

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

Submission ID: 129

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 can also be applied to other scenarios involving the visualisation of 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 had indicated 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 can exceed 45% [Exp12] for some manufacturers. 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 impacted if customers experience two or more different types of problem.

Warranty claims have been split 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 5000 of these experienced problems in three or more categories. Following statistical analysis we can have, for any subset of the problem categories, a value indicating the number of customers who experienced that combination of problems and a value estimating 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 a cardinality ≥ 2 . Our challenge is to find a method for visualising the results of this analysis, which depicts high-dimensional relationships with a variety of cardinalities

During our research into existing visualisation techniques we explored Chord Diagrams, which represent categories as segments arranged around the edge of a circle. Relationships between pairs of category segments are presented using chords, a 2D surface which arcs across the inner circle between two segments. The relative size of segments and chords indicates the frequency of a relationship.

We present a novel work-in-progress approach to visualising the high-dimensional relationships within the customer satisfaction data. Firstly we describe a Multi-Chord Diagram, which extends Chord Diagrams to include chords that describe relationships between three or more category segments. We also describe a Multi-Chord Glyph Diagram, which swaps chords for glyph representations.

2. Related Work

"D3: Data-Driven Documents" presented by Bostock, Ogievetsky and Heer [BOH11] is a visualization framework using common web technologies. We cite D3 as being the inspiration extending Chord Diagrams with high-dimensional data.

D3 cites "Circos: An information aesthetic for comparative genomics" [KSB*09] as their inspiration for Chord Diagrams. Chords are used in Circos diagrams to show the relationship between different sections of a genome. The visualisations display data in a variety of ways, including "scatter, line and histogram plots, heat maps, tiles, connectors and text" [KSB*09].

Kerren and Jusufi presented a radial technique for visualising undirected hypergraphs [KJ13]. Their hypergraph visualiser plots each node in a circle with equal spacing and hyperedges are drawn as arcs around the outside of the node circle. Their technique scales well and notably it features no occlusion, as the external arcs never overlap.

Contingency Wheel [AGMS11] and subsequently Contingency Wheel++ [AAMG12] by Alsallakh et al are designed for the visual exploration of categorical variables in contingency tables. Contingency table columns are mapped to equal-sized sectors of the radial visualisation. The radial ring is used to visualise row-column associations through the use of histograms. More recently, Alsallakh et al. presented Radial Sets, a novel technique for the visual analysis of large overlapping sets [AAMH13]. Arcs/Chords are used to indicate relationships between pairs of sets and hyper-edges visualise overlaps with a degree ≥ 3 . Radial Sets is closely related to our problem of representing high-dimensional relationships between categorical data. The Multi-Chord and Multi-Chord Glyph techniques described in this paper were developed in parallel to Radial Sets, and we feel that our different approach to both technique and application are justification for the novelty of this paper.

We also note the related work of:- Diehl, Beck and Burch compare the use of Radial visualisation layouts against Cartesian layouts. They conclude that a Cartesian layout should be used "unless there are clear reasons to favour a Radial one" [DBB10]. Hierarchical Edge Bundling by Holten [Hol06] is as a potential technique for clutter reduction. Zeng et al. who introduce interchange circos diagrams [ZFAQ13] containing some clever techniques (flyover

ring and chord bundling). Baur and Brandos on crossing reduction in circular layouts [BB04], which could be used to find an improved ordering of the categories in a radial layout. ChromoWheel by Ekdahl and Sonnhammer [ES04] and MizBee by Mayer et al. [MMP09] both demonstrate radial layouts relating to genomic research.

3. Multi-Chord Diagram

The Multi-Chord Diagram overcomes a limitation of Chord Diagrams by visualising relationships between multiple (≥ 3) categories. From Chord Diagrams we retain the 'chord' for depicting the relationship between two categories. For visualising relationships between three or more categories we introduce a new 'multi-chord'. The addition of multi-chords increases the complexity of the visualisation. The N-dimensional relationships require a redesign of the underlying data structure from that of Chord Diagrams. We are visualising a list of sets of problem categories, each with an associated frequency and satisfaction level.

3.1. Arrangement, Size & Ordering

Chords and multi-chords always 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 the include sets of $size = 1$ 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 $size = 1$ 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 $size = 1$. The true proportion is still indicated by the external histogram, which also summarises relationships with other categories.

The ordering along a segment's inner edge impacts upon the visualisation output. We use an algorithm which orders chords joins such that they fan out from the inner edge. Chords and multi-chords which join to an anti-clockwise adjacent segment are placed towards the anti-clockwise end of a segment's inner edge, and vice-versa. This reduces the number of overlaps within the visualisation. We find that this fan approach produces a better result when compared to ordering by frequency or satisfaction level.

