Evaluating the Dot-Based Contingency Wheel: Results from a Usability and Utility Study

Margit Pohl[1], Florian Scholz[1], Simone Kriglstein[1], Bilal Alsallakh[2], and Silvia Miksch[2]

[1] Institute for Design and Assessment of Technology, Vienna University of Technology, Austria
{margit,florian,simone.kriglstein}@igw.tuwien.ac.at
[2] Institute of Software Technology & Interactive Systems, Vienna University of Technology, Austria
{alsallakh,miksch}@ifs.tuwien.ac.at

Abstract. The Dot-Based Contingency Wheel is an interactive visual-analytics method designed to discover and analyze positive associations in an asymmetrically large $n \times m$ contingency table. Such tables summarize the relation between two categorical variables and arise in both scientific and business domains. This paper presents the results of a pilot evaluation study based on interviews conducted with ten users to assess both the conceptual design as well as the usability and utility of the Dot-Based Contingency Wheel. The results illustrate that the Wheel as a metaphor has some advantages, especially its interactive features and ability to provide an overview of large tables. On the other hand, we found major issues with this metaphor, especially how it represents the relations between the variables. Based on these results, the metaphor was redesigned as Contingency Wheel++, which uses simplified and more familiar visual representations to tackle the major issues we identified.

Keywords: Visual Analytics, Evaluation, User Interface, Interview, Contingency Tables.

1 Introduction

Categorical data appear in many data tables both in scientific and in business domains. In contrast to numerical variables, the values of a categorical variable have no inherent order. Therefore, common analysis techniques that handle numerical variables are usually inapplicable to analyze categorical data. The analysis of categorical data is usually based on contingency tables. A two-way contingency table is a matrix that records how often each combination of categories from two categorical variables appears in the database. An example would be the combination of color of hair and color of the eyes. A contingency table would, in this example, contain the frequency of the co-occurrence of blue eyes and blond hair, brown eyes and brown hair etc. An example for a contingency table in the context of medical applications would be types of diseases vs. groups

S. Yamamoto (Ed.): HIMI 2014, Part I, LNCS 8521, pp. 76–86, 2014.
© Springer International Publishing Switzerland 2014

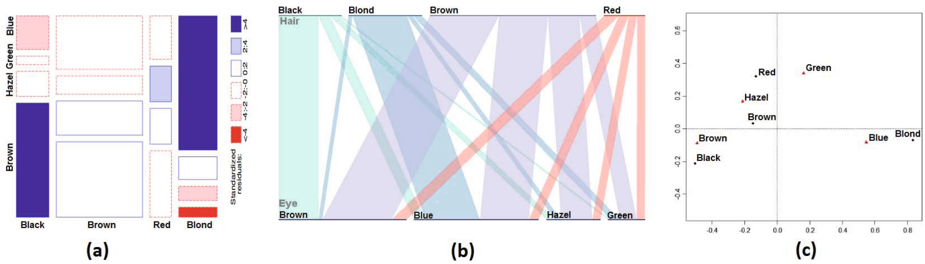


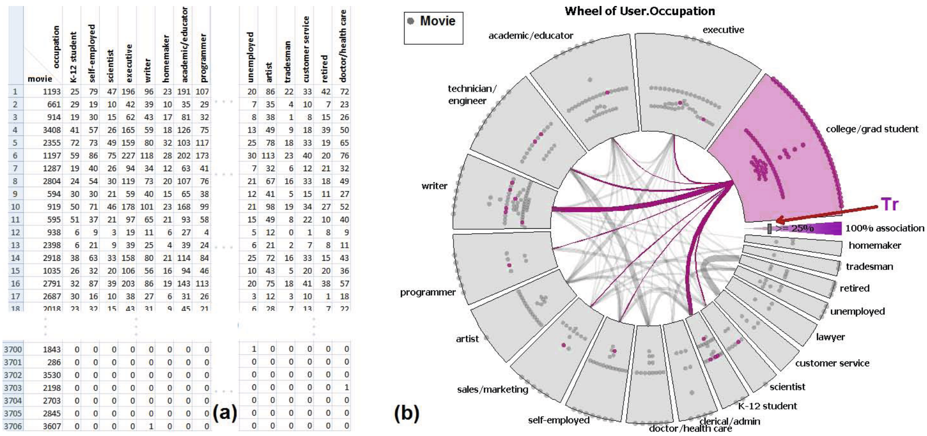
Fig. 1. Three visualizations of for contingency tables: (a) Mosaic Displays [1], (b) Parallel Sets [2], (c) Correspondence Analysis [3] (Screenshot by Alsallakh et al. [4])

