Eurographics Conference on Visualization (EuroVis) (2014)
N. Elmqvist, M. Hlawitschka, and J. Kennedy (Editors) *Short Papers*

Interactive Visualisation of Automotive Warranty Data Using Novel Extensions of Chord Diagrams

Michael Nicholas, Daniel Archambault and Robert S. Laramee

Department of Computer Science, Swansea University, UK

Abstract

Descriptive and predictive analytics enable the automotive industry to be pro-active in its management of warranty repairs and the substantial unknown costs associated with future claims. Understanding what makes their customers most dissatisfied can assist manufacturers to take pro-active steps towards restoring satisfaction and increasing the likelihood of customers making future purchases. Presented in this short work-in-progress paper are two techniques which enable the interactive visualisation of high-dimensional relationships within categorical data. Our Multi-Chord Diagram and Multi-Chord Glyph Diagram extend traditional Chord Diagrams, overcoming the limitation of only depicting relationships between category pairs. We present these techniques in an application, addressing a real-world problem of visualising data relating to customer satisfaction following various combinations of automotive warranty repair. Although presented as a novel application for visualising customer satisfaction, we feel that the techniques described could also be applied to other scenarios involving the visualisation of variable high-dimensional relationships within categorical data.

1. Introduction

This research arose from a real-world problem faced by our industry partner, a company which specialises in analysis and predictive analytics relating to automotive warranty data. A large automotive manufacturer has supplied data collected from customer surveys in which customers indicate whether they are satisfied with their vehicle purchase. Surveys are conducted at various points throughout the warranty period, typically at routine service intervals. Research by Experian indicates that brand loyalty for some manufacturers can exceed 45% [Exp12]. Customer satisfaction is an obvious contributing factor towards this. With 84 million new vehicles sold in 2012 [Int13] there is substantial financial motivation towards being able to understand and improve customer satisfaction levels.

The survey data can be linked by a unique vehicle identification number to the automotive warranty claims made by those customers. Through statistical analysis of this linked data, it is possible to estimate the impact of different problems (resulting in warranty claims) on customer satisfaction and to compare this against a base satisfaction level for customers who experienced no problems. It is also possible to examine how satisfaction levels are affected if customers experience two or more different types of problem.

Warranty claims are divided into 11 problem categories such as Engine, Transmission, etc. In this document we refer to the categories using letters A to K. There are 35,208 survey responses where the customer is flagged as having suffered a problem in one or more categories. Over 5,000 of these experienced problems in three or more categories. We have, for any subset of the problem categories, values indicating the number of customers who experienced that combination of problems and how likely it is that a customer with that combination of problems will report being satisfied with their vehicle purchase. There can be up to 2^n subsets of n categories, so exponential growth could pose problems for larger numbers of categories. In the data there are 829 subsets with cardinalities ≥ 2. Our challenge is to find a method for visualising the results of this analysis, which can depict high-dimensional relationships of varying cardinalities.

In this paper we present a novel work-in-progress approach to visualising automotive warranty customer satisfaction data. In Section 3 we describe a Multi-Chord Diagram which extends Chord Diagrams to include chords that

DOI: 10.2312/eurovisshort.20141167

describe relationships between three or more category segments. In Section 4 we describe a Multi-Chord Glyph Diagram which swaps chords for glyph representations of these high-dimensional relationships. A video accompanies this paper to better demonstrate the visualisations.

2. Related Work

Chord Diagrams is a radial technique for visually exploring pairwise relations in categorical data. Relationships between pairs of category segments are presented using chords, a 2D surface which arcs across a circle, with the relative size of segments and chord thickness indicating the frequency of a relationship. Chord Diagrams feature in 'D3: Data-Driven Documents' [BOH11], a visualization framework using common web technologies. Chords are also used within several other radial visualisation techniques, including 'Circos: An information aesthetic for comparative genomics' [KSB*09] which features radial diagrams that use chords to show the relationship between different sections of a genome. A fundamental difference between previous work and the work presented here is that our visualisations handle variable high-dimensional data.

