Visual Analysis Tool for Categorical Data – Parallel Sets[*]

Fabian Bendix[†]

VRVis Research Center
Vienna / Austria

Abstract

Information visualization (InfoVis) is the part of computer graphics that provides techniques to make visible information in abstract data. Thinking of a typical data set consisting of hundreds or more variables, there are usually a few or more dimensions that are categorical. This work presents a new technique – called Parallel Sets – to visualize this special kind of data; in contrast to numerical values, a categorical scale is not continuous, but it provides a binning. This technique extends the traditional parallel coordinate system (Section 2) to be able to display categorical values adequately. That means, a new visual metaphor is provided, which handles the discrete feature of categorical data by displaying their frequency values. Additionally, this technique attaches importance to the use of meta information. The provided framework facilitates the management of additional information about the data. By that, it is possible to explore the data set, store the found information and successively refined the gained information.

Keywords: Information visualization, categorical data, meta information

1 Introduction

Visualization is the process of transforming data, information, and knowledge into visual form making use of humans' natural visual capabilities [8]. It facilitates cognition by using visual representations to enhance the detection of patterns and by enabling perceptual inference operations. The data domain traditionally classifies the field of visualization into two parts: scientific visualization (SciVis) and information visualization (InfoVis).

Information visualization mostly deals with abstract, heterogeneous data. In contrast to scientific data (e.g., medical data, flow simulation data, weather simulation data, etc.), abstract data has usually no inherent spatial structures, and visual information extraction is a nontrivial problem for these kinds of data. This field of research tries to find new ways to display abstract data, so that a user can explore the data and look for valuable information.

This work focuses on a small, but important part of InfoVis: the visualization of categorical data.

Categorical Data – apart from the structure of the source data (1D, 2D, 3D, temporal data, tree-based or network data [16]), a basic way to differentiate data is to classify data according to their data scales: quantitative, ordinal, or nominal [11].

A quantitative scale is continuous (e.g., measurements) and the latter two scales are discrete, whereas ordinal values are related to each other in terms of ordering (e.g., school marks), and nominal values usually do not have a natural ordering or distance (e.g., names). Categorical variables are closely related to the latter two scales, because they are also discrete; but moreover, categorical means that the variable is binned (a binning can also be introduced for continuous scales).

Thus, from a visualization point of view, categorical data is very challenging, because there is often no natural way of visually arranging categories, because any graphical relation would imply a relationship between these graphical entities, but categories need not contain such relationships implicitly.

One can differentiate between non-transformational and transformation techniques [14] to visualize categorical data, either the categories are directly mapped to visual attributes, or the categories are mapped to numbers, which then are represented by visual attributes (discretization by similarity-based or frequency-based transformations).

2 Related Work

This section presents some currently available solutions in dealing with categorical data. The number of visualization techniques is still quite limited, but each described method has its advantages and outstanding features. The technique that was most influencing for the design of the Parallel Sets layout is the parallel coordinate system.

Parallel Coordinates [12] are designed to display numerical variables. Originally, the axes of a parallel coordinate system represent the Euclidean n-dimensional space R^n [12]. An n-dimensional point is represented by a poly-line, whose vertices are the intersection points on each parallel axis. Each poly-line represents one data item and each intersection point represents the attribute

[*] http://vrvis.at/vis/research/parsets/

[†] Fabian.Bendix@VRVis.at

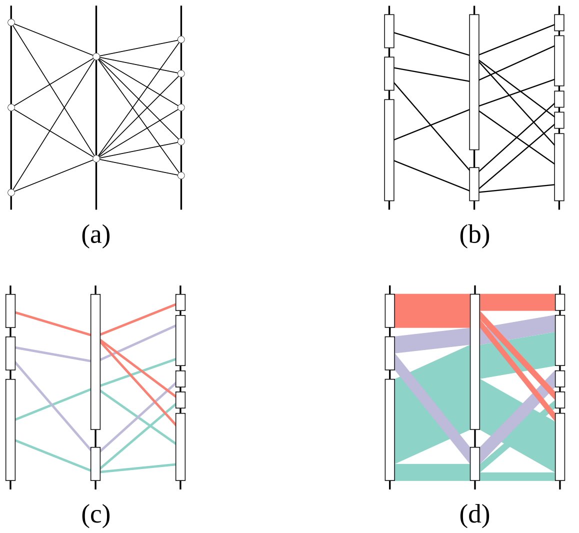


Figure 1: The illustration shows how categorical variables are displayed using traditional parallel coordinates (a), how frequencies can be introduced (b), how color can enhance the visual discrimination (c), and how Parallel Sets accomplish the task (d).