of patients (e.g., female or male). In medical or business domains, the amount of data which has to be handled can get very large. It is difficult for users to find patterns or relationships in contingency tables in tabular form.

Beside several statistical techniques to extract information from such tables, several methods have been developed to visualize these tables such as mosaic displays [1], Parallel Sets [2], and correspondence analysis (CA) [3]. These visualization methods offer insights into the table such as an overview of the distribution of the data or how categories are associated with each other (see Figure 1). For example, a blue tile in mosaic displays, a thick arc in Parallel Sets, and close points in a CA plot mean that the corresponding categories are highly positively associated. Such categories appear more often together than on average in the database.

While the above-mentioned methods offer relatively intuitive representations to visualize contingency tables, they suffer from limited scalability. Only a small number of categories from both variables can be depicted in the visualization. The Contingency Wheel [4] has been proposed as a visual-analytics method designed to discover and analyze positive associations in contingency tables that have a few number of columns but a large number of rows. It uses dots to visually map these associations on a ring chart, as we explain in Section 2.

The following paper presents the results of an evaluation study of the Contingency Wheel which is based on interview data. The motivation of our evaluation study was to find out how users interact with a complex visualization like the Contingency Wheel. The findings of the evaluation help us to identify which functionalities need further improvements. Therefore, we chose a qualitative approach for the evaluation, described in Section 3. Carpendale [5] argues that such methods yield a rich understanding of the various factors influencing the interaction with information visualizations. Qualitative analysis also enables investigators to obtain contextual information about the usage of information visualization or visual analytics methodologies. In Section 4 an overview of the investigation is presented. The prototype we evaluated is described in detail in Section 5. The results we got from the analysis of the interviews are discussed in Section 6, whereas a discussion and a short overview about the improvements



	movie \ occupation	K-12 student	self-employed	scientist	executive	writer	homemaker	academic/educator	programmer	…	unemployed	artist	tradesman	customer service	retired	doctor/health care
1	1193	25	79	47	196	96	23	191	107		20	86	22	33	42	72
2	661	29	19	10	42	39	10	35	29	…	7	35	4	10	7	23
3	914	19	30	15	62	43	17	81	32		8	38	1	8	15	26
4	3408	41	57	26	165	59	18	126	75		13	49	9	18	39	50
5	2355	72	73	49	159	80	32	103	117		25	78	18	33	19	65
6	1197	59	86	75	227	118	28	202	173		30	113	23	40	20	76
7	1287	19	40	26	94	34	12	63	41	…	7	32	6	12	21	32
8	2804	24	54	30	119	73	20	107	76		21	67	16	33	18	49
9	594	30	30	21	59	40	15	65	38		12	41	5	15	11	27
10	919	50	71	46	178	101	23	168	99		21	98	19	34	27	52
11	595	51	37	21	97	65	21	93	58		15	49	8	22	10	40
12	938	6	9	3	19	11	6	27	4		5	12	0	1	8	9
13	2398	6	21	9	39	25	4	39	24	…	6	21	2	7	8	11
14	2918	38	63	33	158	80	21	114	84		25	72	16	33	15	43
15	1035	26	32	20	106	56	16	94	46		10	43	5	20	20	36
16	2791	32	87	39	203	86	19	143	113		20	75	18	41	38	57
17	2687	30	16	10	38	27	6	31	26		3	12	3	10	1	18
18	2018	23	32	15	43	31	9	45	21		6	28	7	13	7	22
⋮	⋮				⋮										⋮	
3700	1843	0	0	0	0	0	0	0	0		1	0	0	0	0	0
3701	286	0	0	0	0	0	0	0	0		0	0	0	0	0	0
3702	3530	0	0	0	0	0	0	0	0		0	0	0	0	0	0
3703	2198	0	0	0	0	0	0	0	0	…	0	0	0	0	0	1
3704	2703	0	0	0	0	0	0	0	0		0	0	0	0	0	0
3705	2845	0	0	0	0	0	0	0	0		0	0	0	0	0	0
3706	3607	0	0	0	0	1	0	0	0		0	0	0	0	0	0

