Common Angle Plots as perception-true visualizations of categorical associations

Heike Hofmann, Marie Vendettuoli

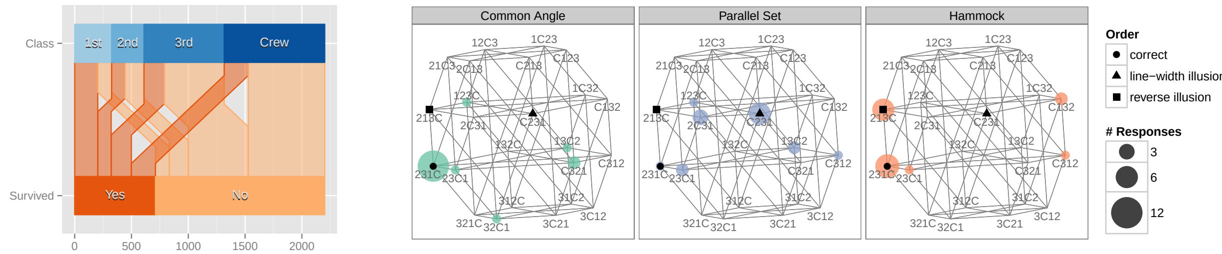

Fig. 1. Who survived on the Titanic? Common Angle plot of survival by class (left). On the right, graphs showing survey results with answers.

Abstract— Visualizations are great tools of communications - they summarize findings and quickly convey main messages to our audience. As designers of charts we have to make sure that information is shown with a minimum of distortion. We have to also consider illusions and other perceptual limitations of our audience. In this paper we discuss the effect and strength of the line width illusion, a Müller-Lyer type illusion, on designs related to displaying associations between categorical variables. Parallel sets and hammock plots are both affected by line width illusions. We introduce the common-angle plot as an alternative method for displaying categorical data in a manner that minimizes the effect from perceptual illusions. Finally, we present results from user studies as evidence that common angle charts resolve problems with the line width illusion.

Index Terms—Linewidth illusion, Data Visualization, High-dimensional Displays, Parallel Sets, Hammock Plots, Müller-Lyer Illusion.

1 Introduction

A well-designed graph is a powerful tool that transcends barriers of language to communicate complex concepts from author to audience. Problems arise, when readers are unable to easily extract a chart's main message or are led to wrong conclusions due to distortions. This endangers the trust between readers and creators of charts, which is based on the main premise that graphics have to be true to the data [28, 29, 25]. There is a lot of discussion on keeping true to the data in the framework of (ab)using three dimensional effects in graphics. Tufte [28] goes as far as defining a *lie-factor* – the ratio of the size of an effect in the data compared to the size of an effect shown. Any large deviation of this factor from one indicates a misuse of graphical techniques. Computational tools help us ensure technical trueness – but this brings up the additional question of how we deal with situations that involve innate inability or trigger learned misperceptions. One example of distortions of this kind is the Müller-Lyer family of illusions, which include contextual illusions, such as differently perceived lengths of line segments depending on the orientation of arrow heads or the sine illusion [9].

Regardless of the cause of distortion, it is the responsibility of the author of a chart to create visualizations that allows readers to extract an accurate interpretation of the underlying data. In order to gauge the extent of distortion due to perceptual limitations, we can employ user studies to provide empirical evidence supporting underlying cognitive models or previously unknown or not anticipated illusions.

Parallel sets (parsets) [20] are a graphical method for visualizing multivariate categorical data. Since their initial publication, parallel sets have spread to mass media outlets [19, 4, 2], have been implemented in various languages [19, 3, 6] and are a reputable resource for further academic work ([20] has 70 citations per Google scholar). While retaining the ability to visualize a large number of dimensions simultaneously that is the parallel coordinates' hallmark trade, parallel sets combine with it a frequency scale that is a well-known feature of other categorical displays such as barcharts or mosaic plots [13, 11, 15, 27]. Unfortunately, the parallel set plot is a victim to distortion due to a contextual illusion: consider the parset plot of Figure 2. This plot shows the relationship between class status and survival

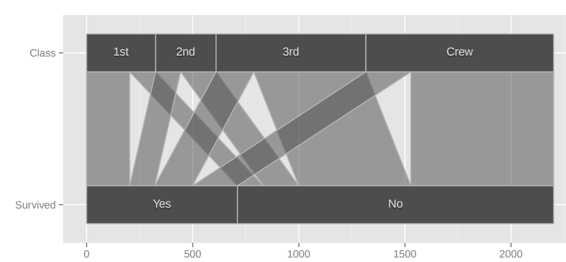

Fig. 2. Parallel sets plot showing the relationship between survival of the sinking of the HMS Titanic and class membership.

on board the HMS Titanic [7]. Class status is recorded as either crew member or passengers in first, second, or third class. The top bar in figure 2 shows the variable Class. The bottom bar shows survival as yes and no. Lines are drawn between top and bottom bar – the (horizontal) width is proportional to the number of survivor and non-survivor they represent. A reasonable task based on this chart is to order levels of the variable 'class' by number of survivors. However, when study participants were asked to perform this task, only 12.5% respondents

- *Heike Hofmann is an associate professor in Statistics and faculty member in the Human Computer Interaction program at Iowa State University, E-mail: hofmann@iastate.edu.*
- *Marie Vendettuoli is a graduate student at Iowa State University, with majors in Human Computer Interaction and Bioinfomatics & Computational Biology. She is a past IGERT Fellow and currently interns with the Statistics section at USDA. E-mail: mariev@iastate.edu.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 27 September 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

	Crew	1st	2nd	3rd
Survivors	212	203	118	178
Non-Survivors	673	122	167	528

Table 1. Correct ordering of variable Class is: crew, first class, third class, followed by second class.

