A Task Based Performance Evaluation of Visualization Approaches for Categorical Data Analysis

Sara Johansson Fernstad and Jimmy Johansson
C-Research, Linköping University, Sweden
{sara.johansson, jimmy.johansson}@itn.liu.se

Abstract
Categorical data is common within many areas and efficient methods for analysis are needed. It is, however, often difficult to analyse categorical data since no general measure of similarity exists. One approach is to represent the categories with numerical values (quantification) prior to visualization using methods for numerical data. Another is to use visual representations specifically designed for categorical data. Although commonly used, very little guidance is available as to which method may be most useful for different analysis tasks. This paper presents an evaluation comparing the performance of employing quantification prior to visualization and visualization using a method designed for categorical data. It also provides a guidance as to which visualization approach is most useful in the context of two basic data analysis tasks: one related to similarity structures and one related to category frequency. The results strongly indicate that the quantification approach is most efficient for the similarity related task, whereas the visual representation designed for categorical data is most efficient for the task related to category frequency.

Keywords—Categorical Data, Quantitative Evaluation, Usability Studies, Parallel Sets, Quantification

1 Introduction
Over the past decades the availability of complex multivariate data sets has increased within a variety of domains. A major challenge for data analysts is to discover patterns and to gain understanding and knowledge from the structures within the data. Different kinds of data have different characteristics that need to be taken into consideration when deciding which method to use for analysis. One data type which requires specialized analysis methods is categorical data. For categorical data there exists no similarity measure comparable with that of numerical variables. Due to this, the analysis of categorical data is often not as straightforward as for numerical data and fewer generic methods are available. Nonetheless, categorical data are common within many areas such as social sciences, biology, chemistry and medicine; some examples being census data, organism classification data, DNA sequence-based data and molecular data. Efficient visualization methods can facilitate and speedup analysis considerably, and so the development of efficient methods for visual analysis of categorical data is an important issue within information visualization.

Two main approaches have been suggested for visual analysis of categorical data. One is to represent each category with a numerical value (quantification) and then analysis using visualization methods commonly employed for numerical data (from here on this approach will be called QuantViz). The other approach is to employ visualization methods specifically designed for the characteristics of categorical data (from here on called CatViz).

Although a range of methods for visual analysis of categorical data are available we cannot assume that all are useful and facilitate analysis and, even more, we cannot assume that all methods are equally useful for all types of tasks. Without proper evaluations of the usability of methods within specific task contexts, there is nothing but presumption to guide analysts as to which method to use. While quantification has been suggested several times as a method for efficient analysis of categorical data, very little evidence of its usefulness has been presented and no formal comparisons between quantification methods and visualization methods for categorical data have, to the best of our knowledge, been made.

1.1 Contributions
Through a usability study we now make a first attempt to evaluate the effectiveness and efficiency of employing quantification of categorical data prior to visualization. We also aim to provide an initial step of guidance in terms of establishing the usability of the two approaches in the context of performing basic analysis tasks. This is approached by comparing the performance of analysis using a QuantViz method with the performance using a CatViz method, while carrying out a number of tasks, details of which can be found in section 4.1.1. In this context performance is defined as a combination of accuracy and re-
A formal comparison of two methods for visual analysis of categorical data.

An experimental study in the context of two basic data analysis tasks.

Guidance as to which method may be most useful for different types of tasks.

It is worth noting that although the main goal of the study is not to compare the performance of one visual representation with another (e.g., parallel coordinates [9] vs. Parallel Sets [13]), but to compare one way of visually analysing categorical data with another (QuantViz vs. CatViz), the results of a comparison of a specific set of visualization methods is not generalizable to all available visualization methods. Hence, the results of this study are only true for the specific visualization methods and tasks of this study. This is a first step in evaluating the performance of quantification as a method for visualization of categorical data and has to be followed by additional evaluations focusing on other aspects before any general conclusions can be drawn.

1.2 Paper Overview

The remainder of this paper is organized as follows: Section 2 presents related research and in section 3 the design decisions of the study are described and motivated. Section 4 describes the study and presents the results, which are finally discussed in section 5.

