very helpful for quickly finding structure in large data sets, in particular geographically referenced data sets from Federal Agencies.

One possible presentation technique for data in a geographic context is the use of linked micromap plots, often simply called micromaps (Carr and Pierson, 1996; Carr et al., 1998, 2000; Fonseca and Wong, 2000). The main idea behind linked micromap plots is to focus the viewer’s attention on the statistical information presented in such a graphical display. Multiple small maps are used to provide the appropriate geographic reference for the statistical data.

Micromap use on the Web was first considered for the U.S. Environmental Protection Agency (EPA) Cumulative Exposure Project (CEP) Web site (<http://www.epa.gov/CumulativeExposure/>). The EPA’s initial goal was to provide convenient Internet access to its hazardous air pollutant (HAP) data for 1990. In this data set, concentrations of 148 air pollutants were estimated for each of the 60,803 census tracts in the 48 contiguous states (Rosenbaum et al., 1999). The Web site was designed to

allow a user to easily navigate the data set to find information on different pollutants at different locations and at different levels of geographic resolution. Unfortunately, no part of the interactive CEP Web site was ever published due to concerns that the 1990 data was outdated at the intended release date in 1998. Only a static version of the CEP Web site without tables and micromaps is accessible. More details on the planned interactive CEP Web site can be found in Symanzik et al. (1999a,b, 2000).

Currently, the U.S. Department of Agriculture — National Agricultural Statistics Service (USDA–NASS) Research and Development Division uses interactive micromaps to display data from the 1997 Census of Agriculture. The USDA–NASS displays acreage, production, and yield of harvested cropland (corn, soybeans, wheat, hay, and cotton) on one of its Web sites (<http://www.nass.usda.gov/research/sumpant.htm>). The user can sort the states by acreage or by yield with respect to the selected crop. However, these micromaps were pre-calculated and stored as *jpg* images. It is not possible to create any new micromap display “on-the-fly” on this Web site. The National Cancer Institute (NCI) is planning to release a Web site in late 2002/early 2003 that provides interactive access to its cancer data via micromaps. This Web site is Java-based and will create micromaps “on-the-fly”.

3. Interactive Tables and Micromaps under nViZn

“nViZn” (<http://www.illumitek.com/>) is a Java-based software development kit (SDK). It is the follow-up to the Graphics Production Library (GPL), described in Carr et al. (1996), developed within the Bureau of Labor Statistics (BLS). nViZn (Wilkinson et al., 2000) is based on a formal grammar for the specification of statistical graphics (Wilkinson, 1999). In addition to the capabilities of the original GPL, nViZn has many extra features. Most useful for the display of data in a geographic context are the capabilities that enable a programmer to create interactive tables and linked micromaps in nViZn. Experiences with nViZn, its advantages and current problems, and its capabilities for the display of Federal data are described in more detail in Jones and Symanzik (2001) and Symanzik and Jones (2001). In the remainder of this section, we will summarize recent additions to our nViZn application that allow a user to display and query Federal statistical data.

About one year ago, our software application only was capable of creating static micromaps and tables from handmade data files containing HAP informa-

tion by state. In addition, it supported a dynamic drill-down map that was an extension of an nViZn example. To display various states and HAPs, a new data file had to be created manually and the nViZn application code had to be modified.

The software is now data-independent: the application code need not be modified to create various graphs. Our application now gathers data from government directory structures (as created for the EPA CEP Web site) with HAP files by state and county, eliminating any need to deal directly with the raw data. As a Federal Agency provides updated data files, the data will be immediately used by our application. It is no longer necessary to create handmade data files.

The created tables and micromaps are now dynamic. The user can sort them in ascending or descending order by six statistical functions. The number of counties that are displayed per micromap can be changed “on-the-fly”.

The individual parts of our application are now integrated to function as a whole. Tables and micromaps can now be created by selecting a HAP and state in the drill-down map. Multiple displays can be viewed simultaneously. The following figures demonstrate some of these extended features.