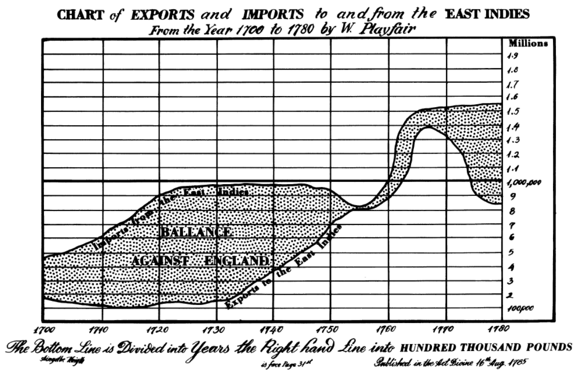

Fig. 3. Playfair's chart from the Commercial and Political Atlas (1786) showing the balance of trade between England and the East Indies. In which years was the difference between imports and exports the highest?

selected the correct order, see table 3.

We believe that readers view parallel sets they are subject to the *line width illusion*, a perceptual distortion that we describe and quantify in this paper. We also propose and test *common angle plots*, an alternative graphing method for visualizing multivariate categorical is not subject to the *line width illusion*.

2 Line width illusion

An example of the *line width illusion* is displayed in figure 3. This chart displays the balance of trade between England and the East Indies as shown by William Playfair in his Commercial and Political Atlas, 1786 [22, 23]. One purpose of this chart is to demonstrate the difference between imports and exports in a particular year and its pattern over that time frame. The difference in exports and imports is encoded as the vertical difference between the lines. When observers are asked to sketch out the difference between exports and imports [5], they very often miss the steep rise in the difference between the lines in the years between about 1755 and 1765. Figure 4 shows the actual difference between imports and exports.

This phenomenon is known and widely discussed in statistical graphics literature [5, 28, 29, 25]. It is due to our tendency to assess distance between curves as the minimal (orthogonal) distance rather than the vertical distance – see sketch 5 for a visual representation of

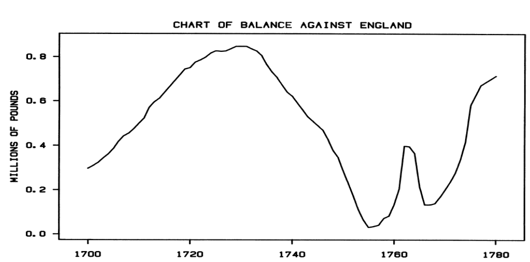

Fig. 4. Difference between exports and imports from England to and from the East Indies in the 18th century – the steep rise in the difference around 1760 comes as a surprise to many viewers of the raw data in figure 3.

both.

In the perception literature, this phenomenon is known as part of a group of geometrical optical misperceptions of a context-sensitive nature classified as Müller-Lyer illusions [9]. Interestingly, there seems to be a general agreement that this illusion exists, but a quantification of it is curiously absent from literature.

The type of chart as shown in figure 3 proposed by Playfair is shown quite commonly, particular in election years – where these kind of charts are used to enable comparisons of support for several candidates, the recommendation from literature is to avoid charts in which the audience is asked to do visual subtractions, and show these differences directly.

2.1 Strength of line width illusion

When visually evaluating lines of thickness greater than one, the line width illusion applies, only now the *edges* of a single line take on the role of the separate curves. As above, there is a strong preference of evaluating the width of lines orthogonal to their slopes as opposed to horizontally (see figure 5) needed for a correct evaluation of parallel sets-style displays.

Orthogonal w_o and horizontal w_h line widths are related – the orthogonal line width depends on the angle (or, equivalently, the slope) of the line:

$$w_o = w_h \sin\theta, \tag{1}$$

where θ is the angle of the line with respect to the horizontal line.

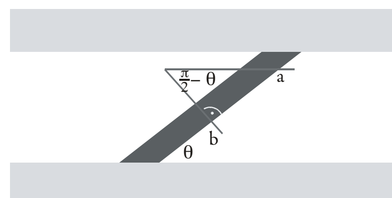

Fig. 5. Sketch of line width assessments: (a) is showing horizontal width, (b) shows width orthogonal to the slope. Survey results in section 5.2 indicate that observers associate line width more with orthogonal width (b) than horizontal width (a).

The perceived slope of a line very much depends on the aspect ratio of the corresponding plot – changing the height to width ratio of a display will change our perception of the corresponding line widths, if they are not adjusted for the slope [5]. This finding is not new, but its strength on our perception is surprising, as can be seen in the example of figure 6. Again, survival and class membership on the Titanic is shown; the same parallel sets plot is shown twice in this figure, but with very different aspect ratios: in the plot on the left the number of surviving 3rd class passengers seems to be about twice as big as the number of survivors among crew members, whereas in the plot on the right the lines have about equal (orthogonal) width. Obviously, this is not due to a change in numbers.

For parallel sets-style displays, the audience has *area of the line segment* an alternate visual cue when evaluating frequencies. Because height (or width for a rotated display) of line segments is constant across the display, the width of a particular segment is proportional to its area. We can therefore employ area comparisons as a proxy or to augment line width evaluations. However, existing literature suggests that this method of comparison is particularly prone to errors in two scenarios commonly seen in parallel sets: (1) extreme aspect ratios of the rectangular shape [14] and (2) when comparing rectangles rotated relative to each other [18]. This incorrect perception and comparison of areas distorts the message readers discern from the graph.