Fig. 2. (a) A large contingency table and (b) the corresponding Dot-Based Contingency Wheel present how the movies are associated with the user groups. Only cells that exhibit positive association larger than $T_r = 25\%$ between the rows and the columns are mapped to dots (on a scale between -100% and $+100\%$).

based on these findings is presented in Section 7. Section 8 concludes the paper and gives an outlook on future work.

2 The Dot-Based Contingency Wheel

The Contingency Wheel [4] is a visualization technique for large contingency tables. To cope with such tables, it adopts a visual-analytics paradigm by first computing associations in the table using statistical residuals. These associations along with the table data are then mapped to the following visual elements:

- **Sectors** represent the table columns and form a ring chart. The size of a sector is proportional to the marginal frequency of the corresponding column, which is equal to the sum of the cell frequencies in this column.
- **Dots** represent a subset of the table cells. A dot is created for a cell in the sector of its column, if its rows and column are positively associated. Such cells contain values that are larger than would be if the row and column variables were independent. The radial position of a dot is proportional to the value of this positive association. The higher the association, the closer is the dot to the outer boundary (the same information is encoded using color in mosaic displays, thickness in Parallel Sets, and proximity in a CA plot). The dots are spread along the angular dimension to reduce overlapping.
- **Lines** represent the existence of shared data between the sectors. A line is drawn between two sectors if at least two cells from the same row result in dots in both sectors. The line is thicker if the two sectors contain more such dots and the higher the associations these dots represent.

– In addition, a purple **slider** shows the value of the threshold T_r that can be adjusted to filter the dots. Only dots that represent associations higher than T_r are retained in the visualization.

Figure 2 shows an example wheel visualization for a large 3706×21 contingency table. The rows represent movies, while the columns represent user groups and are mapped to sectors. The cells represent how many times each movie was rated by users from each user group, and are mapped to dots. One user group is selected to highlight all dots that represent movies positively associated with it. Such movies were more often watched by users from this group than the average of all users. Some of these movies are also positively associated with other user groups, as indicated by the highlighted lines and dots in other sectors.

3 Goal of the Investigation

The motivation of our evaluation study was to find out how users interact with visual-analytics methods, especially with complex visualizations like the Contingency Wheel. The evaluation of the fundamental idea served as an initial test instrument to find out the advantages and drawbacks of the representation. The findings of the evaluation help us to identify which functionalities need further improvements and to check if users missed important features which should be considered for the redesign of the methodology. For this purpose, the main questions which we wanted to investigate are:

1. What are the main advantages of the Contingency Wheel?
2. What are its main disadvantages?
3. Are the concepts underlying the Contingency Wheel clear?
4. How useful are the interactions provided by the methodology?

4 Description of the Investigation

We tested the prototype with ten persons who studied computer science. The reason for this choice of participants was that they were familiar with computers as well as statistical methods. In this way, the focus was on the visualization itself and not on how users get acquainted with basic principles of the tested software. Five participants were additionally visualization experts who worked with visualizations very often. Testing sessions for each participant took about 90 minutes.