Figure 1 shows how a screen might appear after our nViZn application has been used for a few minutes. All data in this and the other figures originates from the EPA HAP data set. The upper left drill-down map provides five main features: The “zoom” button allows a user to zoom into a region of interest. When “tooltip” is active, additional information will be displayed for the region underneath the mouse pointer. When “meta” is active, a window with additional meta data will be displayed upon the click of a mouse button. The “table” and “micromap” buttons allow a user to open tabular and micromap displays of a selected HAP at the state level by clicking the mouse in this state. In each of these two displays, the data can be sorted increasingly and decreasingly with respect to minimum, mean, maximum, quartile 1, median, and quartile 3 of the data for the underlying census tracts. In this figure, two tabular displays for benzene in Idaho (sorted by mean) and for acetaldehyde in Pennsylvania (sorted by quartile 3) and one micromap display for beryllium in Utah (showing the minimum) are shown. A meta data display indicates the exact modeled value of acetaldehyde for Philadelphia/PA.

Figure 2 shows modeled acetaldehyde concentrations for Utah. The micromap display shows the median concentrations (with respect to the underlying census tracts) for each county. Immediately

obvious is the gap between Davis and Utah counties. The tabular display is sorted according to the median but also shows the other summary measures with respect to the underlying census tracts. When carefully comparing minimum, mean, median, and maximum, it appears that for most counties the data is approximately symmetrically distributed.

Figure 3 shows modeled benzene concentrations for the neighboring states Nevada and Utah in tabular form, sorted increasingly by the maximum. A closer inspection of these tables reveals that the sparsely populated counties in eastern Nevada (e.g., Lincoln and White Pine) have very similar benzene concentrations to the sparsely populated counties in western Utah (e.g., Beaver, Millard, and Juab). However, for counties with a somewhat larger city in eastern Nevada/western Utah (e.g., Elko/Elko/NV, St. George/Washington/UT, Cedar City/Iron/UT, and Brigham City/Box Elder/UT), only minimum and quartile 1 are related. A drill-down to the census tract level of these counties would reveal that highest benzene concentrations can be found in urban census tracts.

4. Discussion

In this paper, we have shown how an application based on nViZn can be effectively used to display tabular and micromap displays and query information at different spatial resolutions. While the advantage of dynamic queries, given in Shneiderman (1999), certainly holds for our application, we think that the disadvantage no longer holds — software systems such as nViZn are an effective tool for dynamically exploring and querying a data set or a data base. At this point, it should be mentioned again that we are using the hierarchical file structure initially developed for the EPA CEP Web site as our main data source. However, connecting our nViZn application to statistical data bases represents only a small additional step. One other step is still missing: our application is still working as a standalone application and cannot be accessed through a Web browser yet. Finally, it needs to be determined whether the micromaps showing census tracts within counties that have been created for the EPA HAP Web site are immediately useable or still have to go through some finishing stage.

Acknowledgements

Jürgen Symanzik's work was supported in part by the NSF "Digital Government" (NSF 99-103) grant #EIA-9983461 and by a New Faculty Research Grant from the Vice President for Research Office from Utah State University.

