Common Angle Plots as perception-true visualizations of categorical associations

Heike Hofmann, *Member, IEEE,* and Marie Vendettuoli

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 transends barriers of language to communicate complex concepts from author to audience. It becomes a problem if readers are unable to easily extract the main message, especially when distortion is encoded. The source of a distortion may be due to intrinsic deformities in the graph or simply the perceptual limitations of the audience. Examples include Tufte's *Lie-Factor* [1, p. 57–69] in which the proportion of the physical space occupied by the graphic is inconsistent with underlying data; calculated ratio (of proportions) less than one indicate underrepresentation. Another example is the Müller-Lyer family of illusions such as the sine wave, where viewers perceive extents at the curves to be of different height than in the straight regions even though all regions were of the same height [2].

Regardless of the cause of distortion, the graph author has a duty to create visualizations that allows readers to extract an accurate interpretation of the underlying data. The *Lie-Factor* provides a quantitative method to evaluate distortion due to graph deformities. In order to ascertain the impact of distortion due to perceptual limits, usability studies provide empirical evidence supporting underlying metaphorical models, both known and unpredicted.

Parallel sets (parsets) [3], a graphical method for visualizing multivariate categorical data, presents a case of unspecified distortion due to perceptual limits . Since initial publication, parsets have spread to mass media outlets [4], [5], [6], multiple technical implementations [4], [7], [8] and is a reputable resource for further academic work ([3] 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, parsets introduce the frequency scale that is a well-known feature of other categorical displays such as barcharts or mosaic plots [9], [10], [11], [12].

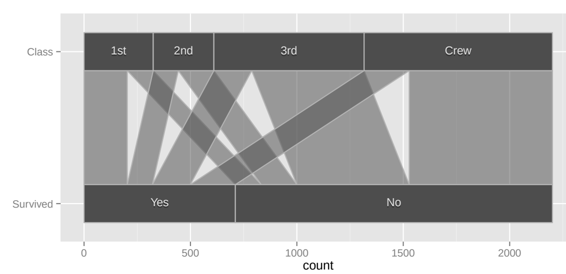

Fig. 1. Parallel sets plot showing the relationship between survival of the sinking of the HMS Titanic and class membership.

Figure 1 gives an example for a two dimensional parallel sets plot investigating the relationship between class status and survival on board the HMS Titanic [13]. Class status is recorded as either crew member or passengers in first, second, or third class. The top bar in figure 1 shows the variable Class. The bottom bar shows survival as yes and no. Between the bars lines are drawn to visualize the relationship between class membership and survival. Based on the number of survivor and non survivors these bands are drawn from each class, and their (horizontal) width is proportional to the number of people they represent. A reasonable task would be to order levels of the variable 'class' by number of survivors. However, when study participants

- *H. Hofmann is faculty with the Department of Statistics, Iowa State University, Ames, IA, 50014. E-mail: hofmann@iastate.edu*
- *M. Vendettuoli is graduate student HCI and BCB programs, Iowa State University and member of the Statistics Section, USDA APHIS CVB Ames, IA 50014*

	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.

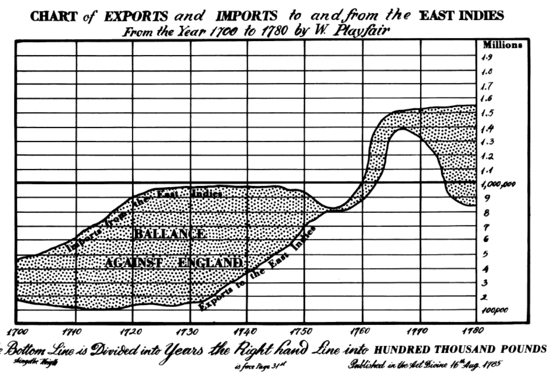

Fig. 2. Playfair's chart from the Commercial and Political Atlas (1786) showing the balance of trade between England and the East Indies. In which years was the difference between imports and exports the highest?

were asked to perform this task, only 12.5% respondants selected the correct order **??**.

We believe that readers view parsets 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 2. 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 [14], [15]. 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 [16], they very often miss the steep rise in the difference between the lines in the years between about 1755 and 1765. Figure 3 shows the actual difference between imports and exports.

This phenomenon is known and widely discussed in statistical graphics literature [16], [1], [17], [18]. It is due to our tendency to assess distance between curves as the minimal (orthogonal) distance rather than the vertical distance – see sketch 4 for a visual representation of 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

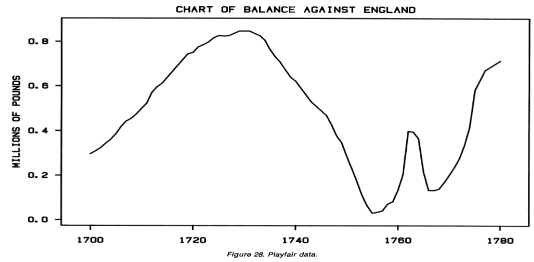

Fig. 3. Difference between exports and imports from England to and from the East Indies in the 18th century – the steep rise in the difference around 1760 comes as a surprise to many viewers of the raw data in figure 2.