The dataset used for the evaluation concerned the answers to standardized psychological tests of 300 young patients, their parents and their teachers. These answers were collected in the course of testing the patient for ADHD ("attention deficit hyperactivity disorder") and mostly diagnosing them with it. The motivation to use this dataset was to ensure that our test persons had no ties to this domain and had the same previous knowledge about the dataset.

From this dataset we extracted an 94×9 contingency table. The rows of this table represent the 94 questions that constitute the psychological test. The 9 columns represent the 3×3 possible combinations of two answers to a question, one from the parent and one from the teacher (answers to a question can be Not True, Often True, or Very True). A cell in the extracted 94×9 contingency table counts for the corresponding question how many times the corresponding parent/teacher answers combination occurred among the 300 the young patients. The aim of the analysis is to find out to which questions there was a higher degree of agreement, partial agreement or disagreement on the answers between the parents and the teachers.

For the evaluation, we conducted semi-structured interviews [6,7,8]. This is a research method which is often used in evaluation studies of information visualizations. It gives a more detailed overview of participants' attitudes to the tested software than more formal methods [9]. The design of our evaluation study was divided into four parts:

– **Introduction:** At the beginning, the structure of the datasets and the basic functions of the prototype were introduced.
– **Tutorial:** The goal of the second subpart was that the participants gained a first impression of the visualization and interacted with the visualization, so that they got familiar with it. Participants had to solve small tasks with the aim to support them to learn the basic functions. For example, one task was to change the value of the slider. After they finished the tutorial part, there was a little break with the goal to clear up misunderstandings about the basic functions.
– **Main Study:** This part included tasks which were designed to assess if the fundamental idea of the visualization was clear. Participants were asked: to merge all sectors representing answer-combinations where the parents checked "Very True or Often True"; to split them up again; to observe how the lines changed when they moved the slider; to find out, if the row-sums were all identical (they were not, differing in this from the tutorial-dataset); to select all answers where the frequency of the parent/teacher combination "Not True"/"Very True or Often True" was above the expected value; to tell us how many of these there were; and to tell us the absolute frequency of the questionnaire-item from the most outlying of the aforementioned set. The tasks were developed in cooperation with domain experts.
– **Interview:** After the subjects had finished the tasks, we asked them about the impressions that they gained during their work with the visualization. The questions used in the interview reflect the research questions described above in Section 3.

The interviews were recorded. Based on these recordings, the analysis of the data was conducted. We looked for significant statements concerning the research questions. Then we compared the subjects' answers and tried to interpret them, following the methods of Bortz and Döring [10].

5 The Evaluated System

The system used for the evaluation differs in some points from the one introduced in Section 2. We removed the interfaces for creating different Wheel configurations and presented our participants only pre-defined Wheel visualizations. Figure 3 shows a screenshot of the evaluation system. Many other features and interfaces where removed with the intent to focus on the evaluation of the basic functionalities, without distracting our participants by the multitude of configuration options.

Moreover, we adapted the existing interfaces for showing the current selection (see F in Figure 3) and the underlying contingency table (see G in Figure 3) to provide multiple views. By that we show these views side-by-side together with the Wheel visualization, and provide new functionality to facilitate the interplay between these views, e.g. to highlight the current selection in the tabular view showing the underlying contingency table.