value for a particular data attribute. With the help of techniques similar to Multiple Correspondence Analysis (MCA), one can transform categorical variables to numeric variables [15]. This way, categories are treated as if they had numerical representations and can be displayed by traditional parallel coordinate systems (Figure 1).

The problem with this approach is that the user expects a discrete model, but gets a continuous one. Hence, it would be more natural to visualize frequency information, rather than displaying the data scores (which have to be transformed to numbers).

The Mosaic Display [5, 6, 18] is a recursive space-subdivision technique, in which the frequency values of categories are represented by particular areas ("tiles") on the screen. Alternating the width and the height of a tile is subdivided into smaller parts, whereas the width, respectively the height, of each part represents the relative frequency of the associated category. This way, with each additional displayed variable, the space is further partitioned into smaller tiles.

The weakness of the mosaic display is that for high-dimensional data, the mapping of which tile belongs to which data variable becomes very difficult. Moreover, the mosaic and the parallel coordinate display do not use any additional information to guide the user. The next presented technique shows an approach to do so.

InfoZoom [17] is a commercial software tool that facilitates the analysis of databases. Data dimensions (which are hierarchically structured) are arranged in rows and the data objects are arranged in columns, whereas all neighboring cells with identical values are combined into a larger cell. The width of a cell indicates the number of values of a particular attribute and the column widths are reduced until all cells fit on the screen. In this compressed view, it is necessary that the user can zoom into particular portions of the data. Therefore, the user selects certain cells and only these are scaled to fit on the screen. To be able to navigate between different levels of detail, InfoZoom stores the history information (which *is* meta information that help exploring the data).

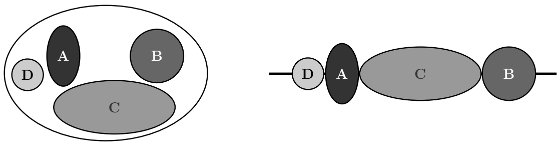


Figure 2: Inspired by Venn diagrams, the used visual metaphor arranges the visual entities on an axis (each entity represents the corresponding category's frequency).

3 Parallel Sets

The contribution of this work consists of three parts: (1) the visual metaphor that finds a natural way of mapping categorical variables to visual entities, (2) the use of additional information to support the user during information extraction, and (3) to provide adequate interaction possibilities and features making exploration possible.

Modified Parallel Coordinate System

Concerning categorical data and their mapping to graphical marks, the goal is to find a good visual metaphor for categories. Two issues have motivated the design of our visual metaphor: the flexibility of parallel coordinate systems and Friendly's statement that areas are natural visual representations for frequency data [6].

Traditional parallel coordinates implement a continuous design model. To integrate categorical variables, the visualization should implement a discrete model to match the discrete user model. Because of that, this approach maps the category's frequency value to the extent of the corresponding visual entity. Then, each dimension is represented by its categories (Figure 2). By using this visual metaphor instead of numerical axes, a discrete design model is implemented.

The idea presented in Figure 1 shows that traditional displays lack important information. Especially when working with nominal variables, the knowledge of how many data objects have a particular data attribute are crucial. Additionally, a line does not represent the relation between two data attributes well enough, because a line only visually says that there is a relation, but gives no hint of how many observations show this relation.

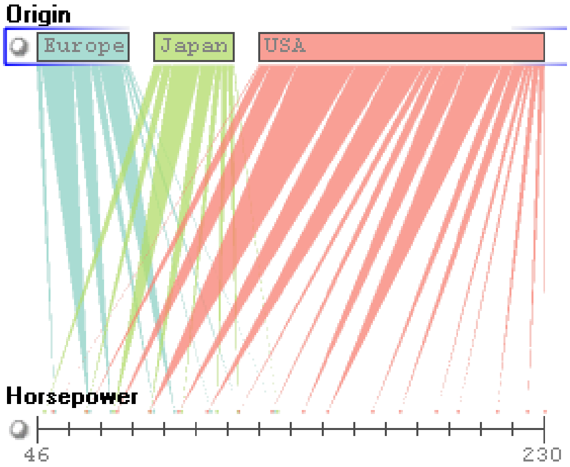


Figure 3: It is shown that numerical variables are also easily integrated, although there actual values are represented and not frequencies.