3 Related work

Hammock plots, introduced by M Schonlau in [26], provide an alternative to parallel sets that is adjusted for the line width illusion. This

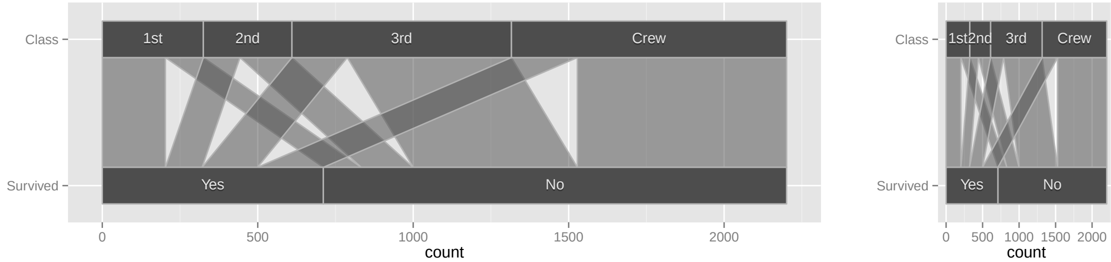

Fig. 6. Parallel sets plots of survival on the Titanic by class. Different aspect ratios seemingly change the thickness of line segments, compare e.g. number of survivors in 3rd class and in the crew.

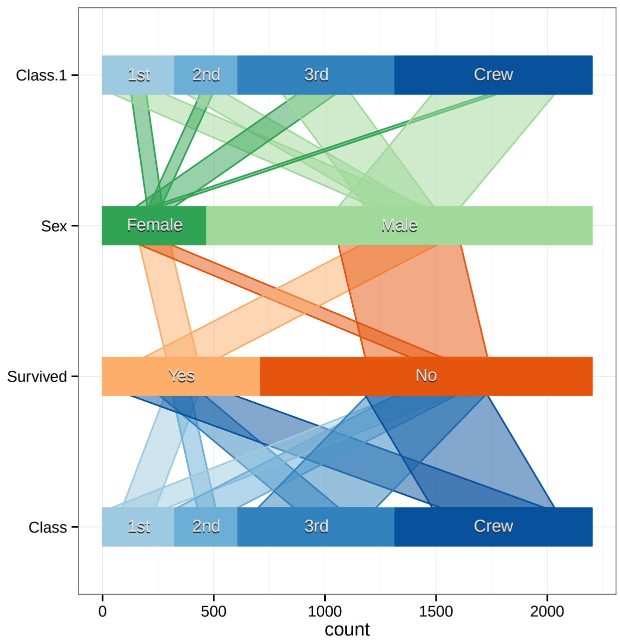

Fig. 7. Hammock plot of the relationship between Class and Survival on the Titanic.

is done by adjusting the –horizontal– line width by a factor of $\sin\theta$, as discussed in equation (1). This adjustment makes the perceived –orthogonal– line width to be proportional to the number of observations it represents. Figure 7 shows an example of a four dimensional hammock plot of the Titanic data. From top to bottom Class, Gender, Survival, and again Class are shown.

Similarly to the parallel sets plot, the bars are divided according to class membership numbers. Lines connect categories between neighboring variables, orthogonal line widths are representing the number of individuals in each combination. Unlike the parallel sets, the lines start from the middle of the bin and connect to the middle of the other variable's bins. This convention is in part due to the fact that the sum of horizontal widths (w_h) after adjustment is greater than the width of marginal bars.

The graph shows that barely any women were in the crew, while male crew members make up the second largest contingent overall. While overall a few more men survived than women, proportionally the situation is much different – a much higher percentage of women survived than men. While more first class passengers survived than not, the survival chances of second class passengers were evenly divided. For third class passengers and crew members fewer members did survive than not.

As the adjustment of line widths is made with respect to the angle θ, which itself depends on the aspect ratio of a plot, we need complete control over these properties of the plotting device when constructing hammock plots – in our implementation (see below for details) we have dealt with this issue by fixing the aspect ratio. This is problematic in some situations, where the rendering has to be done without knowledge of the plotting device.

3.1 Reverse linewidth

A problem that arises in evaluating hammock plots is that if an observer focuses on horizontal line width the plots suffer from a *reverse line width illusion*: judging the number of survivors by class in figures 7 and 8 based on horizontal line width results in an ordering of (largest to smallest) Crew, 3rd, 1st, and 2nd – which is not correct either. Because the lines are centered around the middle of each level, a contextual coordinate system is imposed that encourages comparisons of horizontal width. However, horizontal width is no longer proportional to underlying data, because of the line width adjustment. Rearranging equation 1,

$$w_h = w_o \csc\theta, \tag{2}$$

where w_o is proportional to observations and θ is the angle of the line with respect to the horizontal line.

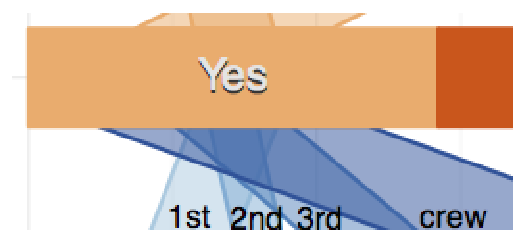

Fig. 8. Lines in hammock plot of Titanic data for survival variable, level yes. Comparing horizontal widths suggests that a greater number of survivors were from third class instead of first, which is inconsistent with underlying data.

Further supporting the poor context is the issue that the bands (unlike parallel sets) are no long proportional in area to the underlying data. Previous work [14, 18] has shown that audiences experience perceptual distortion when comparing (1) extreme aspect ratios of the rectangular shape and (2) when comparing rectangles rotated relative to each other, both of which apply to hammock plots.

4 Common Angles

Figure 9 shows a common angle plot of the same data as the hammock plot.

As in the previously discussed display types ribbons are drawn between categories with widths that are proportional to the number of records they represent.