References

- Carr, D. B., 1994. Converting Tables to Plots. Tech. Rep. 101, Center for Computational Statistics, George Mason University, Fairfax, VA.
- Carr, D. B., Nusser, S. M., 1995. Converting Tables to Plots: A Challenge from Iowa State. *Statistical Computing and Statistical Graphics Newsletter* 6 (3), 11-18.
- Carr, D. B., Olsen, A. R., 1996. Simplifying Visual Appearance by Sorting: An Example using 159 AVHRR Classes. *Statistical Computing and Statistical Graphics Newsletter* 7 (1), 10-16.
- Carr, D. B., Olsen, A. R., Courbois, J. P., Pierson, S. M., Carr, D. A., 1998. Linked Micromap Plots: Named and Described. *Statistical Computing and Statistical Graphics Newsletter* 9 (1), 24-32.
- Carr, D. B., Olsen, A. R., Pierson, S. M., Courbois, J. P., 2000. Using Linked Micromap Plots to Characterize Omernik Ecoregions. *Data Mining and Knowledge Discovery* 4 (1), 43-67.
- Carr, D. B., Pierson, S. M., 1996. Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps. *Statistical Computing and Statistical Graphics Newsletter* 7 (3), 16-23.
- Carr, D. B., Valliant, R., Rope, D., 1996. Plot Interpretation and Information Webs: A Time-Series Example from the Bureau of Labor Statistics. *Statistical Computing and Statistical Graphics Newsletter* 7 (2), 19-26.
- Fonseca, J. W., Wong, D. W., 2000. Changing Patterns of Population Density in the United States. *The Professional Geographer* 52 (3), 504-517.
- Jones, L., Symanzik, J., 2001. Statistical Visualization of Environmental Data on the Web using nViZn. *Computing Science and Statistics* 33, Forthcoming, (CD).
- Rao, R., Card, S. K., 1999. The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. In: Card, S. K., Mackinlay, J. D., Shneiderman, B. (Eds.), *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, San Francisco, CA, pp. 343-349.
- Rosenbaum, A. S., Axelrad, D. A., Woodruff, T. J., Wei, Y.-H., Ligocki, M. P., Cohen, J. P., 1999. National Estimates of Outdoor Air Toxics Concentrations. *Journal of the Air and Waste Management Association* 49, 1138-1152.
- Shneiderman, B., 1999. Dynamic Queries for Visual Information Seeking. In: Card, S. K., Mackinlay, J. D., Shneiderman, B. (Eds.), *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, San Francisco, CA, pp. 236-243.
- Symanzik, J., Axelrad, D. A., Carr, D. B., Wang, J., Wong, D., Woodruff, T. J., 1999a. HAPs, Micromaps and GPL — Visualization of Geographically Referenced Statistical Summaries on the World Wide Web. In: *Annual Proceedings (ACSM-WFPS-PLSO-LSAW 1999 Conference CD)*. American Congress on Surveying and Mapping.
- Symanzik, J., Carr, D. B., Axelrad, D. A., Wang, J., Wong, D., Woodruff, T. J., 1999b. Interactive Tables and Maps — A Glance at EPA's Cumulative Exposure Project Web Page. In: *1999 Proceedings of the Section on Statistical Graphics*. American Statistical Association, Alexandria, VA, pp. 94-99.
- Symanzik, J., Jones, L., 2001. "nViZn" Federal Statistical Data on the Web. In: *2001 Proceedings*. American Statistical Association, Alexandria, VA, (CD).
- Symanzik, J., Wong, D., Wang, J., Carr, D. B., Woodruff, T. J., Axelrad, D. A., 2000. Web-based Access and Visualization of Hazardous Air Pollutants. In: *Geographic Information Systems in Public Health: Proceedings of the Third National Conference August 18-20, 1998, San Diego, California*. Agency for Toxic Substances and Disease Registry, <http://www.atsdr.cdc.gov/GIS/conference98/>.
- Wilkinson, L., 1999. *The Grammar of Graphics*. Springer, New York, NY.
- Wilkinson, L., Rope, D. J., Carr, D. B., Rubin, M. A., 2000. The Language of Graphics. *Journal of Computational and Graphical Statistics* 9 (3), 530-543.

