

Data Exploration with Paired Hierarchical Visualizations: Initial Designs of PairTrees

Bill Kules, Ben Shneiderman and Catherine Plaisant

Department of Computer Science, Human-Computer Interaction Laboratory, and
Institute for Advanced Computer Studies
University of Maryland at College Park
College Park, MD 20742

{wmk,ben,plaisant}@cs.umd.edu

Abstract

Paired hierarchical visualizations (PairTrees) integrate treemaps, node-link diagrams, choropleth maps and other information visualization techniques to support exploration of hierarchical data sets at multiple levels of abstraction. This paper describes several novel applications of PairTrees in the econometric and health statistics domains, as well as some challenges and trade-offs inherent in the technique.

1. Introduction

Coordinated visualizations are an effective way to support exploratory data analysis of multidimensional data sets. Hierarchies are often used to reduce complexity, show structure, and support reasoning at multiple levels of abstraction or aggregation. Our work seeks to integrate hierarchical and coordinated visualizations, taking advantage of the semantics embedded in aggregate, sub-class, containment, and other forms of hierarchy. As part of the NSF Digital Government project “Integration of Data and Interfaces to Enhance Human Understanding of Government Statistics: Toward the National Statistical Knowledge Network”, we are building paired hierarchical visualizations that integrate treemaps, node-link diagrams, choropleth maps and time-series tools.

Data produced by federal statistical agencies often includes explicit hierarchies. For example, within a single statistical dataset (such as a census or survey), geographic attributes are frequently aggregated by state, metropolitan statistical areas (MSAs), or census blocks. Within economic data, businesses are often categorized using the North American Industrial Classification System (NAICS), a 6 level hierarchy used by the United States, Canada and Mexico. Health data often use the International Classification of Diseases (ICD), an extensive hierarchy of disease categories used worldwide for mortality statistics. And an imposed hierarchy can often reorganize data that is not explicitly hierarchical.

Recent research has explored techniques to navigate intersecting hierarchies (Robertson, Cameron et al., 2002), decomposition of tabular data along dual hierarchical axes (Conklin, Prabhakar et al., 2002), methods of zooming visualizations to different levels of abstraction along multiple attributes (Stolte, Tang et al., 2002) and web search that enables users to iteratively construct conjunctive queries using attribute values from separate hierarchies (Hearst, Elliot et al., 2002). GRIDL (Shneiderman, Feldman et al., 2000) provides a two-dimensional display that uses categorical and hierarchical axes to view search results. (Bjork, 2000) describes a hierarchical image browser implementing paired views for text comparison. (Graham, 2001) describes a specific application for comparing taxonomy hierarchies that have a substantial overlap. (Furnas and Zacks, 1994) studied multitrees, a related structure.

2. Applying PairTrees

The following scenarios describe applications for paired hierarchical visualizations that we are building and evaluating. The first scenario applies PairTrees to enhance the immediate usability and comprehension of treemaps. The latter two scenarios support simultaneous multi-level comparisons of aggregate data.

2.1 Enhancing treemap comprehension

Treemaps¹ (Johnson and Shneiderman, 1991) are an effective way to visualize hierarchies of quantitative data, however their structure is often not immediately apparent to users because of the visual complexity of the display. Figure 1 shows US death rates for 43 selected causes of death, a small subset of the ICHS mortality hierarchy. The hierarchy is displayed as a set of nested rectangles, with the innermost rectangles representing leaf nodes and the outermost the root. The size of the rectangle indicates the 1998 death rate per 100,000, and the color indicates the percent change in the death rate between 1981 and 1998. We see by the size of the cardiovascular disease nodes that they remain a significant cause of death, but the green color shows that the rate is declining (e.g., acute myocardial infarctions declined 41%), reflecting advances in this area of medicine. The bright red color of Septicemia and Chronic Obstructive Pulmonary Disease (COPD) & allied conditions show that they increased substantially (up 95% and 86%, respectively) and Alzheimer, with its purple coloring, is shown as an outlier, due to the 1085% increase.