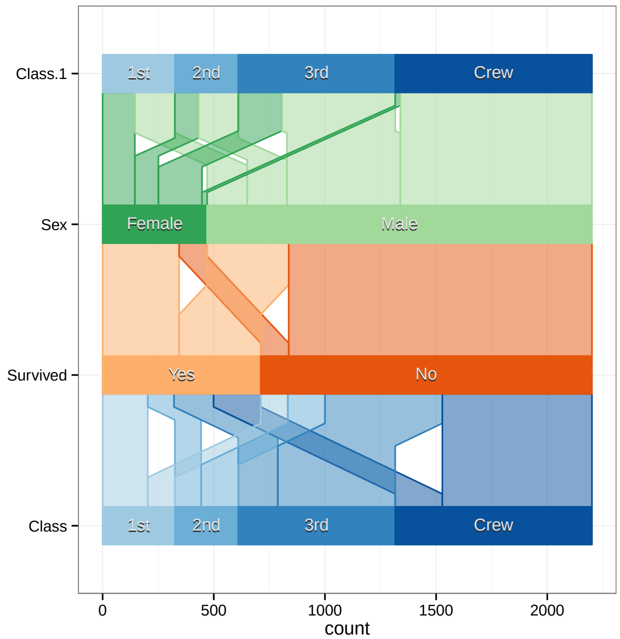

Fig. 9. Common angle plot of the Titanic data.

In order to ensure that widths of all bands are comparable without any distortion, their slopes are artificially made the same in the following manner: assuming a vertical display as shown in figure 9, connecting bands between categories are a combination of a vertical segment, a segment under a pre-specified angle θ, followed by another vertical segment. The pre-specified angle θ (between the line and the vertical band) is given as –at most– the angle of the longest connecting line between two categories of neighboring variables. This makes the width of ribbons comparable without being affected by the distortion, as all ribbons are sharing at least one segment under the same angle.

Common angles, plus the related methods of hammock plots and parallel sets are implemented in the package `ggparallel` based on the `ggplot2` [30] plotting framework in the software R 2.15.1 [24]. The ggparallel package is freely available from CRAN (http://www.r-project.org/). The colors for the plots have been chosen using color schemes from the ColorBrewer project [12], as implemented in the R package RColorBrewer [21].

5 Usability Testing

5.1 Test

To determine the effectiveness of the common angle plot, we conducted a user study in form of a survey asking participants to provide responses regarding the structure in two data sets with predominantly categorical variables. The Titanic data includes class, sex, age, and survival status for each person on board of the Titanic [8]. The gene data was retrieved from the UCSC Genome Browser [17] and includes chromosome location for genes involved in one of three metabolism pathways: steroid biosynthesis, caffeine metabolism and drug metabolism. For each data set, participants were asked to provide responses for three tasks that analysts routinely perform as part of exploratory data analysis:

Task I: simple comparison task, chosen to be unaffected by any illusion. Performance on this task should be comparable across designs.

Task II: simple ordering, involving three pairwise comparisons, some of which are affected by the line width illusion or its reverse.

Task III: more complex ordering task with at least six pairwise comparisons, some of which are affected by either illusion.

Each participant was presented with two of the three types of displays. For each display, the participant was asked to complete each of the three tasks for each data set. All participants were evaluated using the same set of questions with multiple choice options as detailed in Appendix 8, regardless of display type or order. Participants were all shown the Titanic data first, then the gene data. This resulted in a crossover design as shown in Table 2 allowing for comparisons of display types and tasks while simultaneously adjusting for individuals' different skill sets and learning effect.

Titanic Data	PS	CA	H	PS	CA	H
Genes Data	CA	PS	PS	H	H	CA
#responses	8 (9)	6 (7)	8 (9)	6 (7)	10 (11)	8 (8)

Table 2. Overview of study design and participation numbers. The number in parenthesis indicates the number of participants completing the first block, but not the second.

5.2 Results

We are investigating three different aspects of the experiment in this section: first, assessing general performance on the tasks according to percentage of correct responses, second, investigate the extent of variability due to subject-specific abilities, and finally exploration of the space of answers for the more complex ordering Task III.

Correctness of Answers

Answers for each survey question were recoded in binary form according to correctness (with 1 for correct answers, and 0 otherwise). This forms the basis for the evaluation of performance of the different designs.

Table 3 shows percentages of correct answers for each question under each design. Bold numbers indicate significantly different (worse) performance of a design compared to the common angle plot based on a generalized linear model with random effects to adjust for individuals' abilities. The model explains 69.5% of the total variability, corresponding to a highly significant deviance of 435.8 (p-value $\ll 0.0001$).

The observed results are in line with our expectations: as we aimed for, task I does not show any significant differences between the designs and has overall the highest percentage of correctness reflecting its low difficulty level. Generally, difficult levels seem to increase with complexity of the tasks.

parallel sets were affected the most by the line width illusion and show significantly worse performance for tasks II and III in the Titanic data, and for task III in the genes data. The performance on task II in the genes data is borderline non-significant, but shows a negative trend. Hammock plots led to significantly worse performance than common angle plots in the two questions that were affected by the inverse line width illusion, while they show equal performance as common angle plots for the other questions. For task III in the genes data, hammock plots have the overall best performance across designs– but this does not constitute a significant improvement over the performance of the common angle plot. Figure 10 gives an overview of performance of each design on all tasks.

Individuals' skill levels

Figure 11 shows an overview of the predicted skill for each participant under the model. Skills are quite varied between -1.52 and 1.34, but a Kolmogorov-Smirnov test does not show significant deviation from a normal assumption (p-value 0.089). On the scale of the dependent variable the range in individuals' skills translates to a $17.5 = e^{1.34-(-1.52)}$ fold increase in the probability of answering a question on the survey correctly between participants with the best skill set and the worst.