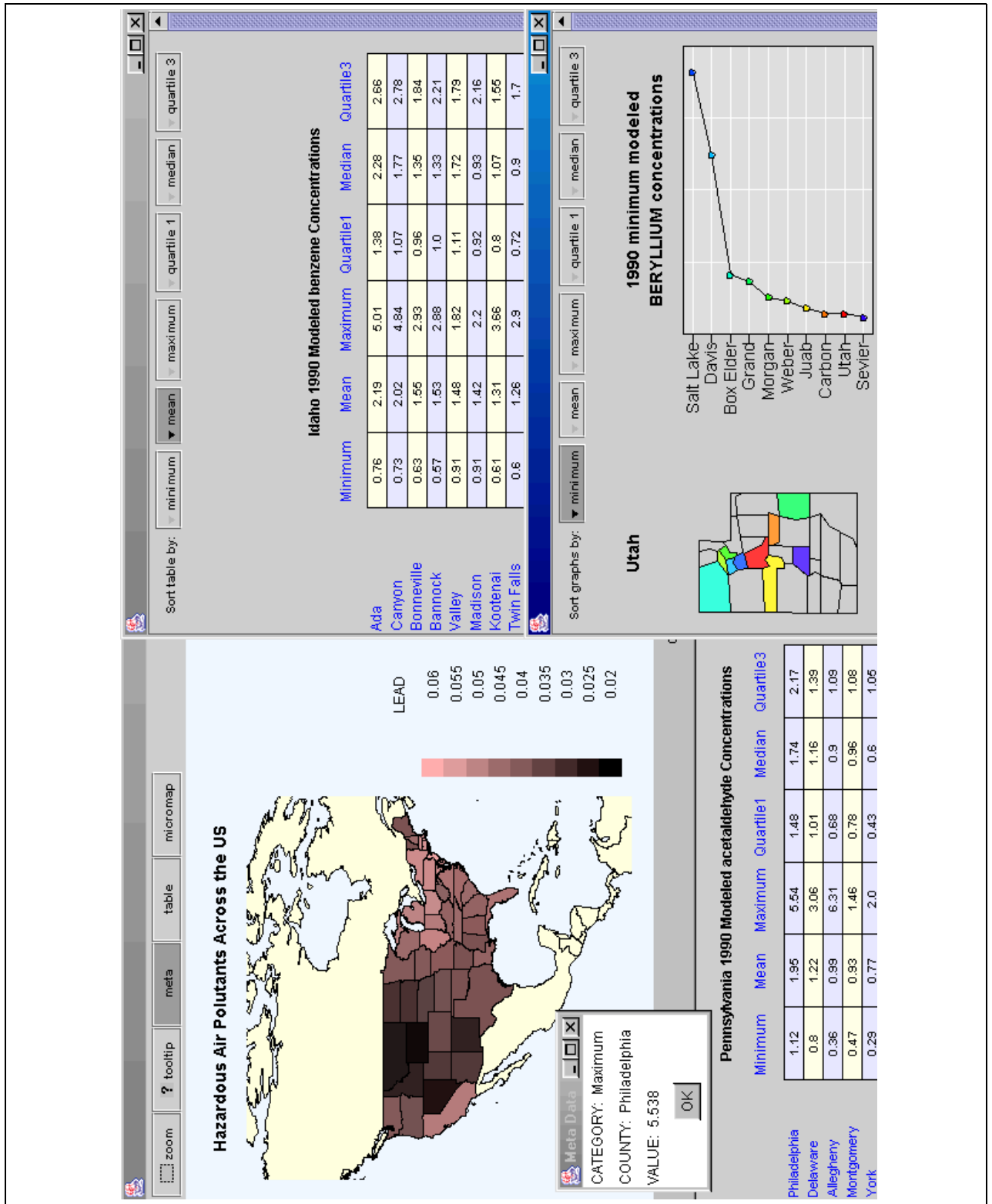


Figure 1: nViZn application, showing a map view, two tabular displays, one micromap display, and a meta data display. Displayed are four different HAPs for different geographic regions.