2 Related Work

This section provides an overview of research in the areas of categorical data visualization, quantification and performance evaluation.

2.1 Visualization

Most visualization methods for categorical data are based on tables of category frequencies (contingency tables). The fourfold display [4] is one example where the cell frequencies of two-by-two tables are represented by quarter circles with size relative to frequency. Other examples are for instance the mosaic plots and mosaic matrices [4] where multivariate tables are represented by tiles whose sizes are proportional to the cell frequencies. In a Cobweb diagram [23] the categories are represented as individual nodes that are pairwise linked together using lines whose lengths represent the deviation from independence of the category pair frequency. Parallel Sets [13] (described in more detail in section 3.2) has a layout similar to parallel coordinates where the categories of a variable are represented with a set of boxes. Another example using the layout of parallel coordinates is presented by Havre et al. [7], where the polylines are spread over additional axes and sorted to display category frequencies and to avoid data overlay. The CatViz method [12] is an example of a tree-map based visualization for representation of hierarchically structured categorical data. In the Attribute Map View [14] categorical values are represented within attribute rows by rectangles whose sizes are relative to category frequency, and in the TreemapBar [8] bar charts are combined with tree-maps by embedding a tree-map visualization inside the rectangular space of the bars.

To conclude, CatViz methods are mostly designed for visualization of category frequencies and their efficiency is often highly dependent on data set size and structure. To the best of our knowledge no formal comparison has been made between the majority of CatViz methods and due to this the selection of method to use in this study is based on our subjective opinion that Parallel Sets (figure 1) is the CatViz method whose efficiency is least affected by data set size, being able to display approximately as many variables as parallel coordinates.

2.2 Similarity Measures and Quantification

Several algorithms usable for quantification of categorical data have been suggested. One of the more common being Correspondence Analysis (CA) [6] which is a method identifying associations between the cells of a frequency table. Tenenhaus and Young [22] describe how a number of different methods for quantification, such as Multiple Correspondence Analysis, Optimal Scaling and Homogeneity Analysis, all lead to the same equation. Hence, they are one method with a number of different names and will all be referred to as CA throughout the remainder of this paper. In Cuadras et al. [3] three methods for quantification are compared; CA, an approach using the Hellinger distance and a log ratio approach. They conclude that although CA and the Hellinger distance approach sometimes provide similar results, CA is the best due to a range of properties, whereas the log ratio approach often provides quite different results from the other two. Shen et al. [20] presents a framework for mapping categorical data to numerical values through calculation of similarities between data items and a reference set. The effectiveness of their method is compared with the effectiveness of other methods in the context of clustering and visualization. Ma and Hellerstein [15] presents a technique for ordering of categorical data, using a combination of clustering and domain semantics. In Rosario et al. [19] methods for numerical representation of categories using CA are presented. Two different CA computations and an arbitrary quantification with uniform spacing are compared for a number of categorical data sets using parallel coordinates. Johansson et al. [11] extended the quantification approach to mixed data sets by incorporating information about relationships among numerical
variables into the process, and included the possibility of interactive modification to make use of the domain knowledge of expert users. Evaluating the performance of different quantification methods is not within the scope of this paper. We used CA in the study since it has been suggested several times as a method for quantification of categorical data, it is a generic method not designed for specific tasks such as clustering, and since a previous evaluation by Cuadras et al. [3] picked it out as a good quantification method compared to some other methods.

2.3 Evaluation

Different approaches can be taken to the evaluation of visualization methods, mainly depending on the objective of the study. In Plaisant [16] four main areas for evaluation in information visualization are listed; 1) controlled experiments comparing design elements, 2) usability evaluation of a tool, 3) controlled experiments comparing two or more tools, and 4) case studies of tools in realistic settings. The study presented in this paper can be defined as a controlled experiment comparing two approaches to visual analysis and is, hence, most closely related to Plaisant’s first and third areas. A large number of studies have been performed within those evaluation areas and a full review is beyond the scope of this paper. In this study the performance of two visualization methods is compared in the context of specific tasks, defining performance as a combination of accuracy and response time. This is a common approach for quantitative evaluations and some previous examples are presented in for instance Stasko et al. [21] where the performance of two space-filling visualization methods for hierarchical data is compared, and in Plaisant et al. [17] where SpaceTree, a tree browser based on node link diagrams, is compared with Microsoft Explorer and a Hyperbolic tree browser. Similarly Johansson et al. [10] investigated the ability of humans to perceive relationships in 2D and 3D parallel coordinates and in Vrotsou et al. [24] a comparison was made between the performance using a 2D representation and a 3D representation for visual analysis of time-geographical data.