In the Parallel Sets visualization, dimensions are horizontally aligned, each dimension is represented by its categories, and each category is represented by a box that is placed side by side with the others. The lines in traditional parallel coordinate systems are replaced by parallelograms; the extent of each visual entity reflects the frequency of the associated category, respectively relation.

This technique is designed for categorical values; however, the visualization is also able to display numerical variables. To save main memory and to facilitate a fast rendering, numerical variables are binned. The user can specify the number of bins for each numerical variable individually. Then, for each stream in the view the mean value for each bin is used to draw a triangle from each neighboring category to that value on the numerical axis (see Figure 3).

Knowledge Extraction

In order to deal with large amounts of data, to be able to structure and combine the data and to extract potentially valuable information, one major contribution of this work is implementing the process of knowledge crystallization [3].

Figure 4 illustrates how information can be extracted from abstract data. The important issue is that the process actually is a loop. Thus, visual analysis means exploring the data, storing the found information, again exploring the data and so on. Hence, the found information has to be stored during this process. For that, meta information is needed; it provides information about the information itself. Meta data provides information associated with the raw data (e.g., names, descriptions) and meta data can be used to structure the raw data hierarchically (e.g., in a file system).

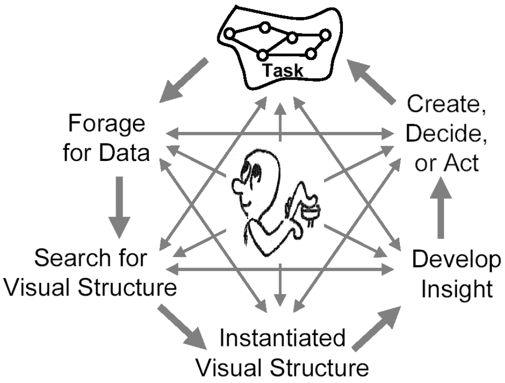


Figure 4: The Parallel Sets technique supports the approved knowledge crystallization loop [3].

Meta Information

In terms of information extraction, it is quite common to use brushing to focus on certain parts of the data that are interesting for the user. Brushing [2, 9, 13] is an InfoVis terminology for selecting portions of the data. A meta language can be used to store brushes or the logical combination of brushes [4]. The saving of meta information during exploration facilitates the reuse of these brushes and facilitates information expansion. Current InfoVis techniques, which focus on the visualization of categorical data, usually lack this feature.

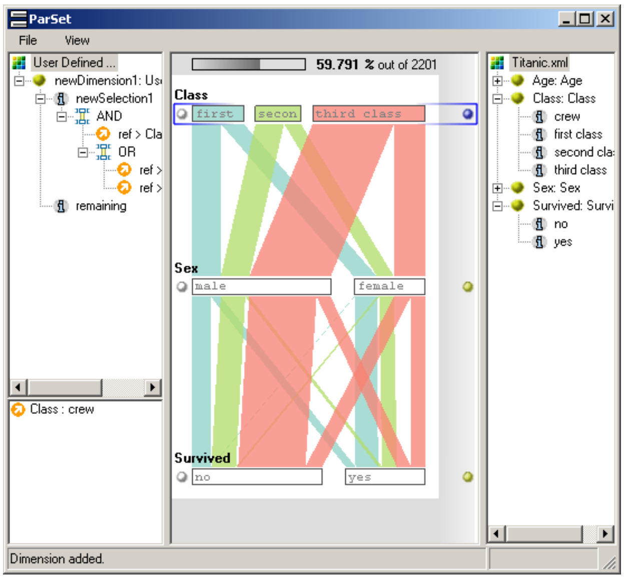
Dimension Reduction

What should be kept in mind when looking at the currently available InfoVis techniques (with regard to categorical data) is the fact that (1) the screen space limits the number of simultaneously displayed dimensions, and (2) human perception abilities limit the dimensionality of the visualization.