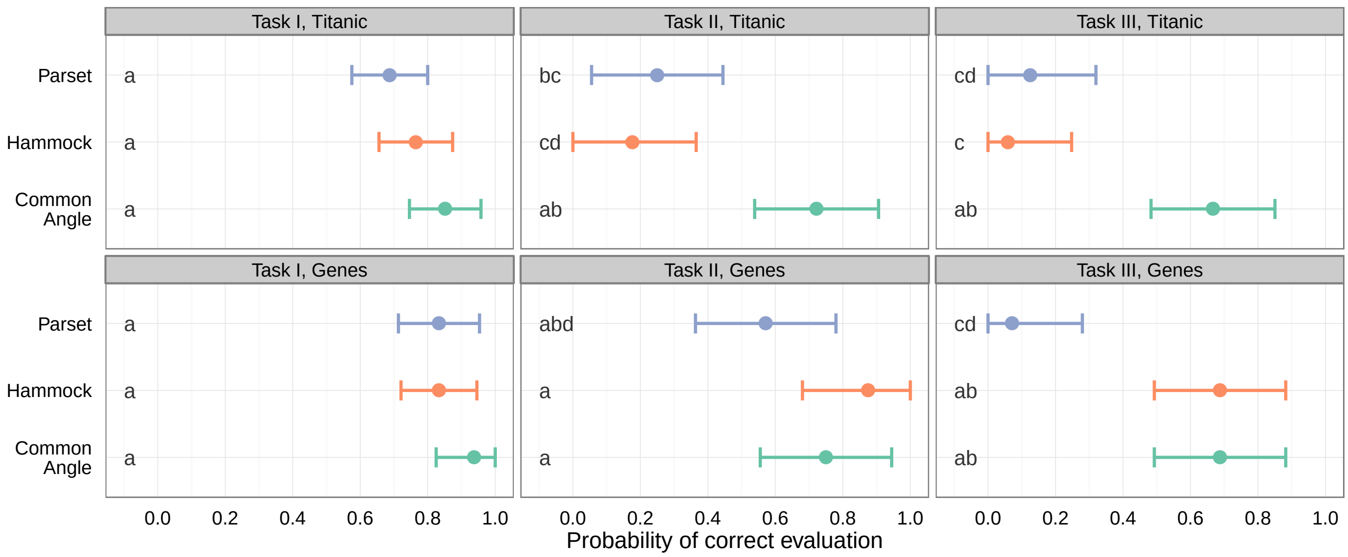

Fig. 10. Overview of performance across tasks and designs. Points show average performance of subjects on each of the tasks, lines represent 95% confidence intervals adjusted for multiple comparisons. The letter at the front of each panel allow for an evaluation of significance of pairwise comparisons: if two averages do not share a letter, they are significantly different at a level of 0.05.

Task	Data	Design		
		CA	H	PS
I	Titanic	85.2 (0.66)	76.5 (0.84)	68.8 (0.98)
	Genes	93.8 (0.51)	83.3 (0.78)	83.3 (0.90)
II	Titanic	72.2 (2.56)	**17.6** (2.31)	**25.0** (2.80)
	Genes	75.0 (2.80)	87.5 (2.13)	**57.1** (3.67)
III	Titanic	66.7 (2.69)	**5.9** (1.43)	**12.5** (2.13)
	Genes	68.8 (2.99)	68.8 (2.99)	**7.1** (1.91)

Table 3. Percentages (standard deviation) of correct responses for each task and design. Bold numbers indicate significant difference from common angle plot performance.

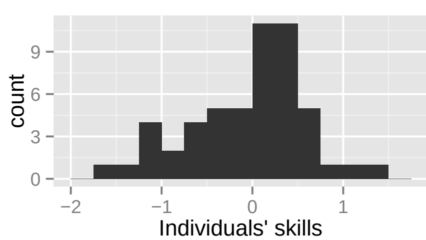

Fig. 11. Histogram of the predictions of subject-specific skills.

Evidence for line width illusions

Task III for the Titanic data required participants to order class levels according to the number of survivors, fewest to highest.

There are 4! = 24 distinct orderings of the levels, corresponding to all permutations of length four. Some orderings are closer to one another than other orderings. The Cayley distance allows us to quantify this distance: the Cayley distance between two orderings is defined as the smallest number of switches necessary to get from one ordering to the other. Visually, this corresponds to a graph: each node represents one ordering, and two nodes are connected by an edge, if only a single switch is necessary to move from one ordering to the other, i.e. if the Cayley distance between these nodes is one. This results in a regular graph of degree six, i.e. each node is connected to six other nodes. Between any two nodes, the Cayley distance on the graph is equivalent to the length of the shortest connecting path between the two nodes.

Figure 12 shows an overview of the permutation space together with an overview of the survey results.

The colored dots on top of the graph correspond to the responses from the survey. The size of these dots is proportional to the number of observers choosing this particular ordering. It becomes obvious from the three graphs in figure 12 that the answers to different designs occupy quite different regions, while answers based on the same design are quite close – usually separated by only one edge.

The correct ordering, as well as the orderings assuming the line width illusion and its reverse are marked by symbols. Answers for the common angle plot are centered around the correct answer, while responses to parallel sets cluster around the response corresponding to the line width illusion. Due to the smaller number of responses to the hammock design a clear clustering of the answers is not recognizable, but the answer for the inverse line width illusion is among the responses. Table 4 gives an overview of all responses to task III for the Titanic data.

