Multi-Attribute Glyphs on Venn and Euler Diagrams to Represent Data and Aid Visual Decoding

Richard Brath

Oculus Info Inc., Toronto, ON, Canada
richard.brath@oculusinfo.com

Abstract. Representing quantities on Venn and Euler diagrams can be achieved through the use of multi-attribute glyphs. These glyphs can also act as an aid to assist in the visual decoding of the membership of segments within the diagrams and convey other data attributes as well.

1 Overview

Instead of area-proportional Venn and Euler diagrams to indicate quantities, this approach uses separate overlaid glyphs to decouple the representation of data from logical combinations. It also uses glyph attributes to assist in visual decoding of membership of regions. This approach can scale to higher-order logical diagrams and potentially offer more accurate visual estimation than area-proportional techniques.

The depiction of data on set diagrams is useful in various applications (e.g. Boolean queries, genetic informatics). Area-proportional set diagrams have become popular in research (e.g. [1-6]) and software (e.g. see eulerdiagrams.org). However, the area-proportional approach has shortcomings, such as:

1. **Visual comparison of irregular areas is difficult**. Information visualization researchers indicate difficulty with visual comparison of areas and/or a preference for using length instead of area for faster visual comparison (e.g. [7,8,9,15]). In our casual test, only 8% (2 out of 25 people) correctly identified the region of different area on a 2 way Venn as opposed to 80% correctly identifying the circle of different area out of three circles, each test having one item of 20% different area.
2. **Area accuracy vs. aesthetic shapes**. Researchers prefer circles and other aesthetic shapes[6], but the areas (particularly circles) may have a degree of error, typically increasing with higher order sets. e.g.[1]. Wilkinson [6] says “Higher-order Venn diagrams can be drawn on the plane with nonconvex polygons, but they are difficult to compute for more than a few sets and are difficult to decode visually.”
3. **Negative values**: Areas cannot represent negative values unless coupled with another visual attribute, such as hue (e.g. red/green) or shape (e.g. arrows).

Also, discussions with prospective users revealed concern for visual decoding of set membership for a region in complex diagrams, such as higher-order Venn diagrams.

3rd International Workshop on Euler Diagrams, July 2, 2012, Canterbury, UK.
Copyright © 2012 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

Instead, a glyph-based approach is considered. The use of glyphs within set diagrams is not new. Glyphs have been used to represent items in a dataset (first 3 in fig. 1). Spoerri's approach [10] reduces each region of a Venn diagram to a glyph, each glyph indicating the particular Boolean combination by its relative position and shape.

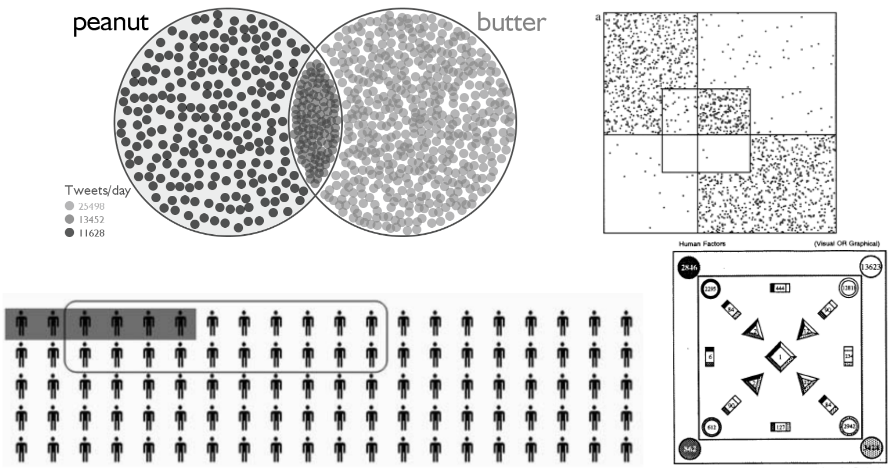

Fig. 1. Top Left: TwitterVenn [11] uses simple glyphs per tweet matching each of three search terms. Top Right: Edwards' Carroll diagram [2] uses both proportional areas (to indicate expectations) with dots (to indicate observations) to draw attention to regions with unexpected results. Bottom Left: [12] depicts a uniform density of glyphs with membership indicated by shading or outlined boxes. Bottom Right: InfoCrystal [10] reduces each region of a Venn diagram to a glyph, where shape and position indicate the Boolean relationship, and the number indicates the count of items within each region.