In Parallel Sets, a new approach to reduce the number of displayed dimensions is presented. The goal is not to preprocess the data (PCA [7]) and to visualize the resulting lower dimensional space (VHDR [19]), but to make use of the domain knowledge of the user. During the knowledge crystallization process, the user explores the data and by means of brushing, relevant information can be focused. The crucial steps are to be able (1) to store the brushes as categorical dimensions and (2) to reuse these created dimensions over and over again, until the information is found that the user has been looking for. The requirements are:

- each two brushes are mutually exclusive
- each data observation is part of at least one brush

By means of dimension reduction, one can create a new dimension that classifies each data object according to the needs of the user. There are two types of combination: generalization or specialization. Sometimes the provided



Dimension added.

Figure 5: The application screenshot shows the basic layout: user and exclusion panel, visualization panel, and data panel (from left to right).

information is too detailed and a coarser classification facilitates a cleaner visualization. On the other hand, the combined dimension can show all relations of the dimension that are combined.

4 Visualization Design

The application consists of two main parts: the visualization itself and the framework that facilitates the handling of meta information. Exploring data involves the use of meta information – especially when working with categorical data. The separation of visualization and meta information management makes the visualization itself exchangeable.

The framework is composed of four panels (Figure 5): data panel (showing the meta information of the available raw data, the dimensions and the categories in a tree view), the user panel (showing the dimensions the user has created in the same fashion as the data panel), the exclusion panel (providing a list of categories the user temporarily wants to be excluded from display), and the visualization panel.

Visualization of Frequency Data

The information that is provided by the visualization is obtained by a crosstabulation [11]. Statistical examinations deal with categorical data quite frequent; there is always a first look at frequency tables (contingency tables) to get a quick overview. Table 1 gives an example of a two-way table of the *titanic* data set [1]; what is displayed by the visualization is the information obtained by such multi-way tables.

Class	Sex		
	female	male	
first	145 44.6% 30.8% 6.6%	180 55.4% 10.4% 8.2%	325 14.8%
second	106 37.2% 22.6% 4.8%	179 62.8% 10.4% 8.1%	285 12.9%
third	196 27.8% 41.7% 8.9%	510 72.2% 29.5% 23.2%	706 32.1%
crew	23 2.6% 4.9% 1.1%	862 97.4% 49.8% 39.1%	885 40.2%
	470 21.4%	1731 78.6%	2201 100%

🟥⬜ ⬜⬜ f_{ij} is the amount of how many observations fall into the combination of the crosstabulated values of the i-th row and the j-th column (e.g., there are 145 women in the first class).

⬜🟥 ⬜⬜ $r_{ij} = f_{ij}/f_{i+}$ are the individual row frequencies, whereas $f_{i+} = \sum_{j=1}^{n} f_{ij}$ is the marginal row count for the i-th row (e.g., 44.6% of the passengers in the first class were women).

⬜⬜ 🟥⬜ $c_{ij} = f_{ij}/f_{+j}$ are the individual column frequencies, whereas $f_{+j} = \sum_{i=1}^{m} f_{ij}$ is the marginal column count for the j-th column (e.g., 30.8% of the female passengers traveled first class).

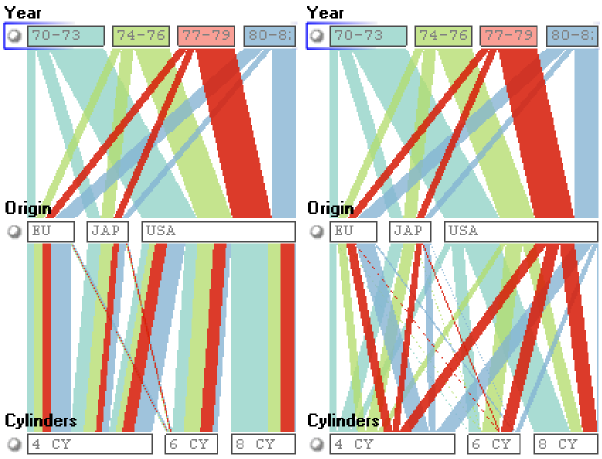
⬜⬜ ⬜🟥 The remaining number is the absolute frequency $p_{ij} = f_{ij}/f_{++}$), whereas $f_{++} = \sum f_{i+} = \sum f_{+j} = \sum f_{ij}$ is the total sum of observations. This value represents the quantum for each combination of *Class* and *Sex* relative to the overall amount of observations (e.g., 6.6% of the passengers are women in the first class).