Related to our work are Contingency Wheel [AGMS11] and subsequently Contingency Wheel++ [AAMG12] for exploring data in contingency tables. Alsallakh et al. also developed the novel and closely related Radial Sets technique for the visual analysis of large overlapping sets [AAMH13], using arcs/chords and hyper-edges to visualise overlaps with a degree ≥ 3. Our Multi-Chord and Multi-Chord Glyph diagrams were developed in parallel to Radial Sets. Despite addressing similar underlying challenges, Radial Sets do not feature the glyph-based visualisations that this work does.

Other related techniques include a radial technique for visualising undirected hypergraphs [KJ13], which draws hyperedges as arcs around the outside of a node circle and features no occlusion as the external arcs never overlap. Hierarchical Edge Bundling by Holten [Hol06] is a commonly used technique for clutter reduction. Interchange circos diagrams [ZFAQ13] use a flyover ring and chord bundling to address occlusion. Baur and Brandos discuss crossing reduction in circular layouts [BB04], which could be used to find an improved ordering of the categories in a radial layout. KronoMiner [ZCB11], ChromoWheel [ES04] and MizBee [MMP09] are additional examples of related techniques which utilise radial layouts. A technique similar to edge stubs in Parallel Tag Clouds [CVW09] could also be adapted for our Multi-Chord Glyph diagram.

3. Multi-Chord Diagram

The Multi-Chord Diagram overcomes a limitation of Chord Diagrams by visualising relationships between multiple ($\geq$ 3) categories. From Chord Diagrams we retain the 'chord'

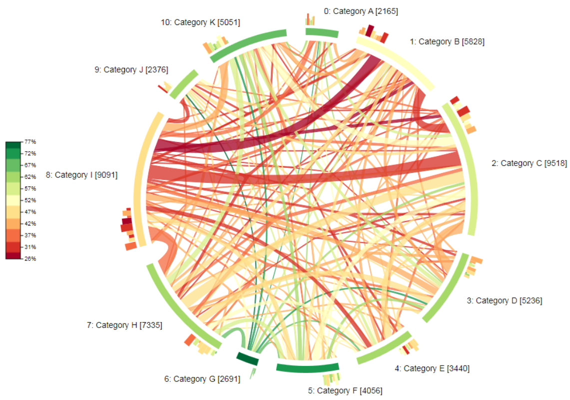

Figure 1: *Multi-Chord Diagram describing the frequency of automotive warranty repairs in various problem categories and the resulting impact upon customer satisfaction. Frequency is represented by the size of outer segments and thickness of chords, while satisfaction impact is represented using colour. The visualisation shows that problem categories B, C and I have the most negative impact upon satisfaction level, and frequently occur together. Categories F and G are both less frequent and have less of a negative impact upon satisfaction.*

for depicting the relationship between two categories. We introduce a new 'multi-chord' for visualising the relationships between three or more categories. The addition of multi-chords increases the complexity of the visualisation, as we are visualising sets of problem categories that vary in cardinality, each with an associated frequency and satisfaction level.

3.1. Arrangement, Size & Ordering

In Chord Diagrams the chords maintain a 1:1 ratio with segment size. A chord with frequency 10 will join to 25% of the segment's inner edge if the segment represents a problem category with frequency 40. The frequency of a problem category is the sum of the frequency of all sets which contain that category. A design decision must be made whether to include sets of cardinality 1 in the summation. If included then the visualisation will accurately reflect problem frequency and it will be the case that only part of each segment's inner edge will join to chords and multi-chords. Chords can be placed adjacently at one end of, or be distributed evenly along, a segment's inner edge. Alternatively if the user were primarily interested in inspecting the relationships between two or more categories then we could choose to exclude cardinality 1 sets from the summation, which will result in full use of each segment's inner edge. In Figure 1 we have chosen to exclude sets of cardinality 1. The true proportion between single and multiple problem categories is still indicated by the circumference utilised by the external histogram, which also gives an indication of the frequency of relationships with each other category.