Müller-Lyer illusions [2]. 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 2 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 4) needed for a correct evaluation of parsets-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.

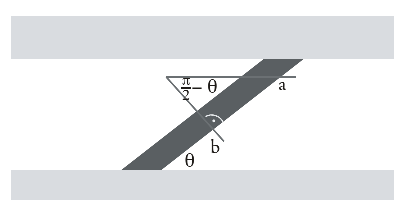

Fig. 4. Sketch of line width assessments: (a) is showing horizontal width, (b) shows width orthogonal to the slope. Survey results in section **??** indicate that observers associate line width more with orthogonal width (b) than horizontal width (a).

The perceived slope of a line very much depends on the aspect ratio of the corresponding plot – changing

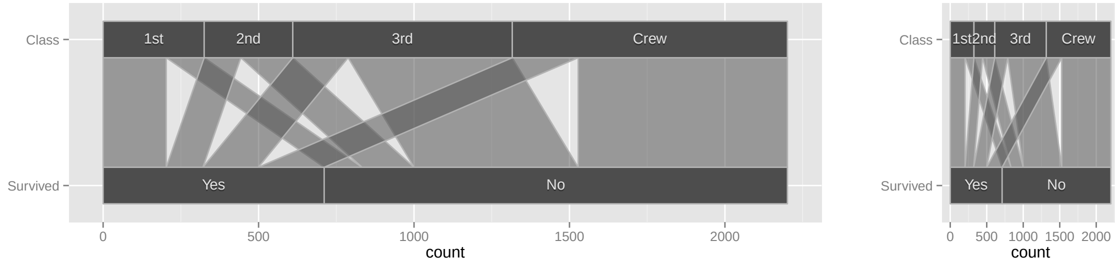

Fig. 5. Parallel sets plots of survival on the Titanic by class. Different aspect ratios seemingly change the thickness of line segments, compare e.g. number of survivors in 3rd class and in the crew.

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 [16]. This finding is not new, but its strength on our perception is surprising, as can be seen in the example of figure 5. 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 parsets-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 [19] and (2) when comparing rectangles rotated relative to each other [20]. 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 [21], provide an alternative to parallel sets that is adjusted for the line width illusion. This 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 6 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.

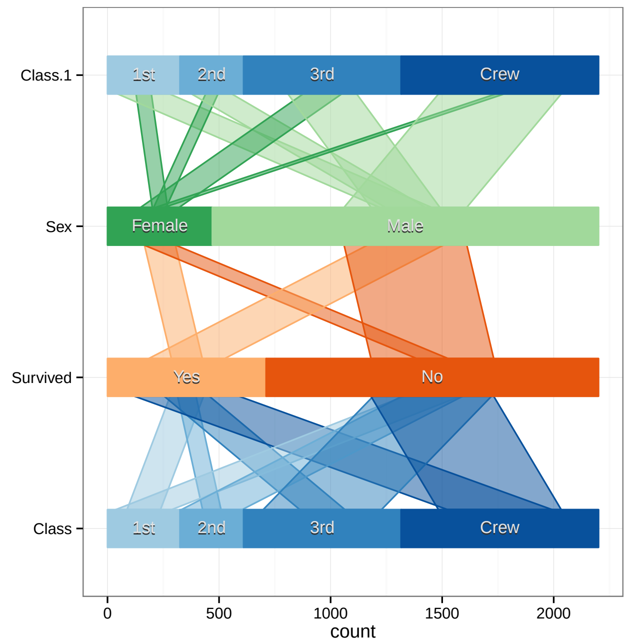

Fig. 6. Hammock plot of the relationship between Class and Survival on the Titanic.

The graph shows that barely any women were in the crew, while male crew members make up the second largest contingent overall. 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 about fifty-fifty. 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 figure 6 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 alone is not proportional to underlying data.

[[If only h-width is considered, the graph has a lie factor; can we quantify this amount in terms of theta?]]

4 Common Angles

4.1 Construction

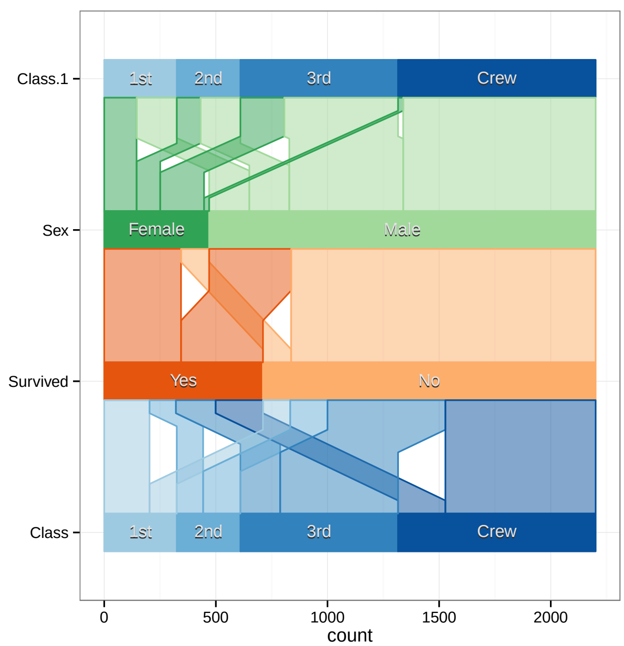

Fig. 7. Common angle plot of the Titanic data.