Table 1: The crosstabulation of the *titanic* data set shows the frequencies for dimension *Class* and *Sex*.

Layout

Into the initially empty view, the user can add axes by dragging dimensions from the user panel and from the data panel into the view and set the position of the new axis. On the right side of the dimension's boxes, a button is displayed. By dragging this button, the position and the ordering of the added dimension can be changed afterwards.

At any point in time, there is one special dimension, the active dimension. The active dimension defines the color-



(a) (b)

Figure 6: The visualization offers the possibility to draw the interconnection (a) sorted (in which all streams between two categories are parallel) or (b) unsorted (the streams *pass through* categories).

coding of the interconnections. Each category of the active dimension gets one color and all passing "streams" obtain the color of the category they pass through. A stream is a group of data records that have identical attribute values in all displayed dimensions. By this color-coding, streams can be differentiated and the streams that pass a particular active category have equal colors assigned.

For rendering the streams, the application offers two modes: sorted and unsorted (see Figure 6), according to the relation of the stream to each other. The sorted mode, which provides a tidy display, renders the streams in a way that all interconnections between each two categories are parallel. The unsorted mode provides the information of how the groups are split by each dimension. Starting at the topmost dimension, there is some amount of data objects, which are also in each of the next dimension's categories.

5 Features and Interaction

The major emphasis of the application is to provide well-suited interaction possibilities. To make use of the user's domain knowledge, the user has to be able to work with a large number of dimensions. Because of that, the key issue is interactivity.

Highlighting is used to provide details-on-demand immediately. For the dimensions and the categories, the names are provided, but one could be interested in the concrete number of data items that belongs to a particular category. Then the user moves the mouse pointer over a category and one second later, a tooltip offers the available meta information for that category. Additionally, all streams that pass through that category are elevated so that

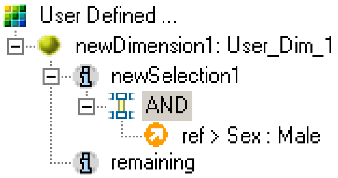


Figure 7: The tree view shows the meta information that is created, if the user starts brushing the data.

they are rendered in front of all other streams, and can be visually traced across all shown dimensions.

Dimension Composition is the process by which the dimensions can be reduced to provide a clearer visualization. If the user starts selecting categories, the meta information, which is displayed in the user panel of the application, shows how the new dimension is composed. As depicted in Figure 7, the user selection creates a new dimension, a new category, a logical operator, the reference, and the default category. The default category is used to satisfy the requirement that each data object has to be part of exactly one brush-category, thus all unselected data items are part of this group. The reference reflects the selection itself and in conjunction with logical operators (that can be modified afterwards), the user can construct arbitrary brushes.

As a next step, the user could either refine the current category or create a new category. Because in this example the operator node is selected, the next selection will be added to the conjunction. If the dimension node were selected, the next selection would form a new category. The categories and also the selections are mutually exclusive. That means the nodes are processed top down and data items that belong to the first category cannot belong to the second and so on.

Category Composition is similar to dimension reduction that is explained above. The user can group certain categories together to form a larger group. This hierarchy is stored in the meta information for each dimension. In the tree view of the data panel, where the meta information is displayed, the user can collapse or expand the group nodes and the visualization reacts immediately on this interaction by displaying either the group or the particular categories.