3 Design Decisions

Due to the fundamental differences between the QuantViz and CatViz approaches, a major concern during the design of this study has been to make decisions in terms of design in a way that makes it possible to perform a fair comparison, while not removing the basic differences and benefits of any of the methods. As CatViz method Parallel Sets is used (figure 1, described in detail in section 3.2) and the quantified data is visually represented using a multiple views set-up of three common visualization methods; parallel coordinates, table lens [18] and scatter plot matrix [1] (figure 2, described in detail in section 3.1). It might appear as an unfair comparison using multiple views for the QuantViz approach while a single view is used for the CatViz approach. This is however a carefully considered decision based on the following:

1. Using multiple views facilitate analysis if the views complement each other. However, the use of multiple views increase the perceptual burden by demanding coordination of several sources of information.
2. Visualization methods for numerical data have different strengths and weaknesses and may complement and strengthen each other in a multiple view system.
3. Visualization methods designed for categorical data are based on category frequencies. They generally have the same strengths and weaknesses and hence do not complement and strengthen each other in the same way as methods for numerical data may.
Our aim has been to present all methods used in the study as advantageous as possible. The benefit of using multiple views depends both on visualization methods and the task at hand. For the tasks performed in this study we found no CatViz method that would complement Parallel Sets in a favourable way since all methods focus on category frequencies. Hence, using a single view was considered most beneficial for the CatViz approach. For the QuantViz approach on the other hand we could clearly appreciate that the features of for instance a table lens, which is able to display category frequencies, would complement the features of parallel coordinates in an advantageous way. Similarly a scatter plot matrix is able to display all pairwise correlations concurrently, which is not possible in parallel coordinates or table lens. Hence, using multiple views was considered most beneficial for the QuantViz approach. Another important issue was to avoid performance differences appearing due to interactive differences and features within the specific tools used in the study. To assure this the participants were presented with pre-prepared screen shots of the tools and were, hence, not allowed to interactively explore the data. Through this the individual differences between the tools were reduced as much as possible without removing the basic differences and benefits of any of the visualization methods. The visualization methods used and the layout rules employed when creating screen shots will be describe in the two following sections.

3.1 Quantification and Visualization

CA was used as quantification method, as this is a commonly used method for quantification purposes. CA is used for analysis of frequency tables where each cell represents the frequency of a combination of categories and it identifies similarity between cells by extracting independent dimensions in the table, using singular value decomposition. The first independent dimension, which explains most of the variance within the frequency table, is used as numerical representations of the row categories in the frequency table [6]. A multiple views layout is used for visualization (figure 2), including a scatter plot matrix, a table lens and parallel coordinates. These methods are commonly available and complement each other, and are hence considered to provide a good representation of benefits using the QuantViz approach. In general the QuantViz approach aims to perform analysis as if the data would have been numerical. Based on this the visualization methods employed for QuantViz in this study are used as if the data were numerical and hence basic implementations of the visual representations has been used. Although the visual representations could have been extended with additional features facilitating categorical data analysis, this is not used since the method would then no longer be a pure QuantViz method but a combined QuantViz/CatViz method. The only visual enhancement used is displaying the category names next to the axes of the parallel coordinates to provide a link between the numerical representation of a category and its name. The layout rules used when preparing screen shots of this method are as follows:

1. Colouring is in all views performed according to the categories of the leftmost variable.
2. When tasks relate to one specific variable that variable is positioned to the left and sorting of rows in the table lens is performed according to it.
3. When tasks relate to two specific variables those are positioned as the two leftmost variables. Sorting of rows in the table lens is performed according to the second variable.