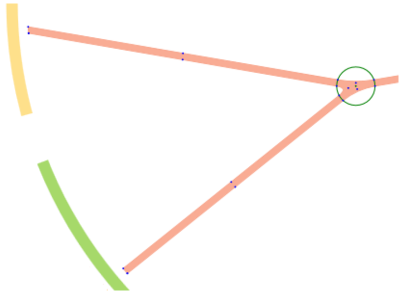

Figure 2: *Multi-Chord construction showing circle overlay and control points in blue.*

The ordering along a segment's inner edge can heavily impact upon the resulting visualisation. We use an algorithm which orders chords to fan out from the inner edge, such that where possible chords and multi-chords which connect to an anti-clockwise adjacent segment are placed towards the anti-clockwise end of a segment's inner edge, and vice-versa. While overlapping chords are inevitable this fanned approach reduces the number of overlaps within the visualisation and produces a better result in comparison to ordering by frequency or satisfaction level.

3.2. Multi-Chord Construction

Multi-Chords are each constructed around a centroid within the inner circle. The centroid is computed by averaging the x, y position of both the start and end points where that multi-chord joins each segment. To prevent multi-chords between adjacent segments from being too close to the boundary of the inner circle we include the centre of the visualisation as a weighted component in the calculation. The coefficient can be customised in an interactive user option, allowing the user to pick the best value based upon the visualisation output.

Around each multi-chord centroid we define a circle (Figure 2) which we use in the construction algorithm. We set the circle radius to be $(2 \times \pi \times VisRadius \times (ChordSize/360) \times Cardinality \times Weighting)$. $ChordSize$ is the thickness of the join to a segment (in degrees), $VisRadius$ is the overall visualisation radius and $Weighting$ is an optional user configurable coefficient parameter.

Each multi-chord arm is constructed using two parallel lines from the segment towards the centroid. Quadratic bezier curves smoothly join adjacent arms together, using the intersection point of the edges as a control point. When the angle between two arms is small the intersection can occur outside of this circle. Lines being joined can be shortened so that they terminate prior to the intersection point and a small bezier curve used to soften the acute intersection angle.

3.3. Interactivity & User Options

As described in the introduction, the number of multi-chord permutations increases as $O(2^n)$. The 829 sets in the customer satisfaction dataset are enough for clutter and occlusion to become an issue.

Opacity can be controlled by the user via a slider, allowing them to select the most appropriate level. Opacity is also used during interactive filtering. Hovering over a chord highlights it against other chords. Users can also hover over a segment to highlight all chords which relate to that category. Clicking a segment activates the segment filter, so the user can hover over and select other segments to drill-down further. Value-based filters are implemented using dual-value sliding controls for filtering chords based upon the Cardinality, Frequency or Satisfaction values. Sliders are also used for setting the coefficient values to manipulate the chord construction algorithm. Increasing the centre radius coefficient can be useful for exaggerating the centre of high-cardinality multi-chords.

For representing the customer satisfaction level we use a green to red diverging colour scheme from Colorbrewer [Bre14] to represent high to low satisfaction levels.

3.4. Chord Grouping Threshold

Applying a filter to frequency can address occlusion by reducing the number of multi-chords and only showing the most significant problems. However, hiding lots of low-frequency sets can mask a large cumulative problem. We address this with a chord grouping threshold user option, which redistributes the value of low-frequency sets to their subsets. This method ensures that the overall number of customers reflected by the visualisation does not decrease when low-frequency chords are hidden.

A customer reporting a problem in categories $S = \{A,B,C\}$ is also a customer who experienced problems $S_1 = \{A,B\}$, $S_2 = \{A,C\}$ and $S_3 = \{B,C\}$. If S occurs f times and $f < t$ (threshold), then we divide and distribute f equally amongst each of its immediate subsets. For example, if S_1, S_2, S_3 each have $f = 5$ and S has $f = 3$. A threshold $t = 4$ would result in S being distributed across S_1, S_2, S_3 giving them $f = 6$ and S having $f = 0$ (not shown). This effect is compounding starting from the highest cardinality, so that if a set which has received some distributed frequency is still below the threshold, its new value will again be distributed to the next level of subsets. Each chord presents both its original value and grouped value so that user can inspect how much of the frequency has been grouped from low frequency supersets.