Hiding Categories means that the user can drag categories into the exclusion panel. All the categories, which are listed there, are skipped during the building of the multi-way table, which is the source for the visualization. This feature is especially useful in conjunction with the dimension reduction feature, because the user can build his own dimension and all uninteresting data items can be summarized in one category. If the dimension is

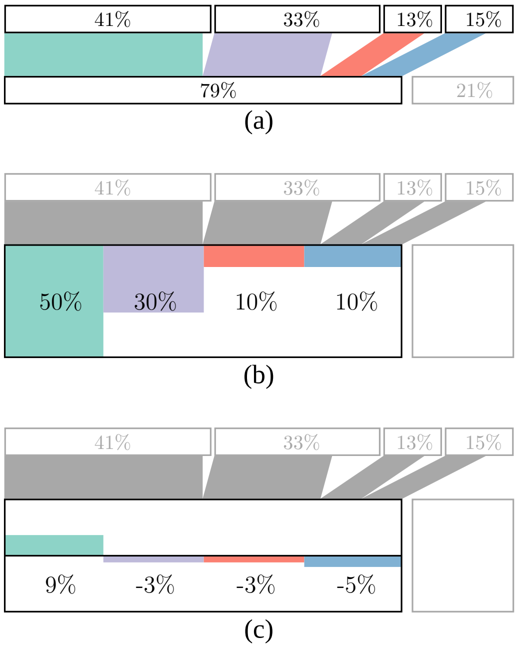


Figure 8: The illustration shows that the interconnections represent absolute frequencies (probabilities) (a), histograms can be used to show relative frequencies (conditional probabilities) (b), and the differences between conditional and unconditional probabilities indicate a degree of independence (c).

displayed, the uninteresting category is dragged into the exclusion panel and these values are no longer displayed in the view.

Reordering means that the position of the dimensions and of the categories is not fixed. At any point in time, the user has the possibility to drag a dimension to some other position and by that, the ordering of the axes can be changed. The same is true for categories: with drag and drop, the user can order the categories according to his needs.

6 Optimizing the Visualization

In contrast to the techniques that are presented in Section 2, the Parallel Sets visualization is a very flexible display in which screen space is still available. The height of each category's box can be enlarged without disturbing the visualization, to gain additional space. There is room to draw a plot that presents special insight. In this work, histograms are drawn inside each category, if the user enlarges the box (similar to [10]). Inside each box, two self-contained histograms are displayed: the top histogram shows the above dimension, and the bottom histogram shows the dimension below – relative to the dimension the histogram is displayed for. Thus, the leftmost bar in the histogram represents the leftmost category of the neighboring dimension; the second bar represents the second category and so on.

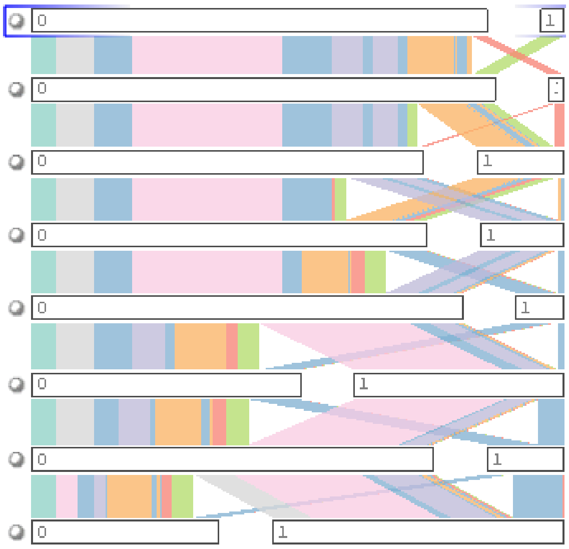
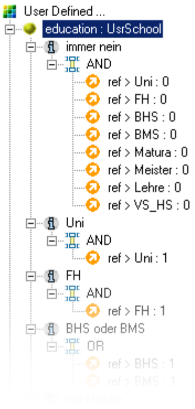


Figure 9: All displayed dimensions reflect the education of persons in each household – ordered from university to primary school (top to bottom); the information is available in *binary* form (yes or no) and is unsuitable for exploration.

The application provides several possibilities of what information is displayed in the histogram (Figure 8). There are two modes, which give especially useful information:

Relative Frequencies can be displayed in the histograms. Considering the crosstabulation of frequency values, the histogram displays the relative row/column frequencies. In statistical terms, the relative frequency actually is the conditional probability (if the marginal frequencies are seen as *posteriori* probabilities). Thus, the height of the histogram bars makes the relative relationship between two dimensions more apparent to the user and provides the probabilities of the relation conditional to each category.