3.2. Multi-Chord Construction

Multi-Chords are each constructed around a central point within the inner circle. The central point is computed by

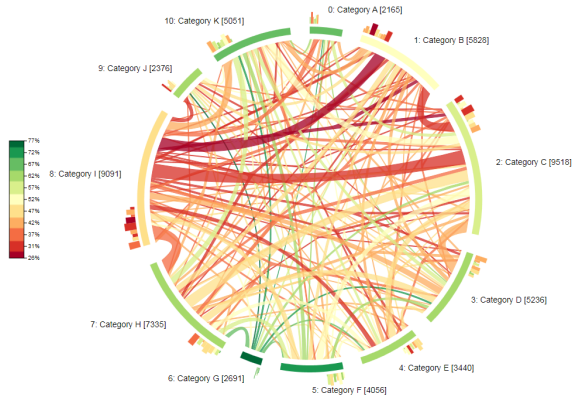


Figure 1: Multi-Chord Diagram describing the frequency of automotive warranty repairs in various problem categories and the resulting impact upon customer satisfaction. Frequency is represented by the size of outer segments and thickness of chords, while satisfaction impact is represented using colour.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

taking an average x, y position of both the start and end points where that multi-chord joins to each segment. To stop multi-chords between adjacent segments being too close to the outer edge of the visualisation we include the centre of the visualisation as a weighted component in the calculation. The co-efficient can be an interactive user option, allowing the user to pick the best value based upon the visualisation output.

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

Each multi-chord arm is constructed with parallel lines from the segment towards the centre point, terminating when they intersect the inner circle. Quadratic bezier curves smoothly join adjacent arms together, using their intersection point as the 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.

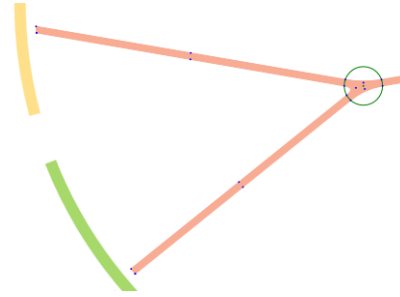


Figure 2: Multi-Chord construction showing circle overlay and control points in blue.

Opacity can be controlled by the user interactively via a sliding control, allowing them to subjectively select the most appropriate level. Opacity is also used for interactive filtering. Hovering over a chord highlights it against other chords. Hovering a segment highlights 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 chord visibility based upon a range of the Cardinality, Frequency or Satisfaction values. Sliding controls 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], providing an intuitive mapping to the notion of good and bad. We also note the Rainbow colour map by Telea [Tel08] as an interesting alternative. Use of a black background also produces an excellent contrast for on-screen viewing.

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 its 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 $S_3 = \{B, C\}$. We propose that if S occurs f times and $f < t$ (threshold), then f can be divided equally amongst each of its immediate subsets. For example, if $S_{1,2,3}$ each have $f = 5$ and S has $f = 3$. A threshold $t = 4$ would result in S being distributed across $S_{1,2,3}$ giving them $f = 6$ and S having $f = 0$ (not shown). This effect

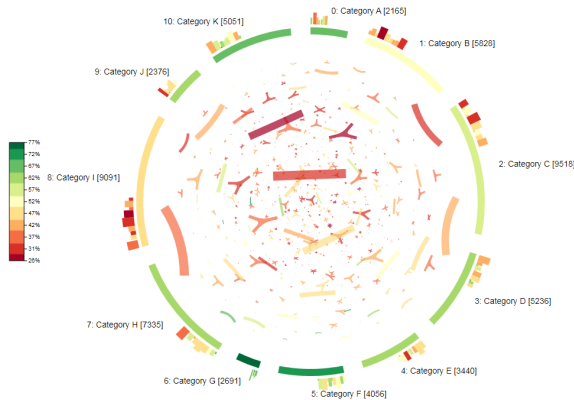


Figure 3: Multi-Chord Glyph Diagram.

is compounding, so 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 set retains the original value so a user can inspect how much of the frequency has been grouped from supersets.

With the Customer Satisfaction data, setting $t = 10$ reduces from 829 to 240 chords. All chords reflect 13415 customers. The 589 redistributed chords reflected 1235 customers $< 10\%$. Each of the discarded chords is relatively insignificant, but cumulatively they represent a large enough percentage of the customers that they should not be discarded. This low threshold 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 4a). 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 $size = 1$ 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 central point places glyphs towards the centre of the

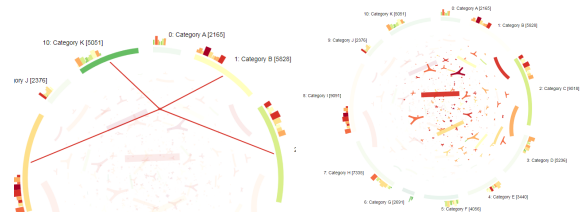


Figure 4: Left: Hovering over a Multi-Chord Glyph transforms it into a full Multi-Chord representation. Right: Hovering over a segment highlights only the glyphs which relate to that category.

visualisation, leaving underutilised space around the edge of the visualisation. To more evenly distribute the central points we use a polylogarithmic function to increase the distance between the visualisation centre and each glyph centre. The result is a more even distribution, while not moving glyphs too close to the edge of the visualisation.

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 glyphs using opacity (Figure 4b). While low frequency Multi-Chords may appear as a visible 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 it is easier to see where more significant relationships exist.

5. Conclusion and Future Work

These two extensions 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 or more categories. Positive feedback has been received from both our industry partner and representatives of the vehicle 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 better. It would also be interesting to explore new applications of these techniques using different data sets.

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
- [DBB10] DIEHL S., BECK F., BURCH M.: Uncovering strengths and weaknesses of radial visualizations—an empirical approach. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 935–942. 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., Weinkauff 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
- [Tel08] TELEA A.: *Data Visualization - Principles and Practice*. A K Peters, 2008. 3
- [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