This has a significant impact with our Customer Satisfaction data, where a threshold of $t = 10$ reduces from 829 to 240 chords. All chords reflect 13,415 customers. The 589 redistributed chords reflected 1,235 customers ($< 10\%$). Each of the removed chords is relatively insignificant, but cumulatively they represent a large enough percentage of the customers that they should not be discarded. Instead their frequency is grouped into the subsets so that users can under-

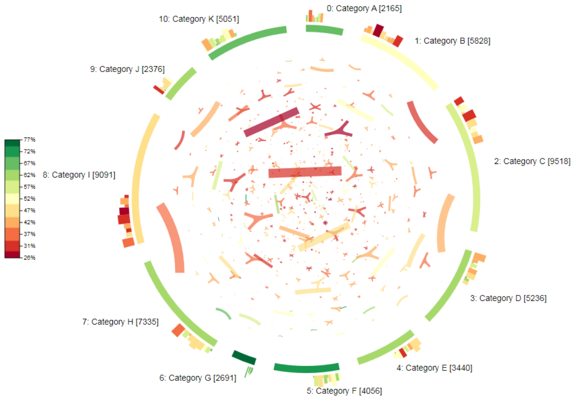

Figure 3: *Multi-Chord Glyph Diagram.*

stand where those customers would fall. This low threshold of $t = 10$ results in 71% fewer chords rendered within the visualisation space. With $t = 30$ only 117 chords are rendered.

4. Multi-Chord Glyph Diagram

Our second extension, the Multi-Chord Glyph Diagram, replaces each chord and multi-chord with a glyph representation (Figure 3). The motive for this alternative approach is to reduce the amount of surface area that each chord requires within the visualisation space, and therefore reduce the amount of clutter and occlusion.

A multi-chord glyph is effectively the central part of a multi-chord, with shortened arms that extrude in various directions to indicate the segments in the relationship. Hovering over a glyph transforms it into the full multi-chord representation (Figure 4). Arm length is 4 times the glyph thickness but this could be made into a user option. Glyphs are aligned towards the centre of each segment. A design decision is still required to decide whether to include or exclude cardinality 1 sets in the segment size calculation. We prefer to exclude them, which results in larger glyphs and emphasises higher frequency chords.

Glyphs generate much less surface area within the visualisation, leaving unused space around the outside of the visualisation. This introduces a problem which isn't observed in the Multi-Chord Diagram. The averaging process to determine the centroid places glyphs towards the centre of the visualisation, leaving underutilised space around the edge of the visualisation. To distribute the glyphs more evenly, we apply a logarithmic interpolation function to spread the glyphs radially. The result is a more even distribution, which places glyphs further from the centre of the visualisation but not too close to the edge.

Glyphs connecting just two segments produce a straight or slightly-curved segment, which is easy to distinguish from a higher-cardinality glyph without the need for a different implementation. Hovering and clicking on segments filters the

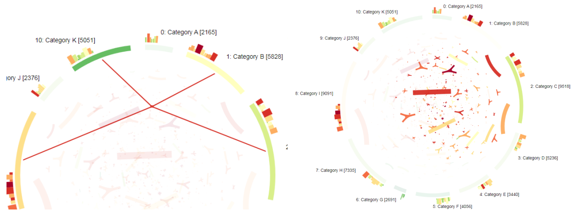

Figure 4: *Left: Hovering over a Multi-Chord Glyph transforms it into a full Multi-Chord representation. Right: Hovering over a segment highlights only the glyphs which relate to that category.*

glyphs using opacity (Figure 4). While low frequency Multi-Chords may appear as a 1-pixel wide line, low frequency Glyphs may generate just a single pixel. This is a negative aspect when compared to the Multi-Chord Diagram, but to artificially increase their size would break the frequency scale. The grouping threshold can be used to redistribute these low frequency relations.