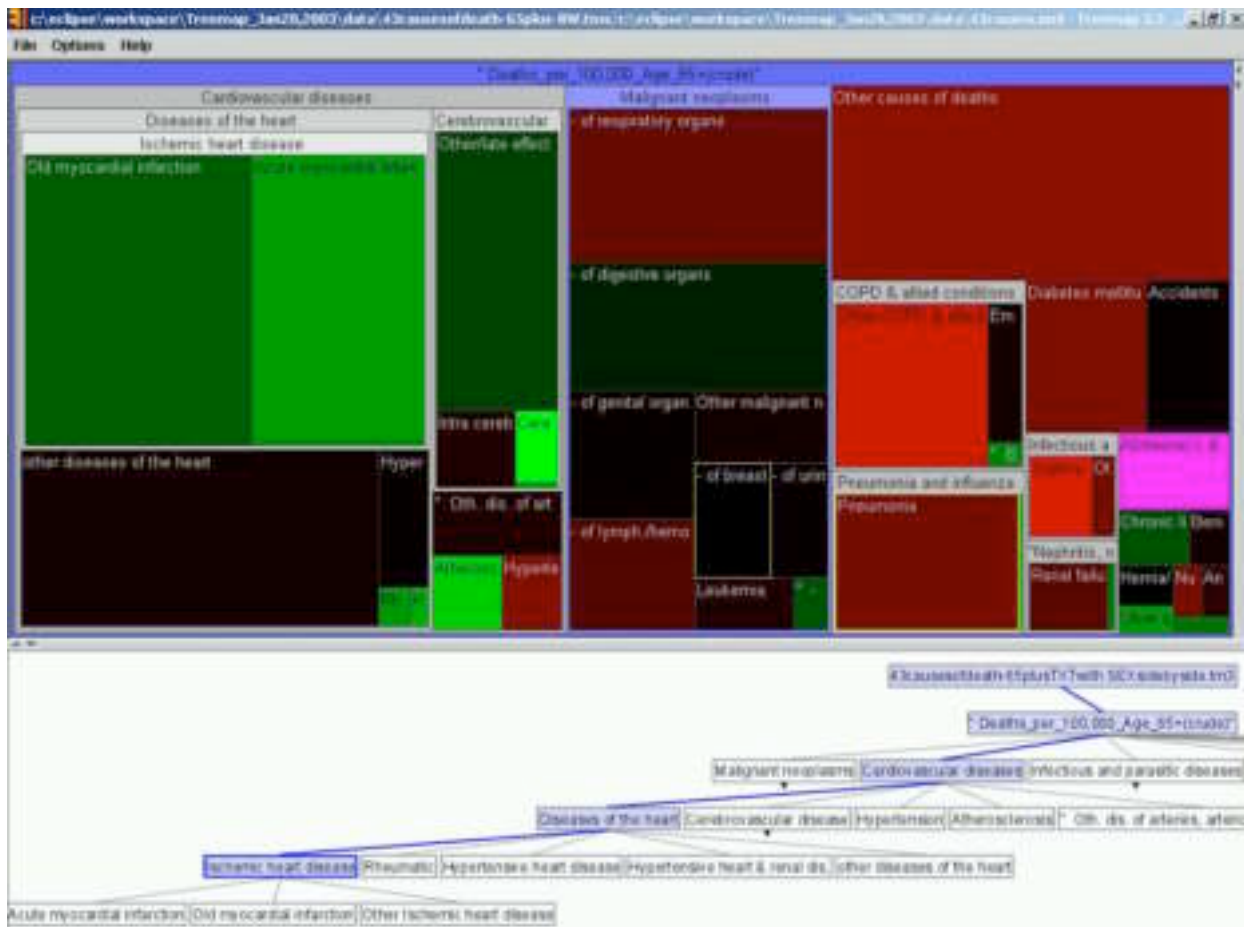


Figure 1. A display of cause-of-death data with the node "Ischemic Heart Disease" highlighted in both views.