Order	CA	H	PS	
Crew, 1st, 3rd, 2nd		2		
2nd, 1st, 3rd, Crew		6		inverse illusion
Crew, 3rd, 1st, 2nd		1	1	
2nd, 3rd, 1st, Crew	12	7	1	correct
2nd, 3rd, Crew, 1st	1	1	2	
Crew, 3rd, 2nd, 1st	2			
3rd, 2nd, Crew, 1st	1			
1st, 2nd, 3rd, Crew	1		1	
1st, 3rd, Crew, 2nd	1		2	
Crew, 2nd, 3rd, 1st			6	line width illusion
2nd, Crew, 3rd, 1st			3	
Total	18	17	16	

Table 4. Responses to task III in the Titanic data: order levels of Class by the number of survivors (smallest to largest).

Common angle plots show the best performance in terms of correctness (66.7% on 18 responses), compared to a correctness of 12.5% for parallel sets plots on 16 responses, constituting a significantly better performance of common angle plots at a level of 0.0027, based on a Mantel-Haenszel test (the difference in performance to hammock plots is also significant at a level of 0.0006; but there is no significant difference in correctness between hammock plots and parallel sets.). While the intuitive assessment of lines by their width orthogonal to their direction is well known, it is surprising to see its strength: in this particular setting, it is strong enough to 'shrink' the horizontally widest line for six out of 16 participants by at least 44%, from 212 to below 118, and a further three participants perceived a shrinkage to below 178, a distortion factor of at least 16%.

Opinion on common angle plots

Answers to the question of 'which chart did you like better?' are shown in table 5. There is a clear endorsement in favor of common angle plots versus the other two types of displays. The most common reason cited for the choice was a facilitated comparison of width, area or "size", The only consistent complaint against common angle was a preference for straight lines. This purely aesthetic preference is deeply rooted and in our opinion the biggest challenge for common angle plots.

			Which chart did you like better?	
			Chart 1	Chart 2
PS	vs	**CA**	2	**6**
CA	vs	PS	**4**	2
H	vs	**CA**	3	**5**
CA	vs	H	**8**	2
H	vs	PS	3	5
PS	vs	H	1	5

Table 5. Preferences for first or second chart across all six combinations of questions and chart types.

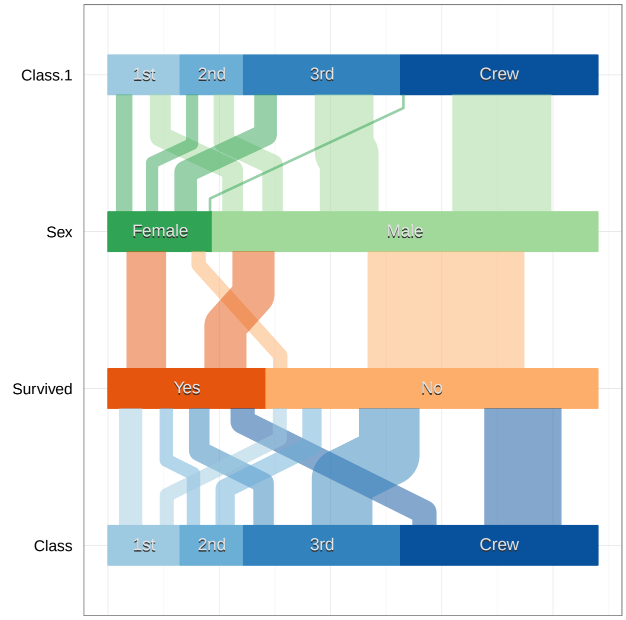

Fig. 13. Common angle plot of Titanic data using hammock correction.

5.3 Methods

The survey was created using the Qualtrics Labs, Inc software (www.qualtrics.com). For survey contents, see Appendix 8. The study design is presented in section 5.1. All models are fit in the `lme4` package [1] within the software frame work of `R` 2.15.1 [24]. Comparisons are adjusted for multiple testing using the `multcomp` package [16] and evaluated for pairwise significances using the `effects` package [10].

6 Discussion

For data with a large number of variable levels, common angle plots may introduce more line crossings than hammock plots, while the number of crossings is the same between parallel sets and common angles. This may affect the effectiveness of the overall display. Use of color may also be problematic with many variable levels, as readers may have difficulty resolving a palette of colors separated by small intervals. Certainly, resolving many colors is difficult for audience with a history of color blindness and may be technically limited in print applications. The issue of color is consistent regardless of display choice between parallel sets, hammock plots or common angles. Further study is necessary to resolve the potential for distortion in different application, especially the impact of bands displayed with extreme values for θ, a known source of perceptual inaccuracy [14].

One opportunity for improvement lies in the algorithm to determine thickness of the connecting line. In the tested version of common angle plots, the line width was not explicitly defined - the line width is a byproduct of specified θ for a band that connects marginal bars. Using equation 1 to determine the band thickness while keeping θ consistent across all bands is a common angle-hammock plot hybrid that bears further investigation. A drawback of this approach is that sum of ribbon w_h will no longer match widths of marginal bars. This may create a additional processing burden placed on the audience to map the relationship between band width and the with of marginal bars. Both hammock plots and this modification of common angles face the issue of band *area* as context to support reader interpretation of w_o. Since the band area is now related to the incident angle θ, changes in the display aspect ratio may have a distortion effect. In the study described in this paper, this effect was not evaluated and aspect ratio was kept constant.

In the original paper, parallel sets were introduced to reflect a hierarchy of variables. Prior examples in this paper show sets of two-dimensional plots to focus on the association between pairs of variables. With color coding, it is possible to show hierarchies in all of the types of displays. Figure 14 shows a common-angle plot with a hierarchy: survivors of the disaster are marked in blue, non-survivors by orange. From top to bottom of the plot a hierarchy is drawn, considering first survival, then gender, followed by age and finally class membership. The coloring tracks survival status throughout the hierarchy, the layout in a common angle plot makes comparisons valid across all levels. This is of particular importance in hierarchical displays, which by definition have a larger number of smaller groups than displays without a hierarchy exacerbating problems induced by the line width illusion.