The contribution of this paper explores, in section 2, the use of glyphs (pictographic and scaled glyphs) to indicate quantities and the use of additional visual attributes to indicate set membership or other data. Results are discussed in section 3.

2 Glyph-based Approach

Our approach is focused on the use of glyphs to decouple the depiction of logical relationships (e.g. Venn and Euler diagrams) separate from the depiction of quantities. By decoupling the quantity from the set diagram, visual attributes more amenable to fast estimation (e.g. size, origentation and color) can be used [7-9]. This approach enables the use simple aesthetically pleasing diagrams of set representations to show the logical relationships between the sets; while using separate glyph(s) within each region to indicate a) quantity of items within a given region, b) indicate set membership to aid visual decoding and c) potential additional attributes.

2.1 Sketches and Real-Data Mockups

To quickly iterate through conceptual ideas, loose sketches were followed by mock-ups using simple sets of real-data. Loose sketches can reveal limitations of promising ideas when implemented with real-data, (e.g. occlusion, imperceptible differences, large dynamic ranges, etc). For rapid mock-ups, we divided the Titanic passenger list into 4 sets for a Venn diagram and 3 sets for an Euler diagram which resulted in useful properties such as empty segments, small segments and large segments.

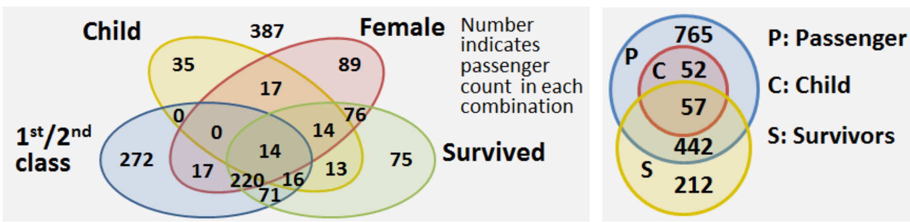

Fig. 2. Left: Venn diagram of Titanic passengers by 4 attributes showing passenger counts. Right: Euler diagram of Titanic passengers and crew.

2.2 Unit Markers and Pictographs

Markers of a fixed size can be repeated to represent quantities ranging from simple dots to pictographs e.g. Isotype [13]. Additional data can be represented on each marker, e.g. using color or sub-shapes. For example, “Social Stratification in the United States” [14] uses pictographic markers with human figures indicating five variables, through 1) background color (occupation), 2) shape (gender), 3) pairing (marital status), 4) extra outline (dependents), and 5) figure color (race). We have also used this approach successfully, e.g. [15].

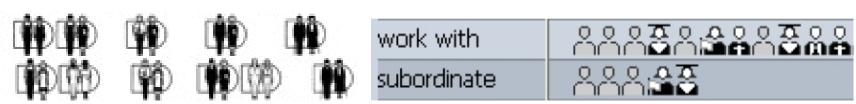

Fig. 3. Pictographic glyphs indicating multiple data variables, from [13,14].

However, a pictograph approach has some challenges:

- Some regions of the set diagram are too small to fit the pictographs. The addition of the leader line could increase the effort to visually decode the relationships.
- The irregular shape of some regions of the set diagram requires an irregular placement. Pictographs organized linearly can be visually estimated by length, which is preferred to visual estimation of area (e.g. [8]).

