Survis

Seth Horrigan

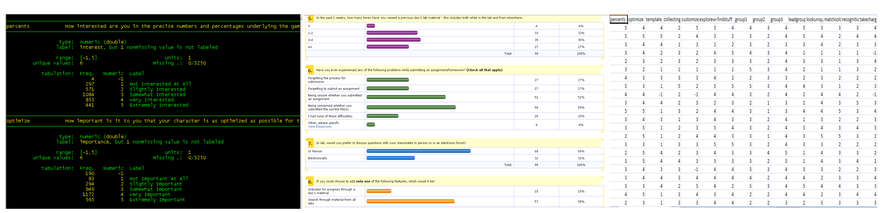
Visualizing comparisons of many ordinal variables

Objective

Visualize many ordinal or categorical variables and their interrelation.

Motivation

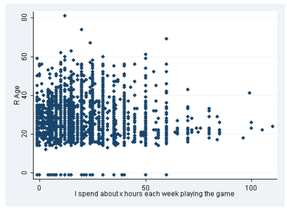
Questionnaire surveys often use banks of Likert-type questions or questions that allow only simple categorical responses. Once collected, the data must be analyzed. Analysis often involves observing distributions of answers and relations between questions. Most often data is represented as tables of numbers. Some better tools will provide bar graphs of each individual question and the distribution of responses.

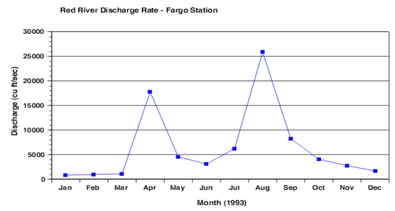


Such methods do not directly provide the viewer insight into variable interrelation. Many are also hard to interpret or only severely limit the number of variables simultaneously viewable.

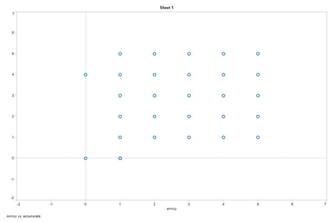
Challenges

Two variables can easily be visually compared using position in flat 2-D space.





Using attributes like shape, color, and size, additional variables can be added to these types of graph. Unfortunately even with only two variables, traditional position encoding like scatterplots and line graphs break down when applied to two categorical variables.



Methods like dimensional stacking, dense pixel arrays, and parallel coordinates visualize high-dimensional interval data but they are difficult to interpret and do not work well with categorical data.

Solution

Nested Category Map

Nested category maps are based on the design of squarified treemaps. Squarified treemaps visualize hierarchical data by grouping hierarchically equivalent items into rough approximations of squares.

Nested category maps apply this idea to ordinal, non-hierarchical data by imposing an artificial hierarchy on it.

This method provides a visual nesting of responses, allowing the viewer to see how each category of response contributes to the next higher level and to the whole.

The implementation provides variable reordering and details-on-demand.

Correlation Map

Correlation maps compose the idea of bar graphs with the structure of scatterplots. Each tile is built as a grid of bars. Each bar's height corresponds to the sum of respondents at the intersection of the values on the two axes.

Tiles are color-coded according to the significance of the Spearman's rank ordering correlation coefficient. A green hue represents significance at p = 0.01, and a red hue represents no statistical significance. The color saturation represents the strength of the correlation or lack thereof.

The map relies on the advantages of small multiples. Each bars is scaled to match similar tiles allowing visual comparison. The implementation also provides details-on-demand through tooltips and secondary detail displays.

Future Work

- Evaluation with potential users for ease of learning, ease of use, and efficacy in identifying relationships among variables and improving perception of the composition of responses.
- Better algorithms for arranging correlation tiles.
- Methods to compose more variables in the same visualization; nested category maps lose effectiveness past $\log_{10}$(number of responses) simultaneous dimensions under normal distributions.
- Optimizing label placement.
- Automatic variable ordering according to value of comparison.

Berkeley Institute of Design