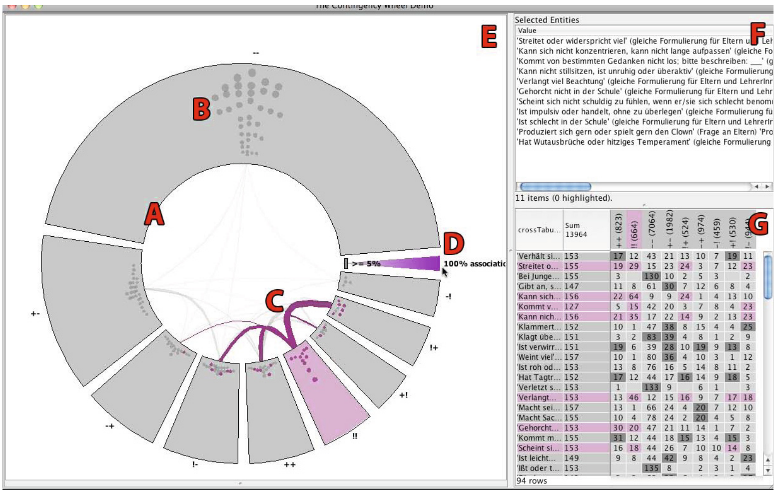


Fig. 3. The system used for the evaluation. A) sectors represent the answer-combinations from parents vs. teachers (the table columns), B) dots represent the psychological test questions and are positioned in different sectors depending on how often the sector's answer-combination appear in the database, C) lines between sectors indicate if shared questions exist in both of them, and D) is slider to filter the dots depending on an association strength. The area indicated by E) contains the Wheel visualization, F) the selection list, and G) the corresponding contingency table.

6 Interview Results

In this section we present the results of the interviews according to the research questions described in Section 3.

6.1 Advantages

Users mentioned that, compared to the tabular presentation, the wheel visualization can present large amounts of data without scrolling. They saw this as a considerable advantage. Most of the participants stated that the wheel visualization gave them a good first overview about the datasets, so they could analyze the points in the sectors more easily. They pointed out, for example, that it was easier for them to see the extreme values and outliers in the wheel visualization than in the table.

Moreover, it was mentioned that the merging and splitting functionality of sectors is useful (e.g., it was mentioned as "quick, simple and uncomplicated"). The participants also appreciated the performance of this functionality, e.g., that merging and splitting happened immediately without noticeable delay.

The interplay between the wheel visualization and the table was noted as very useful. Participants also stated that the table gave them the possibility to deal with the values in more detail (e.g., to see the information about absolute frequency). Moreover, they noted that working with the table increased their confidence in the reliability of the data and helped them to understand the meaining of the wheel visualization better from the beginning.

Furthermore, the usage of slider was noted as very helpful for progressively filtering and showing only the more outlying dots.

6.2 Disadvantages and Limitations

Although the subjects pointed out that the slider was very useful to filter out specific points, most of the participants mentioned that it was hard for them to find the slider, because of its confusing design. For example, one subject noted that s/he identified the graphical representation of the slider as sector for the remaining values and not as interaction element. Furthermore, we observed that the participants often forgot that they filtered out points. Therefore, the subjects did not interpret the visualization correctly.

Moreover, it was noted that the meaning of the different sizes and positions of the points was not clear for them at the beginning. They stated that it was confusing that information about the relative frequency was double-coded. Although the subjects found it generally difficult to understand the meaning of the connecting lines at the beginning, the double-coding of the line thickness was one reason that the participants had problems to interpret the connecting line between sectors correctly. For example, one participant noted that s/he was confused that a line was thicker although the connected sectors had fewer shared points than two other connected sectors.

Several usability issues were mentioned. For example, they missed the possibility to deselect items and it was for the participants unusual not to click on the selected item to open the context menu. Another design problem was that the points were too small if they were in small sectors or too close together so that the subjects had problems to select a point or that they were not sure if they selected the desired point. Furthermore we could observe that participants often did not notice that sectors had been merged.

6.3 Comprehension of the Concepts

Participants felt that the visualization was very complex and sometimes felt confused, especially in the beginning. Some expressed confusion due to the fact that size and (radial) position of the dots convey the same meaning and also that their angular position does not have any inherent meaning as such (it is determined by the decluttering algorithm). It was stated that after overcoming these obstacles the dots were generally understandable. In a similar way, some users understood the meaning of sectors only after a certain period of interaction with the system.

The way how the selected dots and sectors were represented in the table was often found to be confusing. For example, when selecting a single sector, sometimes table cells in a column of a different sector would be marked as selected, too. The meaning of differently colored cells in the table was rather difficult to understand.