Independence is also an interesting question, regarding the conditional probabilities. If the conditional probabilities are equal to the *posterori* probabilities, then the two dimensions are completely independent. In this kind of histogram, a straight line would visualize an independent relation, because each histogram bar displays the deviation of these two probabilities. For instance, the degree of independence can be used to introduce ordering to a dimension in relation to another dimension and provide very valuable information (this is done by reordering categories, so that the histogram shows a monotonic distribution).



```
User Defined ...
education : UsrSchool
  immer nein
    AND
      ref > Uni : 0
      ref > FH : 0
      ref > BHS : 0
      ref > BMS : 0
      ref > Matura : 0
      ref > Meister : 0
      ref > Lehre : 0
      ref > VS_HS : 0
  Uni
    AND
      ref > Uni : 1
  FH
    AND
      ref > FH : 1
  BHS oder BMS
    OR
      ref > BHS : 1
      ref > BMS : 1
  nur Matura
```

Figure 10: This meta information represents a possible classification to reduce the dimensionality of Figure 9.

7 Reality Check

To show how the presented features work and why they are useful, the exploration of a real data set should reinforce the implementation of each of them.

The Data is a questionnaire data set. 93.872 households were asked 99 questions about their living standard and regarding to pet-care and homecare. The questions are grouped according to certain topics: general information (school qualification, number of children, income), pet care information (whether there is a pet in the household) and homecare information (what is the favorite washing agent, supermarket).

The Task is to explore the categorical data that contains special structures: the data variables are hierarchically organized; the data consists of a large number of dimensions with low cardinality; and it is quite common that some questions are not answered, because of several reasons: privacy concerns, people do not want to answer questions, etc.

The Exploration

Figure 9 gives an example for dimension composition. To convert the raw data into a more convenient form, the dimensions (each gives the information, if one has obtained a particular qualification or not) are combined to one dimension that has one category for each qualification. By this, the data is not transformed, but the data is classified differently (the meta information that represents this new dimension is shown in Figure 10).

During this process, the user makes use of his domain knowledge and of the interaction possibilities: reordering of dimensions and categories and highlighting (to see detailed information and to enhance the relationships by elevating the streams). As a result of this first part, the user gets a new categorical dimension, with which he can work better – by means of adding it to the visualization again.

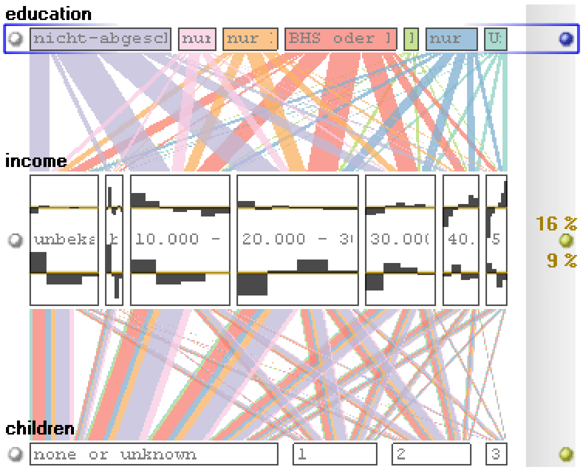


Figure 11: One can see a clear relationship between high income and good educational qualification and the fact of having one or no children.

This can be done repeatedly with different dimensions, to form ones that are more expressive. Then the user can add all the interesting dimensions into the view (conveniently three to four) and take a closer look at the relation among the variables. Figure 11 gives an example, of the relation between education, qualification, and number of children. The bottom dimension has an inherent order, whereas the top dimension is ordered, so that in the middle, the histogram of deviation between conditional and unconditional probabilities of the rightmost category shows a monotonic relation. This reveals that a better financial standing correlates with better education and the fact of having only one child or no children at all.

8 Conclusion

Categorical data sets are quite common in the field of visual data mining, but to display such data efficiently is a big challenge to InfoVis. This work presents a technique that provides a very intuitive approach: categories are mapped to their frequency values and these numbers are used to determine the visualization.

The two-dimensional layout and the discrete design model facilitate the understanding of the visualization and of the relationship between categorical attributes. An analysis is only possible if a user can properly handle the data visually. That is why interaction, visual feedback and the use of meta information is essential for dealing complex relations.