Another opportunity for extending common angle plots is to add interactivity. It is important to note that any additions of functionality via interactivity should not come at the expense of developing distortion-free displays. Simply augmenting a plot that has distortion of the *line width illusion* variety does not eliminate the presence of that illusion, rather it obfuscates the message of the display. A display with visual cues in conflict with the interactive feedback introduces cognitive load by asking the audience to decide which source is correct. In the case where interactive feedback (e.g. summary data on mouse hover) is accurate when the visual cues suggest alternate interpretation, a mistrust of the graphical presentation may develop. In an extreme case, the user may develop a mistrust of their own perception, which would reduce effectiveness of all data displays, regardless of the presence of any perceptual distortion. In the alternate case, where audience chooses to rely on perception over interactive feedback, it is possible that distortions in displays that are part of initial exploratory analysis may lead to research choices that are unsupported by data. This is both a waste of resources and, when human or animal subjects are involved, may lead to ethical violations.

7 Conclusion

We have proposed a new chart type for visualizing multivariate categorical data, common angle plots, and tested its usability compared to existing charts that perform a similar function. Results from user testing indicate that common angle plots effectively communicate underlying data without encouraging perceptual distortion of the *line width*

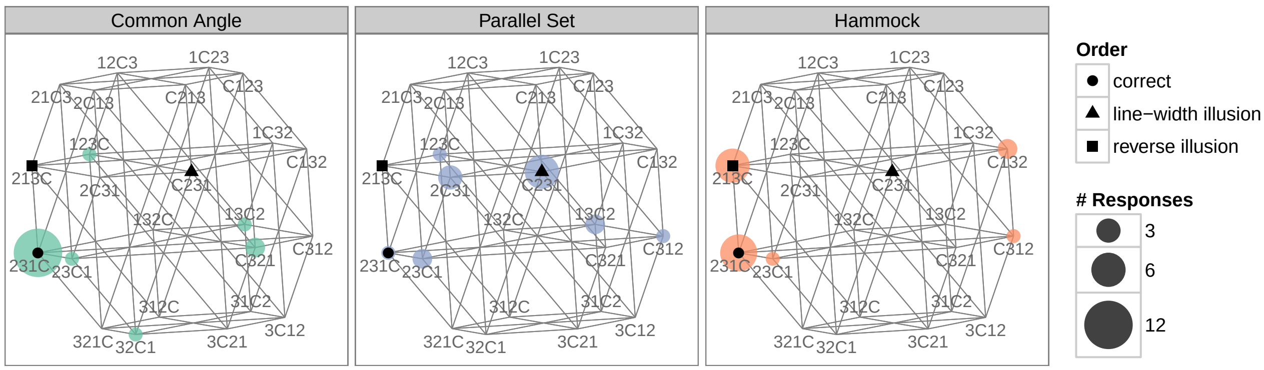

Fig. 12. Answers to task III in the Titanic data – each node corresponds to a single ordering of the levels in variable 'Class'. Lines are drawn between orderings that are only one swap of levels apart. The colored dots show responses from the survey, their sizes depend on the number of responses for each ordering.

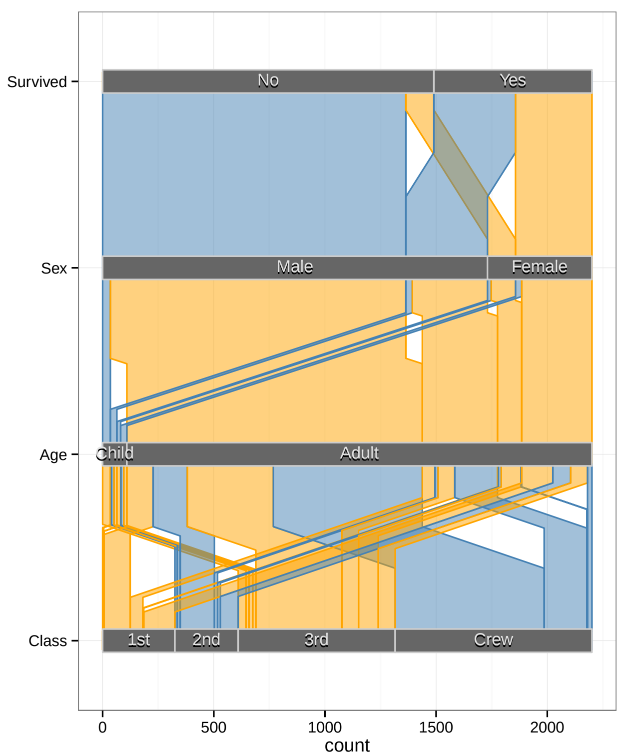

Fig. 14. Common angle plot of the Titanic data using a hierarchical structure in the variable (cf. to parallel sets chart in [6]).

illusion. Two other chart types which address visualization of multivariate categorical: parallel sets and hammock plots, are subject to the line width illusions due to contextual framework. Audiences perceive parallel sets with distortion due to a natural tendency to evaluate line width in the orthogonal direction while data is mapped to the horizontal width. For hammock plots a correction is made to map data to the orthogonal width, however the centered line intersection with axes creates a strong contextual encourage evaluation of line width using the horizontal measure (*reverse line width illusion*). Common angles avoids the perceptual distortion associated with either version of the illusion regardless of the underlying data set.

8 Survey

At the survey start, participants were presented a brief tutorial regarding the different plot types. The tutorial can be found at `http://mariev.net/tutorial.html`

The survey consists of two blocks of questions each pertaining to one data set (Titanic data or gene data). Each block was presented with a single plot to use as reference when responding. Two different plot types were shown for the two different blocks, yielding a total of six unique orderings of plot types as shown in table 2.