3.2 Parallel Sets

Parallel Sets [13] (figure 1) is a visualization method for analysis of categorical multivariate data where the variables of the data set are represented by horizontal parallel axes. Instead of displaying individual data items Parallel Sets focus on displaying category frequency. Within each variable the categories are represented by vertical boxes whose width corresponds to the relative frequency of the category. The width of the bands stretching between the categories of adjacent variables represent the relative frequency of a combination of categories. For this study a publicly available implementation of Parallel Sets has been used (http://eagereyes.org/parallel-sets). All screen shots of Parallel Sets within the paper have been created using this implementation, displaying synthetic data sets designed for this study. The tool’s available features of re-ordering variables as well as categories have been used to prepare screen shots as follows:

1. Colouring is performed according to the categories of the topmost variable.
2. When tasks relate to one specific variable, that variable is always positioned as the top variable, and when tasks relate to two specific variables, those variables are always positioned as the two top variables.
3. For all variables the categories are internally ordered according to frequency, with highest frequency to the left and lowest to the right.

4 The Study

The aim of this study is to evaluate the effectiveness and efficiency of utilizing quantification followed by visualization for analysis of categorical data. More specifically, we want to evaluate the performance when carrying out basic data analysis tasks using the quantification approach compared to carrying out the same tasks using Parallel Sets.
is motivated by the fact that the ability to efficiently perform the basic tasks is fundamental for the ability to efficiently perform more complex tasks. The overall task of data analysis is to identify structures and patterns within data. Most patterns, such as correlation and clusters, can be defined in terms of similarity. Hence, the most relevant general task to focus on are, in our opinion, the identification of relationships in terms of similarity. Additionally, when it comes to analysing categorical data the frequency of categories, i.e. the relative number of items belonging to specific categories or combinations of categories, is often of major interest and is, as mentioned previously, the main property of focus in categorical data visualization. Based on this, the tasks of this study are separated into two classes: 1) identification of category frequencies, and 2) identification of similarity patterns. Within the two classes questions are asked relating either to structures within one variable or structures between two variables (from here on called 1VQ and 2VQ respectively). To be more precise, four tasks are defined as follows:

- Frequency task, 1VQ: Which category in variable X is the most common?
- Frequency task, 2VQ: Which combination of categories in variables X and Y is the most common?
- Similarity task, 1VQ: Which two categories in variable X are most similar to each other?
- Similarity task, 2VQ: Which variable is most similar to variable X?

In terms of similarity the following definitions were used:

1. Two categories are similar if they (in general) belong to the same categories within all (or most) variables in the data set (example displayed in table 1).
2. Two variables are similar if the category combinations for the variables are clearly separated, that is, if the data items of each category in variable X (in general) belongs to only one category in variable Y, and vice versa (example displayed in table 2).

### 4.1 Method

To evaluate the performance the participants of the study carried out a number of tasks (described in detail in section 4.1.1) using the two visualization approaches. Performance was measured in terms of accuracy and response time, which were recorded for all tasks and stored in log files. Questionnaires were used to receive the subjective opinion of the participants in terms of difficulty of performing tasks and preference of visualization method.

#### 4.1.1 Tasks

The tasks used in the study were selected to represent the basic elements of typical data analysis tasks, both specific to categorical data analysis and more general tasks. We focus solely on the basic elements of the tasks, which

<table>
<thead>
<tr>
<th>Item</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>D</td>
<td>G</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>D</td>
<td>G</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>E</td>
<td>H</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>E</td>
<td>H</td>
</tr>
</tbody>
</table>

Table 1: A data set including three variables (X, Y, Z) and six data items. If analysing the similarity of the three categories in variable X categories A and B are considered more similar to each other than they are to category C. This is true since the items belonging to categories A and B belong to the same categories in variables Y (D) and Z (F and G), whereas the items belonging to category C in variable X belong to a different set of categories in variable Y (E) and Z (H).

<table>
<thead>
<tr>
<th>Item</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>D</td>
<td>G</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>D</td>
<td>G</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>E</td>
<td>H</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>E</td>
<td>H</td>
</tr>
</tbody>
</table>

Table 2: An example data set including three variables (X, Y, Z) and six data items. If analysing the similarity of the three variables, variable X and Z are considered more similar to each other than they are to variable Y. This is true since the data items belonging to the three categories in variable X belong to one category each in variable Z, and vice versa. Whereas they are spread over the two categories in variable Y.