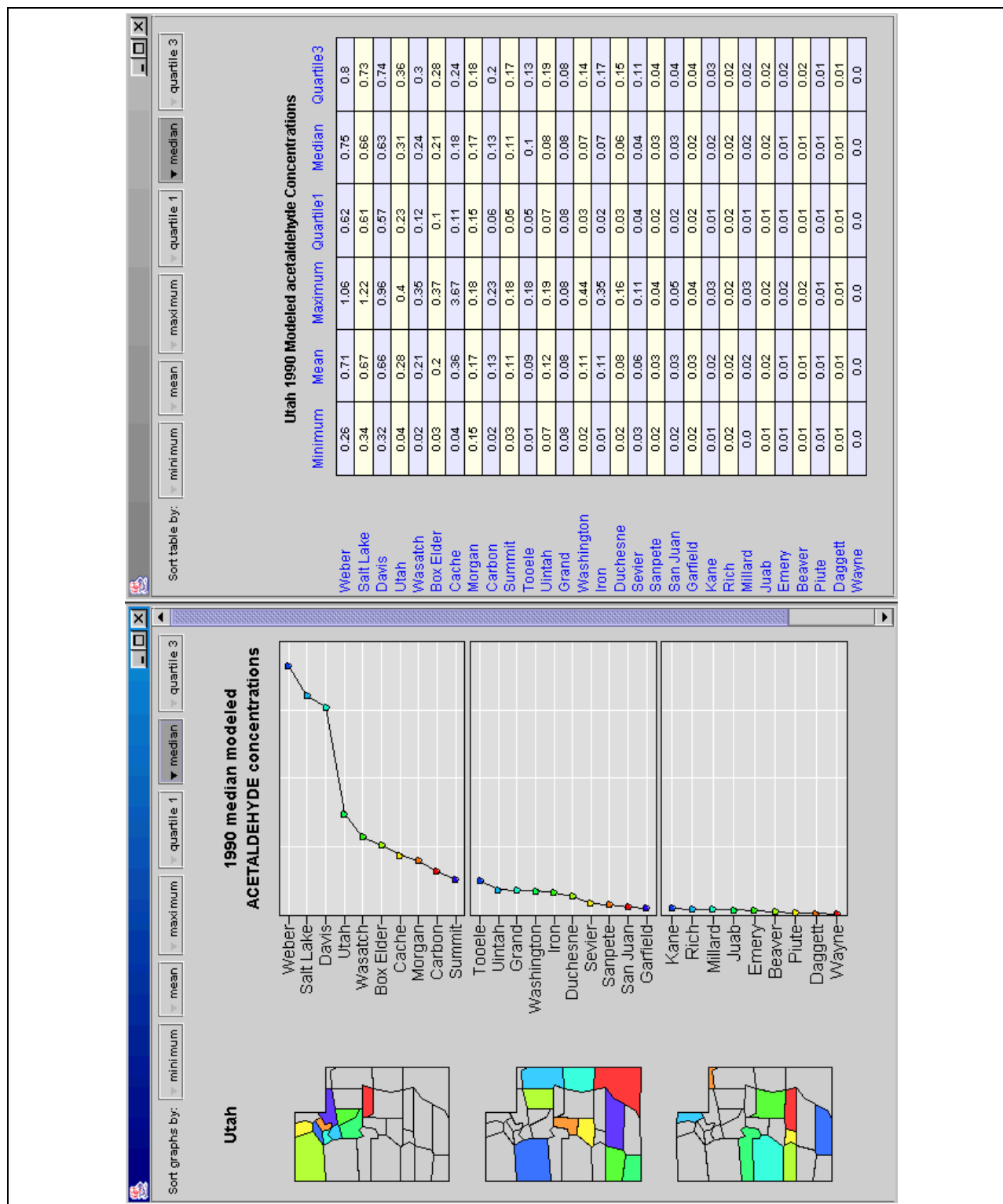


Figure 2: Side-by-side micromap and tabular displays of 1990 modeled acetaldehyde concentrations in Utah, sorted by median.