Participants were randomly assigned to one of these six combinations. This study design structure was imposed, in part, to encourage participation by reducing the amount of time for survey completion. Completion of all survey questions was anticipated to take 10 - 15 minutes. No personally identifiable information was collected, nor was any compensation offered.
The questions pertaining to the Titanic data were:

Task 1: *Agree, Disagree or Don't Know/Can't Determine with the following statements:*

- There were an approximately equal number of Male and Female Survivors
- The group with largest number of travelers was Female Survivors
- There were more Male Non-Survivors than number of males in First and Second Class Combined

Task 2: *Order the following groups by number, fewest to most*

- 1st Class female passengers
- Male Survivors
- Crew Survivors

Task 3: *Order the categories of Class by number Survived, fewest to most.*

- 1st
- 2nd
- 3rd
- Crew

The questions pertaining to the gene data were:

Task I: *Agree, Disagree or Don't Know/Can't Determine with the following statements:*

- There are about the same number of genes in the group "steroid biosynthesis:chromosome 1" as in the group "caffeine metabolism: chromosome 8"
- The group with the greatest number of genes is "drug metabolism:chromosome 4"
- there are more genes involved in the group "drug metabolism: chromosome 1" than all genes involved in the caffeine metabolism pathway

Task 2: *Order the following chromosomes by number of genes involved, fewest to most.*

- steroid biosynthesis :: chromosome X
- steroid biosynthesis :: chromosome 4
- drug metabolism :: chromosome X

Task 3: *Order the following chromosomes by number of genes involved in steroid biosynthesis pathway, fewest to most.*

- chromosome 1
- chromosome 4
- chromosome 8
- chromosome X

9 Participants' demographics

All students, staff and faculty from Iowa State University programs in Statistics, Bioinformatics and Computational Biology and Human Computer Interaction were invited to participate by email. 93 individuals accessed the survey; 86 participants gave consent, 15 of those dropped out right after, 20 went to the training site and did not return. Out of the remaining 51 participants, 46 individuals submitted responses for all questions and five gave responses to the first block of questions.

Participants used their own personal computing devices to access the survey, a majority of participants used Intel Mac OS X (versions ranging from 10.6.8 to 10.8.2), while Windows was the next most common operating system. The preferred choice of browser was Firefox, followed by Chrome. For two participants, the Qualtrics survey software was unable to capture operating system or browser information.

Acknowledgments

The survey for this study was carried out with approval from IRB-ID 12-204.

References

[1] Douglas Bates, Martin Maechler, and Ben Bolker. *lme4: Linear mixed-effects models using S4 classes*, 2012. R package version 0.999999-0.
[2] Michael Blastland. Go figure: How to understand risk in 13 clicks, March 11 2009.
[3] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.
[4] Shane Carter and Mike Bostock. Over the decades, how states have shifted, Oct 15 2012.
[5] William S. Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):pp. 531–554, 1984.
[6] Jason Davies. Parallel sets.
[7] Robert J. Dawson. The 'unusual episode' data revisited. *Journal of Statistics Education*, 3, 1995.
[8] Robert J. MacG Dawson. *Journal of Statistics Education*, 3, 1995.
[9] Ross H Day and Erica J Stecher. Sine of an illusion. *Perception*, 20:49–55, 1991.
[10] John Fox. Effect displays in r for generalised linear models. *Journal of Statistical Software*, 8(15):1–27, 2003.
[11] M Friendly. Visualizing categorical data: Data, stories and pictures. In *SAS User Group Conference*, 1992.
[12] Mark A. Harrower and Cynthia A. Brewer. ColorBrewer.org: An Online Tool for Selecting Color Schemes for Maps. *The Cartographic Journal*, 40(1):27–37, 2003.
[13] JA Hartigan and B Kleiner. Mosaics for contingency tables. In *Proceedings Symposium on the Interface*, 1982.
[14] J Heer and M Bostock. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *CHI 2010: Visualization*, 2010.
[15] H Hofmann. Exploring categorical data: Interactive mosaic plots. *Metrika*, 2000.
[16] Torsten Hothorn, Frank Bretz, and Peter Westfall. Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363, 2008.
[17] WJ Kent, CW Sugnet, TS Furey, KM Roskin, TH Pringle, AM Zahler, and D Haussler. The human genome browser at UCSC. *Genome Research*, 12(6):996–1006, June 2002.
[18] N Kong, J Heer, and M Agrawala. Perceptual guidelines for creating rectangular treemaps. *Transactions on Visualization and Computer Graphics*, 2010.
[19] Robert Kosara. Eagereyes.
[20] Robert Kosara, Fabian Bendix, and Helwig Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):558–568, July 2006.
[21] Erich Neuwirth. *RColorBrewer: ColorBrewer palettes*, 2011. R package version 1.0-5.
[22] William Playfair. *Commercial and Political Atlas*. London, 1786.
[23] William Playfair, Howard Wainer, and Ian Spence. *Playfair's Commercial and Political Atlas and Statistical Breviary*. Cambridge University Press, 2005.
[24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.
[25] Naomi Robbins. *Creating More Effective Graphs*. Wiley, 2005.
[26] Matthias Schonlau. Visualizing categorical data arising in the health sciences using hammock plots. In *Proceedings of the Section on Statistical Graphics*. RAND Corporation, American Statistical Association, 2003.
[27] M Theus, H Hofmann, B Siegl, and A Unwin. *New Techniques and Technologies for Statistics II*, chapter MANET: Extensions to interactive statistical graphics for missing values. IOS Press Amsterdam, 1997.
[28] Edward Tufte. *The Visual Display of Quantitative Information*. Graphics Press, USA, 2 edition, 1991.
[29] Howard Wainer. *Visual Revelations*. Psychology Press, 2000.
[30] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.