To improve the immediate usability of this information-rich display and reduce learning time, the treemap is coupled with a SpaceTree² (Grosjean, Plaisant et al., 2002) to show the same hierarchy, now viewed as

¹ See <http://www.cs.umd.edu/hcil/treemap>

² See <http://www.cs.umd.edu/hcil/spacetree>

a node-link diagram. Node-link diagrams are widely used and clearly show the structure of the hierarchy. There is a one-to-one correspondence between nodes in the visualizations. Brushing the pointer over the Ischemic Heart Disease category in the treemap highlights the corresponding node in the node-link diagram, as well as emphasizing the path back to the root disease node. This clearly shows where the Ischemic Heart Disease category appears in the hierarchy. This brushing and linking technique is bi-directional, enabling users to quickly find a small node such as Hypertensive Heart Disease in the treemap by simply brushing it in the node-link diagram, where it is easily seen. Double-clicking on a cause of death category zooms the treemap display in to that category, while re-focusing the node-link diagram around the corresponding node. Right-clicking zooms the treemap display out a level, leaving the node link diagram unchanged. The simple clarity of the node-link diagram complements the density of the treemap display.

2.2 Visualizing geospatial data with a hierarchical attribute

The previous scenario used two views of the same hierarchy. This scenario and the next demonstrate PairTrees built on separate hierarchies. Figure 2 shows the ICHS disease hierarchy again, this time paired with a choropleth map. The tuples being visualized are (disease, location, death rate). The treemap leaf nodes are size and (redundantly) color coded by deaths per 100,000 (lighter colors indicate a higher value). The choropleth map is color coded by death rate, using the same color scale. The initial treemap view shows US death rates for each disease, while the geospatial map shows state level values aggregated across all diseases. Clicking on a state changes the tuples being displayed in the treemap to reflect values

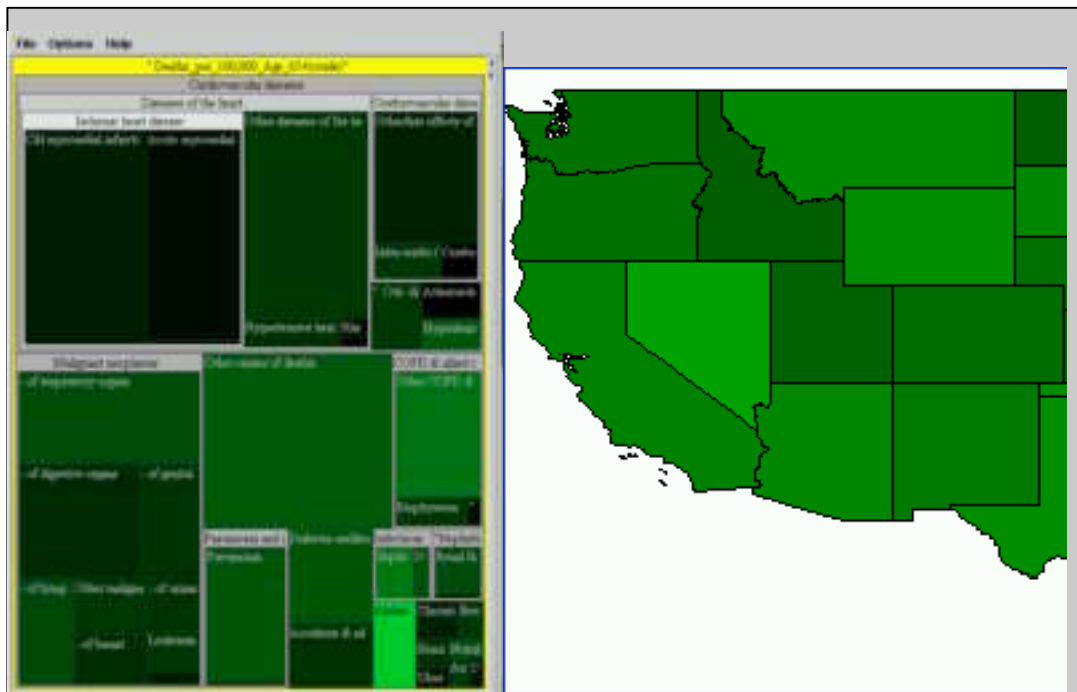


Figure 2. A treemap and choropleth map display death rates by disease and state (mock-up).

for the selected state and zooms the map to that state. Clicking next on a disease or disease category in the treemap will display death rates just for the selected disease (for the selected state) in the geographic map at the more detailed county level. We can thus move up and down each hierarchy to explore the data at multiple levels of detail.

