Parallel Sets - Critical Analysis

Anudeep Nallamothu

CMSC 636 Data Visualization

Assignment 2

---

1. Introduction

Recent years has seen exponential growth of data collection. We are in a data rich information poor era. Analysis and Visualization are two important things in the conveying information to the user. Creating usable and interactive Visualizations using these huge amounts of data has become a challenge. There are many visualizations like Parallel coordinates, hierarchial edge bundling and Parallel coordinates are some of the visualization methods that can be used to deal with large datasets. In this paper, we will deal with Parallel sets which is quiet useful in dealing with datasets having categorical data. Parallel sets is optimized to deal with categorical data. Interaction is an integral part of Parallel sets data.

2. Analysis of Parallel Sets

2.1. Effectiveness

Parallel sets can deal fitting the higher dimensions of data into the screen space in more practical way. This technique gives power to user by giving interaction with the visualization so that user can see specific data he wants to see.

2.2. Technique

Dataset consists of high dimensional data. In data, each dimension is represented by axis parallel to each other. On these parallel axis, each category is represented by a box and is scaled according to the frequency value of that data. This uses a hybrid of frequency based techniques and parallel coordinates. The interaction scheme helps in manipulating, filtering and changing data according to the user need to gain more information from abstract data. This interaction technique allows user to consider only certain dimensions(columns) in the screen space.

2.3. Limitations

Though this method discusses visualizing categorical data, it doesn't deal with data having a mix of categorical and continuous data which is more common in real world. For example, if data about population in a country is considered, it contains both categorical data like gender as well as continuous data like population in a region. Another point to be noted is, outliers cannot be seen in the visualization. This might arise inaccuracy in the information conveyed.

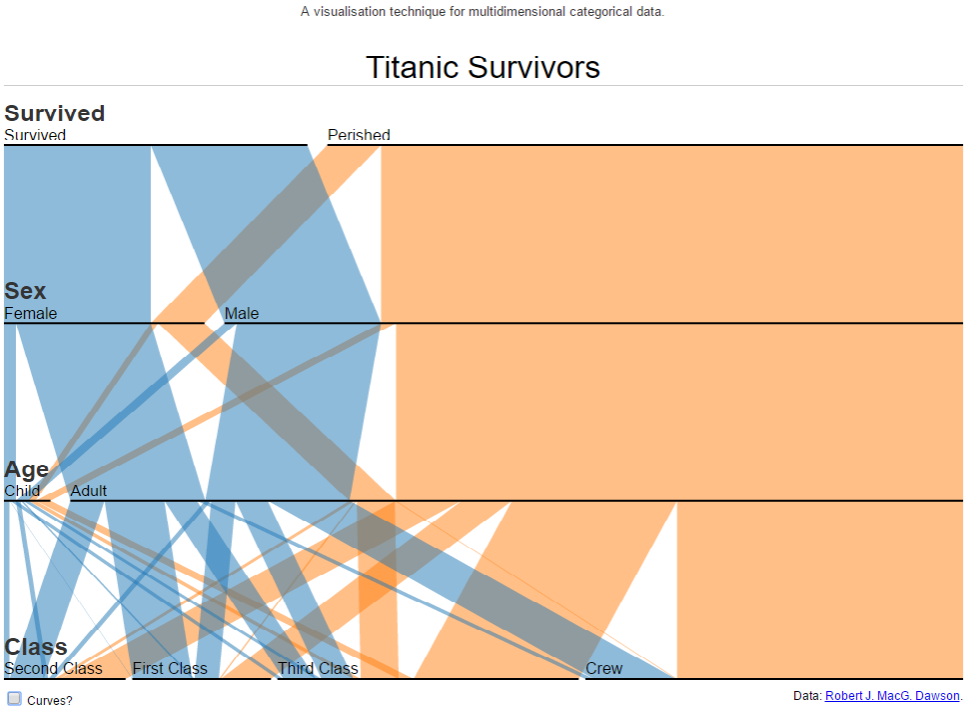

Figure 1: Parallel Sets Visualization

3. Conclusion

Parallel sets addresses the key problems faced in representing categorical data. On a whole, it is used to visualize complex data by giving a interaction workflow as well as visual metaphor of the abstract data.