#### 4.1.2 Materials

The tests were carried out on an HP EliteBook 8540p laptop with an Intel i5 2.53GHz CPU, 4 GB RAM and an Nvidia NVS 5100M graphics card. An external Dell 20” monitor was used, set to a resolution of 1680 x 1050 pixels. The participants were presented with screen shots of the visualization methods together with the task to perform, included in figures of size 1100 x 800 pixels. The experimental environment used was designed in Matlab. To retain full control over relationships in the data synthetic data sets were created using Matlab. A total of 32 data sets are included.
were created, 16 designed for frequency related tasks and 16 for similarity related tasks, half of which where smaller data sets including three to four variables and three to five categories within each variable, and half being larger data sets including six to seven variables and six to eight categories within each variable. The data sets were designed to provide relatively clear answers to the tasks.

### 4.1.3 Experimental Design

The study was designed as a 2-factor within-subject design, where the two factors were type of analysis (QuantViz vs. CatViz) and type of task (Frequency vs. Similarity). The four separate experimental phases were hence:

1. QuantViz + Frequency
2. QuantViz + Similarity
3. CatViz + Frequency
4. CatViz + Similarity.

The presentation order of the four phases was counterbalanced using a Latin-square procedure [5], resulting in four different phase orders. The participants were randomly assigned to one of the four phase orders, assigning equally many participants to each.

Performing the tasks of this study includes interpretation of fairly complex visual representations, and hence some visualization experience was set as a requirement for participants. Due to this the participants were 15 researchers and students within areas related to visualization, aged between 22 and 35 years. Each participant performed 32 tasks in total (8 per phase), and no data set was used more than once per participant, avoiding that any structures asked for had been identified by chance during previous tasks. The presentation order of data sets and questions was randomized within each phase. However, since each task type includes two different kinds of questions (1VQ and 2VQ) and since the data sets were of different sizes, the randomization was limited by a certain level of control. During each phase it was ensured that equally many 1VQ and 2VQ were asked, and that half of each question type was asked using a smaller data set and half using a larger data set. No major performance differences were expected due to data set size and type of question, thus they were not treated as factors within the experimental design. However, information on size and question type was recorded to be able to identify unexpected results related to this.

### 4.1.4 Procedure

To assure that all participants possessed the basic knowledge needed to be able to interpret the visual representations, an initial introduction was held for all participants, either in small groups or individually. The introduction was in the form of a small scripted lecture where the visual representations were described in detail as well as the basic concept of quantification. Furthermore, the rules used for screenshot layout was described. The concepts of frequency and similarity in the context of the study was also explained, and the basic task types were presented to the participants together with example images displaying how the structure asked for can be identified using each of the analysis methods. The explanations and example images were printed out and made available throughout the test. The participants were told that their response time would be measured and that they should try to answer as quickly as possible but to maintain high accuracy.

The test itself was performed individually and consisted of a training period and the experiment. The training period included a small number of test tasks using both analysis types and was used as a means for getting familiar with the experimental environment, methods and tasks. During the training period the participants were encouraged to ask questions to make sure they had understood the tasks. The experiment consisted of four separate phases, as described in section 4.1.3, where the participant was presented with 8 tasks and screen shots. Once the answer to the task was found the participant pressed a button, the screen shot was hidden and an interface for submitting the answer was displayed. For each task the response time was recorded from when the screen shot was displayed until the answer button was pressed. The answers provided, as well as response time and data sets used, were stored in log files. When a phase was finished a pause screen was displayed, allowing a break between the phases, and the participant was asked to answer two questions; 1) Did you find it difficult or easy to answer the questions using the visualization method? 2) Did you find it more difficult to answer the questions when the number of variables and number of categories increased? The questionnaire also encouraged the participants to provide additional comments on the method used and tasks performed. When all four phases were finished the participants were asked to answer a final questionnaire, including questions as to which analysis type the participant preferred using for each task type.

### 4.2 Result

This section will present the results of the study in detail. Results are reported both in terms of performance measures and questionnaire answers.