2.3 Visualizing records with multiple hierarchical attributes

Our next example is drawn from the US Census Bureau’s Economic Census. Figures 3 and 4 show revenue data for the “Professional, Scientific and Technical Services” sector (sector 54) of the North American Industry Classification System (NAICS). The left side shows each industry within this sector, size coded by revenue (color coding is not used here). The right side shows the type of service provided, also size coded by revenue, using the “Receipt or Revenue Line” hierarchy. This provides an immediate overview of revenues within each hierarchy. If we select “Management and Consulting Services” in the Revenue Line view, the Industry view is updated to show only those industries earning revenue by from those services. Figure 4 shows the result: The NAICS window now shows the revenue for firms that provide management consulting services. Not surprisingly, management consulting firms provide the bulk of such services, but accounting firms are clearly providing a significant fraction, too – a fact that recent accounting scandals have highlighted.

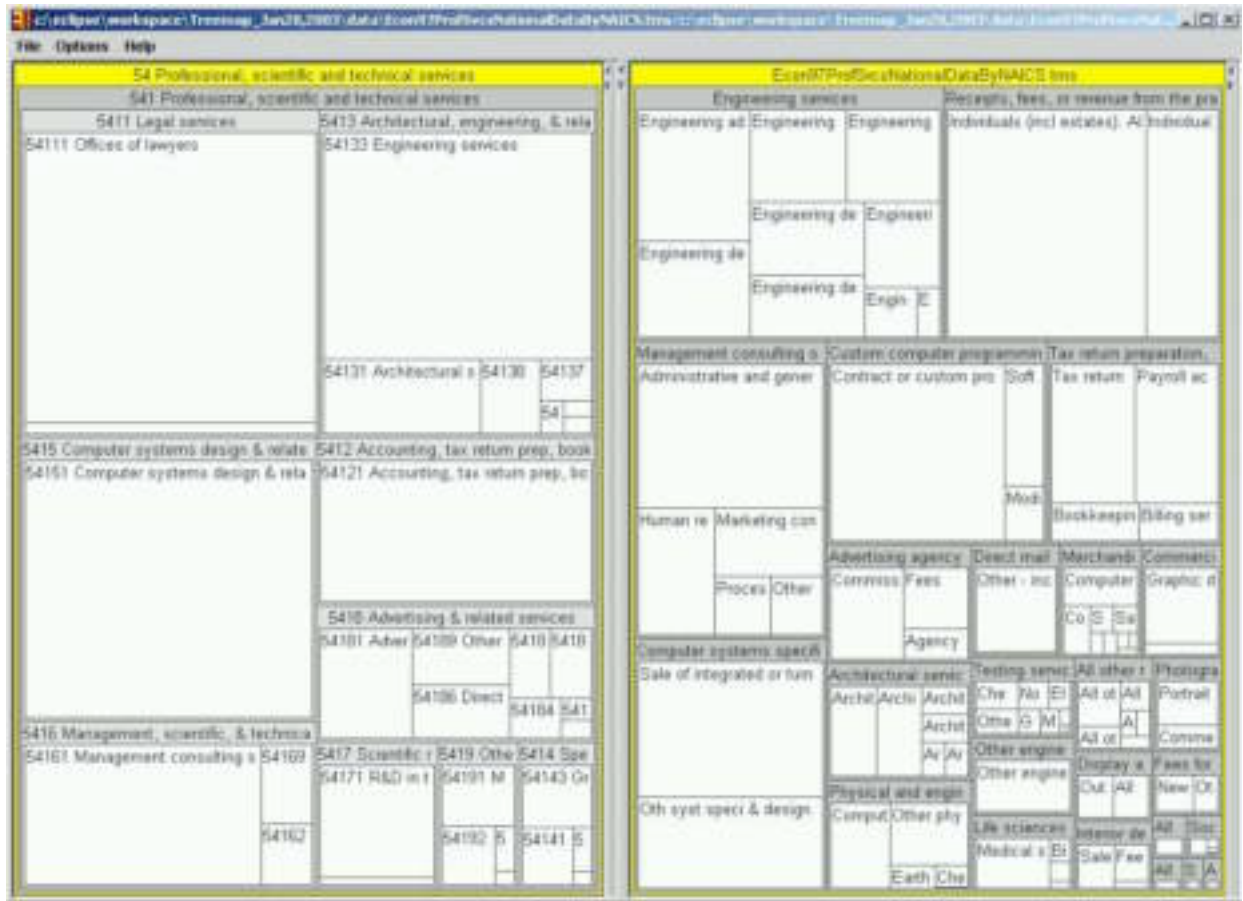


Figure 3. A paired treemap shows an overview of revenues by NAICS code (left) and revenue line categories (right).

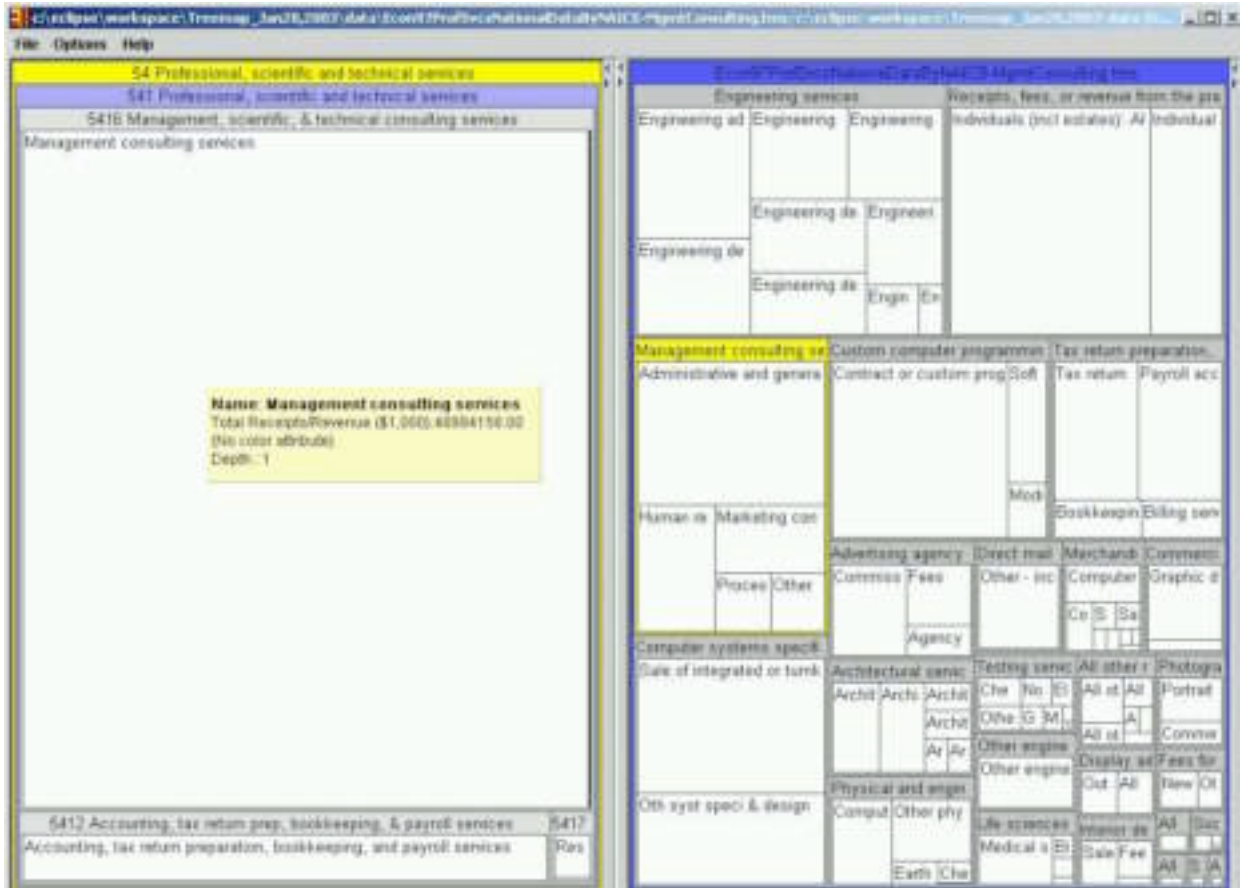


Figure 4. After selecting the Management Consulting revenue line (on the right), the NAICS window is filtered to show just firms providing such services.