Glyphs require the user to mentally map them against their respective category segments. Feedback from users indicates that it is more difficult to follow the relationships, but that the visualisation suffers less from occlusion and gives a better overview to see where more significant relationships exist.

5. Conclusion and Future Work

Presented as a work-in-progress, the Multi-Chord and Multi-Chord Glyph diagrams form a novel application for visualising high-dimensional data relating to customer satisfaction following automotive warranty repairs, overcoming a limitation of Chord Diagrams by enabling visualisation of relationships between three and more categories. Positive feedback has been received from both our industry partner and representatives of the manufacturer who supplied the data. Most users express a preference for one extension over the other, however the choice of preference was quite balanced.

In future work we would like to more formally compare the two extensions to assess whether one is measurably better at presenting information. It will also be interesting to explore how the techniques we have developed for this application can be applied to other real-world data sets, which will include analysing the scalability of the methods.

6. Acknowledgements

The authors would like to thank David Chung and also the KESS programme. This work is part-funded by the European Social Fund (ESF) through the European Union's Convergence programme administered by the Welsh Government.

References

[AAMG12] 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 (12/2012 2012), 2849–2858. 2

[AAMH13] Alsallakh B., Aigner W., Miksch S., Hauser H.: Radial Sets: Interactive Visual Analysis of Large Overlapping Sets. *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2496–2505. 2

[AGMS11] Alsallakh B., Gröller E., Miksch S., Suntinger M.: Contingency wheel: Visual analysis of large contingency tables. 53–56. 2

[BB04] Baur M., Brandes U.: Crossing reduction in circular layouts. In *WG* (2004), Hromkovic J., Nagl M., Westfechtel B., (Eds.), vol. 3353 of *Lecture Notes in Computer Science*, Springer, pp. 332–343. 2

[BOH11] Bostock M., Ogievetsky V., Heer J.: D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2301–2309. 2

[Bre14] Brewer C. A.: Colorbrewer 2.0. http://colorbrewer2.org, 2014. Last Accessed: Mar 2014. 3

[CVW09] Collins C., ViÃl'gas F. B., Wattenberg M.: Parallel tag clouds to explore and analyze faceted text corpora. In *IEEE VAST* (2009), IEEE, pp. 91–98. 2

[ES04] Ekdahl S., Sonnhammer E. L.: ChromoWheel: a new spin on eukaryotic chromosome visualization. *Bioinformatics 20*, 4 (Mar. 2004), 576–577. 2

[Exp12] Experian Automotive: Automotive Industry Insights Summary: Q4 2012. http://www.experian.com/assets/automotive/brochures/experian-auto-q4-2012-market-trends.pdf, 2012. Last Accessed: Mar 2014. 1

[Hol06] Holten D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 741–748. 2

[Int13] International Organization of Motor Vehicle Manufacturers (OICA): Worldwide motor vehicle statistics. http://www.oica.net/category/production-statistics/2012-statistics/, 2013. Last Accessed: Mar 2014. 1

[KJ13] Kerren A., Jusufi I.: A Novel Radial Visualization Approach for Undirected Hypergraphs. Hlawitschka M., Weinkauf T., (Eds.), Eurographics Association, pp. 25–29. 2

[KSB*09] Krzywinski M. I., Schein J. E., Birol I., Connors J., Gascoyne R., Horsman D., Jones S. J., Marra M. A.: Circos: An information aesthetic for comparative genomics. *Genome Research*, 9 (June 2009), 1639–1645. 2

[MMP09] Meyer M., Munzner T., Pfister H.: MizBee: A Multiscale Synteny Browser. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 897–904. 2

[ZCB11] Zhao J., Chevalier F., Balakrishnan R.: Kronominer: using multi-foci navigation for the visual exploration of time-series data. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 1737–1746. 2

[ZFAQ13] Zeng W., Fu C.-W., Arisona S. M., Qu H.: Visualizing Interchange Patterns in Massive Movement Data. *Computer Graphics Forum 32*, 3 (2013), 271–280. 2