Figure 7 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.

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 7, we modify the connecting bands between categories from a straight band to 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.

5 Usability Testing

5.1 Test

5.2 Results

6 Conclusion

The conclusion goes here.

Acknowledgment

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

Appendix one text goes here.

References

[1] E. Tufte, *The Visual Display of Quantitative Information*, 2nd ed. USA: Graphics Press, 1991.
[2] R. H. Day and E. J. Stecher, "Sine of an illusion," *Perception*, vol. 20, pp. 49–55, 1991.
[3] R. Kosara, F. Bendix, and H. Hauser, "Parallel sets: Interactive exploration and visual analysis of categorical data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 4, pp. 558–568, Jul. 2006. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2006.76
[4] R. Kosara. Eagereyes. [Online]. Available: http://eagereyes.org/parallel-sets
[5] S. Carter and M. Bostock. (Oct 15 2012) Over the decades, how states have shifted. [Online]. Available: http://www.nytimes.com/interactive/2012/10/15/us/politics/swing-history.html
[6] M. Blastland, "Go figure: How to understand risk in 13 clicks," March 11 2009. [Online]. Available: http://news.bbc.co.uk/2/hi/uk{_}news/magazine/7937382.stm
[7] M. Bostock, V. Ogievetsky, and J. Heer, "D3: Data-driven documents," *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011. [Online]. Available: http://vis.stanford.edu/papers/d3
[8] J. Davies. Parallel sets. [Online]. Available: http://www.jasondavies.com/parallel-sets/
[9] J. Hartigan and B. Kleiner, "Mosaics for contingency tables," in *Proceedings Symposium on the Interface*, 1982.
[10] M. Friendly, "Visualizing categorical data: Data, stories and pictures," in *SAS User Group Conference*, 1992.
[11] H. Hofmann, "Exploring categorical data: Interactive mosaic plots," *Metrika*, 2000.
[12] M. Theus, H. Hofmann, B. Siegl, and A. Unwin, *New Techniques and Technologies for Statistics II*. IOS Press Amsterdam, 1997, ch. MANET: Extensions to interactive statistical graphics for missing values.
[13] R. J. Dawson, "The 'unusual episode' data revisited." *Journal of Statistics Education*, vol. 3, 1995. [Online]. Available: http://www.amstat.org/publications/jse/v3n3/datasets.dawson.html
[14] W. Playfair, *Commercial and Political Atlas*, London, 1786.
[15] W. Playfair, H. Wainer, and I. Spence, *Playfair's Commercial and Political Atlas and Statistical Breviary*. Cambridge University Press, 2005.
[16] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. pp. 531–554, 1984. [Online]. Available: http://www.jstor.org/stable/2288400
[17] H. Wainer, *Visual Revelations*. Psychology Press, 2000.

[18] N. Robbins, *Creating More Effective Graphs*. Wiley, 2005.
[19] J. Heer and M. Bostock, “Crowdsourcing graphical perception: Using mechanical turk to assess visualization design,” in *CHI 2010: Visualization*, 2010.
[20] N. Kong, J. Heer, and M. Agrawala, “Perceptual guidelines for creating rectangular treemaps,” *Transactions on Visualization and Computer Graphics*, 2010.
[21] M. 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.
[22] H. R. Sankey, “Introductory note on the thermal efficiency of steam-engines,” Report of the Committee appointed on the 31st March, 1896, to Consider and Report to the Council upon the Subject of the Definition of a Standard or Standards of Thermal Efficiency for Steam-Engines: with an Introductory Note. Minutes of Proceedings of The Institution of Civil Engineers, pp. 278–283, 1898.
[23] C. Minard, “Carte figurative des pertes successives en hommes de l’Armée Française dans la campagne de Russie 1812-1813,” 1869.
[24] M. Schmidt, “The Sankey diagram in energy and material flow management,” *Journal of Industrial Ecology*, no. 12, pp. 173–185, 2008.
[25] L. L. National Laboratory and the U.S. Department of Energy, “The efficiency solution,” *Scientific American*, June 2009.
[26] M. Fischetti, “Water in, water out,” *Scientific American*, June 2012.

Heike Hofmann is an associate professor of Statistics at Iowa State University. She is a member of the inter-disciplinary Human Computer Interaction program at Iowa State and serves as faculty member for the Bioinformatics and Computational Biology program. Her research interests are the visual exploration of high-dimensional and large data. Her research group has been the recipient of several awards for their data detective and visualization skills.

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. Her research interests involve the use of interactive visualizations in exploratory data analysis.