3. Conclusion

Paired hierarchical visualizations enable users to explore heterogeneous data sets at multiple levels of abstraction, using hierarchies based on aggregation, sub-class (“is a”) and other relationships.

They can be used to show alternative views of a single hierarchy or separate hierarchies. The records underlying each visualization can be the same or related via a well-defined relationship.

There are several challenges to address when applying PairTrees. As noted in (Baldonado, Woodruff et al., 2000), multiple views add to a user’s cognitive load, and the gain in comprehension must make this trade-off worthwhile. A consistent layout between views is desirable, but not always practical. For example, in figure 1, both the treemap and node-link diagram have a consistent top-down orientation, but this is not always feasible, especially with deep hierarchies. Providing feedback and cues to ensure that users’ mental models are accurate is also a challenge, especially with the more abstract data.

We are continuing to develop the PairTree interfaces to evaluate and refine the presentation and interaction dynamics. We also plan to pair hierarchical visualizations with time series data, for example, market capitalization by sector with historical stock prices. We will continue to apply PairTrees to additional data domains.

Acknowledgements

We would like to thank Jesse Grosjean and Gouthami Chintalapani for their help in understanding the SpaceTree and Treemap code. This material is based upon work supported in part by the National Science Foundation under Grant No. EIA 0129978 (see also <<http://ils.unc.edu/govstat/>>), the US Census Bureau and the National Center for Health Statistics.

References

- Baldonado, M., Woodruff, A. and Kuchinsky, A. (2000). Guidelines for Using Multiple Views in Information Visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces*, Palermo, Italy, ACM Press. 110-119.
- Bjork, S. (2000). Hierarchical Flip Zooming: Enabling Parallel Exploration of Hierarchical Visualizations. *Proceedings of the Working Conference on Advanced Visual Interfaces*, Palermo, Italy, ACM Press. 232 - 237.
- Conklin, N., Prabhakar, S. and North, C. (2002). Multiple Foci Drill-Down through Tuple and Attribute Aggregation Polyarchies in Tabular Data. *Proceedings of IEEE Symposium on Information Visualization 2002 (InfoVis 2002)*
- Furnas, G. and Zacks, J. (1994). Multitrees: enriching and reusing hierarchical structure. *Proceedings of the SIGCHI conference on Human factors in computing systems*, Boston, Massachusetts, USA, ACM Press. 330-336.
- Graham, M. (2001). Visualising Multiple Overlapping Classification Hierarchies (Ph.D. thesis), Napier University.
- Grosjean, J., Plaisant, C. and Bederson, B. (2002). SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. *Proceedings of IEEE Symposium on Information Visualization*, Boston, MA. 57-64.
- Hearst, M., Elliot, A., English, J., Sinha, R., Swearingen, K. and Yee, P. (2002). Finding the flow in web site search. *Communications of the ACM* **45**(9): 42-49.
- Johnson, B. and Shneiderman, B. (1991). Tree-maps: A Space-filling Approach to the Visualization of Hierarchical Information Structures. *Proceedings of the IEEE Visualization '91*. 284-291.
- Robertson, G., Cameron, K., Czerwinski, M. and Robbins, D. (2002). Polyarchy visualization: visualizing multiple intersecting hierarchies. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press. 423-430.
- Shneiderman, B., Feldman, D., Rose, A. and Grau, X. F. (2000). Visualizing Digital Library Search Results with Categorical and Hierarchical Axes. *Proc. 5th ACM International Conference on Digital Libraries (San Antonio, TX, June 2-7, 2000)*. 57-66.
- Stolte, C., Tang, D. and Hanrahan, P. (2002). Multiscale Visualizations Using Data Cubes. *Proceedings of IEEE Symposium on Information Visualization 2002 (InfoVis 2002)*.