Doubts mentioned about understanding the slider concerned the value that it was set to and its meaning. For illustration, during the experiment we observed a participant setting the slider to a value of "5%" with the intention of showing only significant dots (in a statistical sense).

Lines seemed to have been a challenging concept for our participants. They initially were not able to tell what the line connecting two sectors means. If the lines were understood they were perceived as useful. Nevertheless, the meaning of the strength/thickness of a line was not always clear. One participant expressed being confused by a situation where one line would be stronger than another, even though the pair of sectors connected by it actually showed fewer common dots. This happens because the thickness does not only depict the count of common dots, but also how highly these dots are associated with the sectors.

6.4 Usefulness of Interactions

Our participants appreciated the slider and its filter functionality. They found the feedback of the dotted line that indicated the cutoff when one moved the slider as useful. Being able to select both dots and sectors and having the related data highlighted in the contingency table was mentioned as very helpful.

7 Reinventing the Contingency Wheel

Some of the limitations we found, like the problems with deselecting the dots, are usability problems, and are very specific to the prototype we tested. But other problems we found are more interesting and seem to relate to the metaphor of the Dot-Based Contingency Wheel itself.

Based on our results, Alsallakh et al. [11] introduced a redesign of the Dot-Based Contingency Wheel to cope with the issues we found. The new design, called Contingency Wheel++ simplifies the visualization by replacing the dots with histograms along the radial dimension (Figure 4). By analyzing the limitations reported in Section 6, they found that dots are not suitable as a metaphor

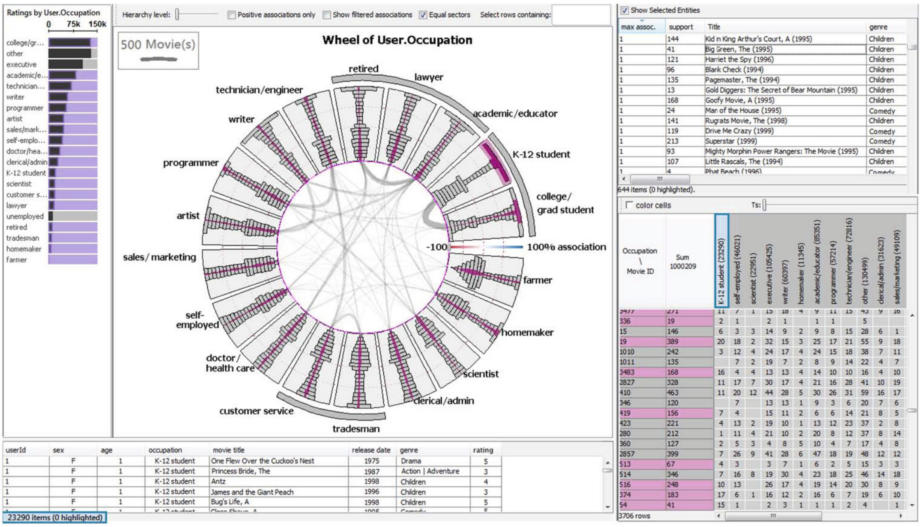


Fig. 4. The Contingency Wheel++ showing the same data as in Fig. 2 (screenshot by Alsallakh et al. [11])

to represent the table cells. They encode the deviations from the expectancy values, while the users expected to see the absolute values when they look at the dots.

Histograms are better suited to show the distribution of entities along a dimension than item-based representations. This solves the issue of what a dot represents and/or "means", and also both the problem with the double-coding of the association value (as dot radial position and size) and with the implied meaning of the angular position of the dot (which encodes no information). The angular dimension encodes the bar lengths of the histograms instead. Furthermore, when using histograms, the lines have clearer semantics than when using dots. The line thickness encodes the similarity between two distributions rather than high associations of shared dots.

The redesign of the Wheel metaphor does not associate information with sector sizes or with cell colors in the contingency-table-view by default. Other visual aids were introduced and made more salient, to address cases like forgotten slider-settings.