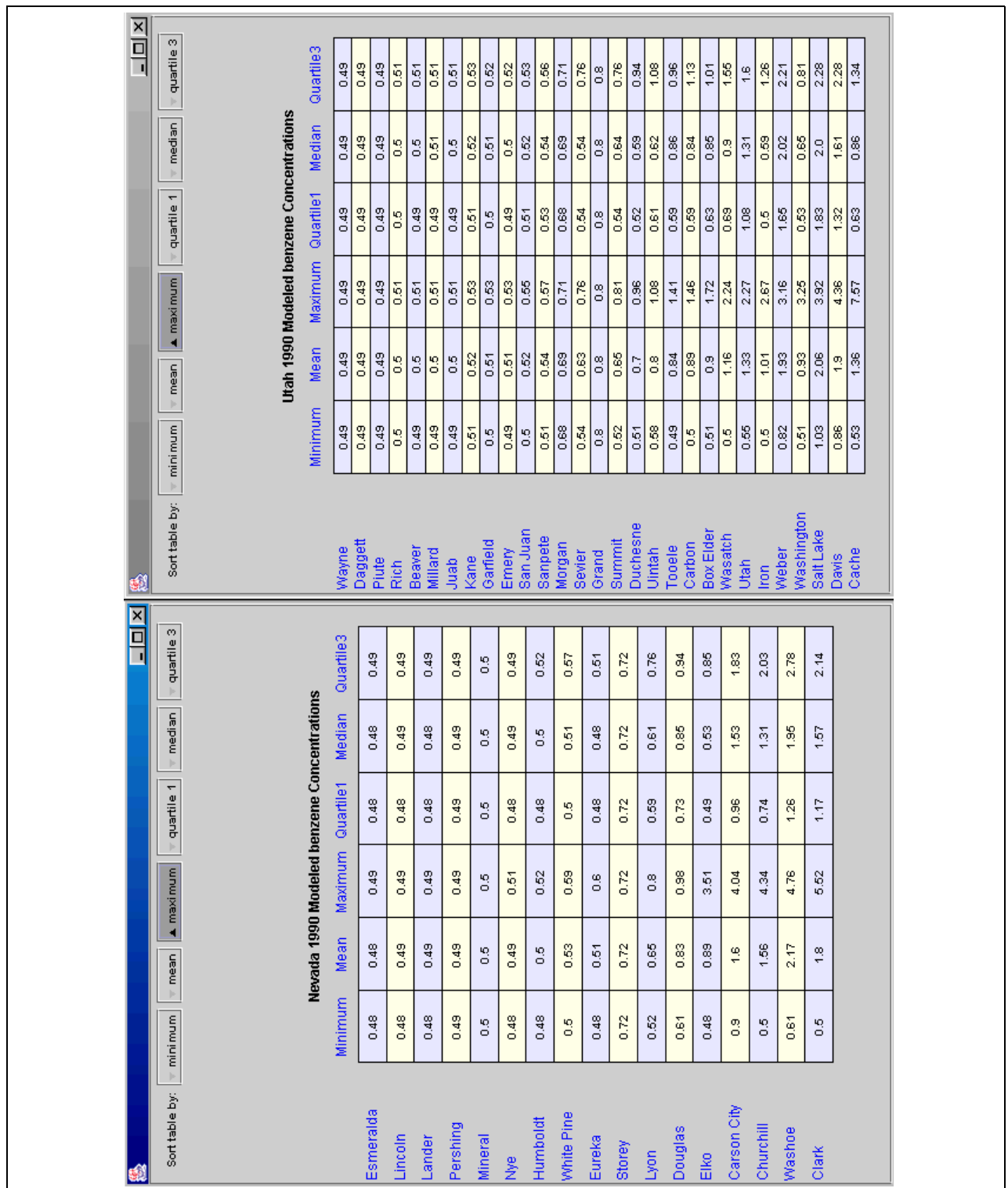


Figure 3: Side-by-side tabular displays of 1990 modeled benzene concentrations in Nevada and Utah, sorted increasingly by maximum.