To conclude, it remains to be mentioned that the presented technique is an innovative idea in coping with categorical data that should influence future approaches as preceding approaches have influenced this work.

9 Acknowledgements

This work has been done in the scope of the basic research on visualization (http://www.VRVis.at/vis/) at the VRVis Research Center in Vienna, Austria (http://www.VRVis.at/), which is funded by an Austrian research program called Kplus.

Special thanks are given to Robert Kosara for his steady support and Helwig Hauser for giving the finishing touches.

References

[1] Titanic data (statlib – datasets archive: http://lib.stat.cmu.edu/s/harrell/data/descriptions/titanic.html).

[2] Richard A. Becker and William S. Cleveland. Brushing scatterplots. *Technometrics*, 29(2):127–142, 1987.

[3] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. Using vision to think. *Readings in information visualization: using vision to think*, pages 579–581, 1999.

[4] Helmut Doleisch, Martin Gasser, and Helwig Hauser. Interactive feature specification for focus+context visualization of complex simulation data. In *VISSYM '03: Proceedings of the 5th Joint IEEE TCVG - EUROGRAPHICS Symposium on Visualization*, pages 239–248. Eurographics Association, 2003.

[5] Michael Friendly. Visualizing categorical data: Data, stories and pictures. *SAS User Group International Conference Proceedings*, pages 190–200, 1992.

[6] Michael Friendly. *Visualizing Categorical Data*. SAS Publishing, 2001.

[7] Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., 1990.

[8] Nahum Gershon, Stuart Card, and Stephen G. Eick. Information visualization tutorial. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 149–150. ACM Press, 1999.

[9] D. L. Gresh, B. E. Rogowitz, R. L. Winslow, D. F. Scollan, and C. K. Yung. WEAVE: a system for visually linking 3-D and statistical visualizations, applied to cardiac simulation and measurement data. In *VIS '00: Proceedings of the conference on Visualization '00*, pages 489–492. IEEE Computer Society Press, 2000.

[10] Helwig Hauser, Florian Ledermann, and Helmut Doleisch. Angular brushing of extended parallel coordinates. In *INFOVIS '02: Proceedings of the IEEE Symposium on Information Visualization*, pages 127–130. IEEE Computer Society, 2002.

[11] StatSoft Inc. *Electronic Statistics Textbook*. http://www.statsoft.com/textbook/stathome.html, 2004.

[12] Alfred Inselberg and Bernard Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *VIS '90: Proceedings of the 1st conference on Visualization '90*, pages 361–378. IEEE Computer Society Press, 1990.

[13] Robert Kosara, Gerald N. Sahling, and Helwig Hauser. Linking scientific and information visualization with interactive 3D scatterplots. In *Short Communication Papers Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, pages 133–140, 2004.

[14] Anilkumar Patro, Matthew O. Ward, and Elke A. Rundensteiner. Seamless integration of diverse data types into exploratory visualization systems. Technical report, Worcester Polytechnic Institute, 2003.

[15] Geraldine E. Rosario, Elke A. Rundensteiner, David C. Brown, Matthew O. Ward, and Shiping Huang. Mapping nominal values to numbers for effective visualization. In *INFOVIS '03: Proceedings of the IEEE Symposium on Information Visualization*, pages 80–95. IEEE Computer Society, 2003.

[16] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *VL '96: Proceedings of the 1996 IEEE Symposium on Visual Languages*, pages 336–343. IEEE Computer Society, 1996.

[17] M. Spenke and C. Beilken. Visualization of trees as highly compressed tables with InfoZoom. *INFOVIS '03: Proceedings of the IEEE Symposium on Information Visualization*, pages 122–123, 2003.

[18] M. Theus, H. Hofmann, B. Siegl, and A. Unwin. Manet: Extensions to interactive statistical graphics for missing values. In *New Techniques and Technologies for Statistics II*, pages 247–259. IOS Press Amsterdam, 1997.

[19] J. Yang, M. O. Ward, E. A. Rundensteiner, and S. Huang. Visual hierarchical dimension reduction for exploration of high dimensional datasets. In *VISSYM '03: Proceedings of the 5th Joint IEEE TCVG - EUROGRAPHICS Symposium on Visualization*, pages 19–28. Eurographics Association, 2003.