Finally, Contingency Wheel++ adopts the multiple views we introduced in our evaluation prototype (Section 5). Our participants found the multiple views connected by brushing-and-linking mechanisms very useful. This could confer benefits not only to users relatively new to the system but also to experts [11].

8 Conclusions

According to our evaluation study, we found that the users appreciated the utility of the Contingency Wheel to gain an overview of large contingency tables and to quickly find extreme values in them. Despite some usability issues, the users also appreciated the abilities to filter and interact with the data, and the combination of the Wheel visualization and the contingency table side-by-side.

The more critical issues we found with the Dot-Based Contingency Wheel are related to the design of its visual metaphor, in particular the dot metaphor to represent table cells. The dots were not straightforward to understand, and sometimes caused the users to draw wrong conclusions about the data. The lines between the sectors were also difficult to understand, mainly because their interpretation was related to the dots.

Subsequently, the presented system was redesigned by replacing the dots with histograms [11]. Histograms serve as abstraction of the dots, and can hence resolve the major issues with the original design. We also recommend such aggregations to deal with large amount of data in similar systems. Histograms are familiar representations that offer an effective alternative to simplify cluttered item-based representations.

Our work demonstrates the value of qualitative evaluation methods, especially for evaluating the first design of a new visual-analytics technique. They help not only in finding usability issues, but also in quickly assessing the clarity of the conceptual design of this technique. This enables spotting major issues with this design, and provides valuable guidance to iteratively refine the design, before conducting a thorough quantitative evaluation.

The interview study is part of a larger evaluation study. We also asked the users to think aloud during their interaction with the prototype. Moreover, we captured the activities of the users on the screen and recorded log files for a detailed analysis. One of our next steps is to substantiate the findings from the interview study with this data and carry out additional investigations on cognitive strategies adopted by the users.

Acknowledgments. This work is conducted in the context of the Centre of Visual Analytics Science and Technology (CVAST). It is funded by the Austrian Federal Ministry of Economy, Family and Youth in the exceptional Laura Bassi Centres of Excellence initiative. We thank Susanne Ohmann and Christian Popov for providing the medical data set and for the assistance in formulating the tasks.

References

1. Hartigan, J., Kleiner, B.: Mosaics for contingency tables. In: Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface, New York, vol. 22, pp. 286–273 (1981)
2. Bendix, F., Kosara, R., Hauser, H.: Parallel sets: visual analysis of categorical data. In: Proc. of the IEEE Symposium on Information Visualization, pp. 133–140 (2005)
3. Benzécri, J.P.: Correspondence Analysis Handbook. Marcel Dekker, New York (1990)
4. Alsallakh, B., Gröller, E., Miksch, S., Suntinger, M.: Contingency wheel: Visual analysis of large contingency tables. In: Proceedings of the International Workshop on Visual Analytics, EuroVA (2011)

5. Carpendale, S.: Evaluating information visualizations. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 19–45. Springer, Heidelberg (2008)
6. Courage, C., Baxter, K.: Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques. Morgan Kaufmann (2005)
7. Wilson, C.: User Experience Re-Mastered: Your Guide to Getting the Right Design. Morgan Kaufmann (2009)
8. Stone, D., Jarrett, C., Woodroffe, M., Minocha, S.: User Interface Design and Evaluation User Interface Design and Evaluation. Morgan Kaufman (2009)
9. Adams, A., Cox, A.: Questionnaires, in-depth interviews and focus groups. In: Cairns, P., Cox, A. (eds.) Research Methods for Human-Computer Interaction, pp. 17–34. Cambridge University Press, Cambridge (2008)
10. Bortz, J., Döring, N.: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4th edn. Springer, Heidelberg (2006)
11. Alsallakh, B., Aigner, W., Miksch, S., Gröller, E.: Reinventing the contingency wheel: Scalable visual analytics of large categorical data. IEEE Transactions on Visualization and Computer Graphics, Proceedings of IEEE VAST 2012 18(12), 2849–2858 (2012)