#### 4.2.1 Performance

The measured data were initially not normally distributed. Due to this the response time was logarithmically transformed prior to statistical testing, and a repeated measures ANOVA [5] was carried out on the logarithmically transformed data, using an $\alpha$ value with a significance of $p < 0.05$. For accuracy a Friedman test was used for significance testing, since this data could not be logarithmi-
Table 3: Average response times in seconds for the four phases, overall average at the top and below split into one and two variable questions. Standard deviation is displayed within parentheses.

<table>
<thead>
<tr>
<th></th>
<th>QuantViz + Frequency</th>
<th>QuantViz + Similarity</th>
<th>CatViz + Frequency</th>
<th>CatViz + Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>26.73 (14.42)</td>
<td>16.15 (7.75)</td>
<td>10.55 (5.12)</td>
<td>32.45 (15.67)</td>
</tr>
<tr>
<td>1VQ</td>
<td>13.91 (10.58)</td>
<td>16.40 (9.68)</td>
<td>7.45 (3.31)</td>
<td>27.04 (11.88)</td>
</tr>
<tr>
<td>2VQ</td>
<td>39.56 (21.63)</td>
<td>15.90 (7.38)</td>
<td>13.66 (7.02)</td>
<td>37.85 (22.67)</td>
</tr>
</tbody>
</table>

Table 4: Descriptive statistics of the measured accuracy for the four phases, displaying median, minimum, maximum, 1st quartile and 3rd quartile. The maximum number of correct answers is 8.

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>QuantViz + Freq.</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>QuantViz + Sim.</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CatViz + Freq.</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CatViz + Sim.</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

The two factors of the study were visualization type (QuantViz vs. CatViz) and task type (Frequency vs. Similarity), resulting in four separate phases. Average response times of the phases are displayed in the top row of table 3 and the statistics of accuracy are displayed in table 4. In terms of overall response time, the average time for the two visualization methods were equal, as displayed in figure 3, and statistical analysis reports no significant effect of visualization type, $F(1, 14) = 3.112, p < 0.099$. However, there is a significant effect of task type, $F(1, 14) = 52.998, p < 0.001$, with frequency tasks being performed almost 25% faster than similarity tasks, as displayed in figure 3. Furthermore, statistical analysis reports a significant interaction effect between visualization type and task type, $F(1, 14) = 122.007, p < 0.001$. This is also visible in figure 3 with similarity questions being performed considerably faster using QuantViz whereas frequency tasks are performed faster using CatViz. Accuracy results generally agree with response time, with high performance for similarity tasks using QuantViz and for frequency tasks using CatViz, as visible in table 4 and from the box plots in the left part of figure 4. Statistical testing confirms that the difference between the four phases is significant, $\chi^2(3, N = 15) = 32.929, p < 0.001$. Furthermore, post-hoc testing shows a significant difference between using QuantViz for frequency tasks and CatViz for similarity tasks, whereas there was no significant difference between using QuantViz for similarity tasks and CatViz for frequency tasks. These results indicate performance differences depending on the combination of visualization type and task type, with higher performance using CatViz for similarity tasks and QuantViz for frequency tasks. However, looking more closely into the results it appears as if question type also had a strong influence on performance.

The two bottom rows of table 3 display the response times of the question types, showing a noticeable difference between performing 1VQ frequency tasks and 2VQ frequency tasks using QuantViz. A similar relationship is identified for 1VQ and 2VQ frequency tasks using CatViz. The difference between 1VQ and 2VQ for similarity tasks is however not as explicit. Similarly, the difference in terms of accuracy between the question types is noticeable, as displayed in table 5 and in the right part of figure 4. Most obvious is the difference between 1VQ and 2VQ frequency tasks using the QuantViz approach. Here the accuracy of 1VQ is almost as high as the accuracy of performing frequency tasks using CatViz, whereas the accuracy of performing 2VQ frequency tasks using QuantViz has a median value of zero. Statistical testing of response time, with three within-subject factors; visualization type (QuantViz vs. CatViz), task type (Frequency vs. Similarity) and question type (1VQ vs. 2VQ), was performed using pairwise comparisons according to Conover [2], to identify which phases are significantly different from each other.
Figure 3: Average response time in seconds for visualization type (left), task type (centre) and the four combinations of task and visualization types (right).

