Redesigning Parallel Sets

Robert Kosara, rkosara@uncc.edu

1   Introduction

Parallel Sets [2] is a visualization technique for categorical data sets. The main idea is to visualize subsets and subsets of subsets to show the structure of the data, rather than individual data items.

The visualization consists of two main parts (Figure 1, left): the bars that represent the distribution of categories within each dimension (e.g., *crew*, *first*, *second*, and *third class* on the Class dimension) and the ribbons that connect dimensions and represent subsets of sets defined by combinations of categories between different dimensions (e.g., the thin ribbon between *crew* and *female*, showing how few members of the crew were women).

While this method was designed by computer scientists, we did use what design and perception knowledge we had to make it as effective as possible. We used perceptually uniformly bright colors (from Color Brewer [1]), tried to keep the display as clutter-free as possible, and used lines and shading to indicate elements that belong together. We also used color for only a single purpose: to show different top-level subsets.

2   Shortcomings

While Parallel Sets work well, we sometimes found it difficult to communicate the technique to users.

Also, when there are several dimensions in the display with more than four or five categories, the display gets cluttered from the many line crossings, and it becomes difficult to see which subset belongs to which category.

3   Redesign

A first result of an ongoing redesign is shown in Figure 1, right. By reducing the bars to lines, we have moved the focus from the marginal probabilities of categories within their dimension to the connections between them, and thus to the much more interesting multi-dimensional subsets. The redundancy in coloring the bars has been removed, and clutter from the box outlines reduced. Type and text color are also used more effectively.
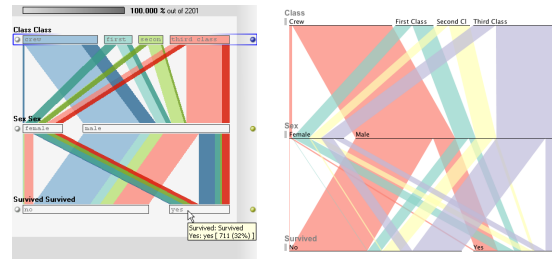

Figure 1: The original Parallel Sets design (left) and the preliminary result of a recent redesign (right).

In addition, the ribbons are aligned so that the new subribbons below a dimension axis seem to be coming directly from the containing ribbon above. This mode existed before, but with the boxes out of the way, the structure of the subsets is a lot more visible. Crossing ribbons are less of a problem in this mode, because they are seen as continuations of larger structures rather than random colored bars.

While we have not tested the new design with users, we are confident that this relatively minor change has made the technique much easier to understand and provides less clutter with the same amount of information shown.

4   Conclusion

There is undoubtedly a lot of room for improvement in many visualization techniques. Small, informed changes can mean a large difference in how a method's difficulty is perceived and how well it communicates the data it was designed for. Design knowledge as well as the design process (like critiquing) provide rich resources for visualization to tap into.

References

[1] C. Brewer and M. Harrower. Color Brewer. http://colorbrewer.org/ (July 2008).

[2] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *Transactions on Visualization and Computer Graphics (TVCG)*, 12(4):558–568, July/August 2006.