Figure 4: Box plots displaying number of accurate answers for the four phases (left) and phases divided based on question type (right). The red lines represent median values.

The questionnaire answers strongly implied that CatViz was the preferred method for performing frequency tasks, as stated by twelve of the participants, and that QuantViz was the preferred method for performing similarity tasks, as stated by fourteen participants. Tables 6 and 7 provide more information on the participants opinions regarding the visualization approaches and tasks. Table 6 displays the number of participants finding it easy to solve the tasks for the four phases and table 7 displays the number of participants who found the tasks more difficult as the data set size increased.

Some participants found it both difficult and easy performing the tasks. Using QuantViz for frequency tasks a main comment on this was that it was easy for 1VQ (i.e. find the most common category in variable X) but difficult for 2VQ (i.e. find the most common combination of categories in variables X and Y). Regarding the increase in difficulty as the data set size grew, comments indicate that this was mostly true for 2VQ. Comments from the participants regarding using the QuantViz approach for similarity tasks indicate that the difficulty mentioned by some participants was mainly due to examples where more than one variable could be interpreted as similar. Using CatViz all participants found it easy to answer the frequency tasks, and most did not find the tasks more difficult as the data set size increased. An explanation made as to why it was easy to use
Table 6: Number of participants for each phase who found it easy to perform the tasks. Numbers within parentheses are total number of participants either finding it easy or both difficult and easy.

<table>
<thead>
<tr>
<th></th>
<th>QuantViz</th>
<th>CatViz</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>3 (8)</td>
<td>15</td>
<td>18 (23)</td>
</tr>
<tr>
<td>Similarity</td>
<td>12 (14)</td>
<td>4 (10)</td>
<td>16 (24)</td>
</tr>
<tr>
<td>Total</td>
<td>15 (22)</td>
<td>19 (24)</td>
<td></td>
</tr>
</tbody>
</table>

this visualization approach was that the answers were obvious due to size and variable sorting. When it came to performing similarity tasks using CatViz a comment made as to what made it difficult was that it was almost impossible to follow the bands between the variables as they spread. Regarding increase in number of variables and categories, the main comment was that more bands, colours and categories generated more clutter.

5 Discussion and Conclusions

In general the performance results, both in terms of accuracy and response time, agree well with the subjective experiences reported by the participants through the questionnaires. For similarity related tasks a majority preferred using QuantViz, which was also the visualization approach for which the recorded performance was best for similarity tasks. For frequency related tasks on the other hand, a majority of participants preferred using CatViz, which also agrees well with the recorded performance. A result worth emphasizing is that when the tasks where split up into questions concerning the frequency of categories within one variable or a combination of categories in two variables, the performance was noticeably higher for the one variable questions using QuantViz. This was also mentioned by several participants who found the one variable questions easy during this phase, whereas two variable questions were perceived as more difficult. In terms of data set sizes no major performance differences were found. However, the participants reported an increase in perceived difficulty as the size increased. More specifically, a majority of participants experienced an increase in difficulty for phases that were generally not perceived as easy, as can be identified through the inverse relationships between tables 6 and 7. Possibly the limitations of both approaches became more obvious as the tasks got more difficult.

As a final point of discussion it is important to again emphasize that although the study presented in this paper aims to compare the two main approaches to visual analysis of categorical data, and not to compare two visualization tools or two visual representations, the results can not be directly generalized to all visualization methods designed for categorical data, nor to all visualizations of quantified categorical data. However, it can be seen as an initial attempt to compare the effectiveness and efficiency of the two visualization approaches; not aiming to find an ultimate solution as to how categorical data is best analysed visually, but aiming to provide guidance as to which method might be most useful for a specific type of task and to encourage further evaluations of the two approaches.

Acknowledgements

We would like to thank Camilla Forsell for her kind advice on study design and analysis of results. This work was partly supported by the Swedish Research Council in the Linnaeus Centre CADICS and by the Visualization Programme coordinated by the Swedish Knowledge Foundation.

References