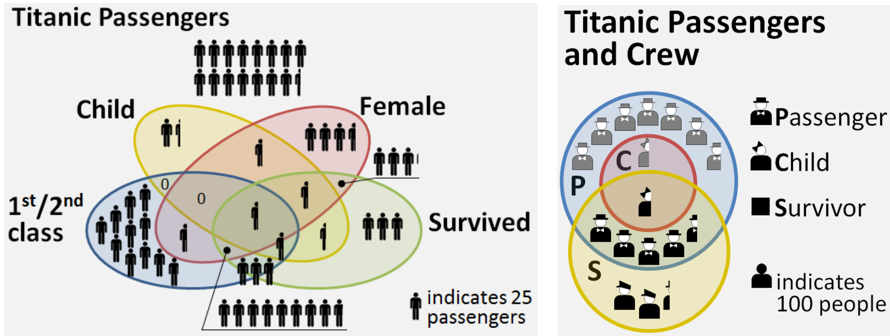

Fig. 4. Pictographs on set diagrams. Leader lines are used when pictographs do not fit regions.

2.3 Scalable Glyphs

Scalable glyphs are a single glyph for each region, sized by the quantity of items associated with that region. Simple glyphs, such as bars varying in length or circles varying in radius, can effectively convey quantities [8,9,16]:

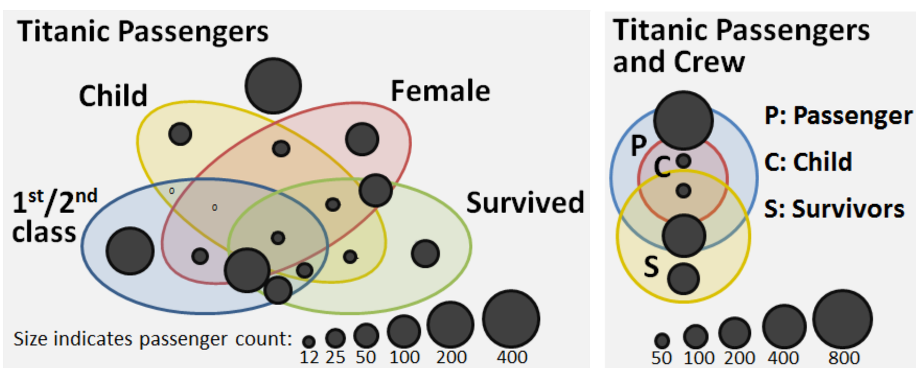

Fig. 5. Scaled glyphs indicating quantities on both Venn and Euler diagrams.

2.4 Indication of Set Membership

With higher order set diagrams, it can be more difficult to perceptually decode the membership for a given component of interest [6]. There are many possible approaches to indicate set membership using either the set diagram or the glyphs.

Background Color: Color can be used, but is problematic. It is challenging to decode the color in intersections as color is not understood as separable [16].

Background Texture: Textures have been used to aid in identification of set membership e.g. [17, 18]. Distinguishing regions by a heterogeneous channel-based approach [19] could be more effective. However, in small regions, textures may not be clearly distinguishable or the glyphs may occlude textures.

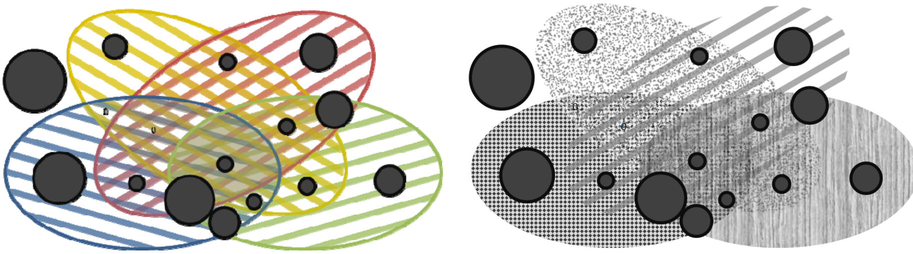

Fig. 6. Background textures can assist in decoding membership.

Glyph with Colors: The same coloring used in the set diagram can be reused in glyphs to indicate membership. Rather than blend colors, however, the colors can be kept separate within the glyph. The layout of the color could be organized as stripes, or radially resembling a bullseye or pie.

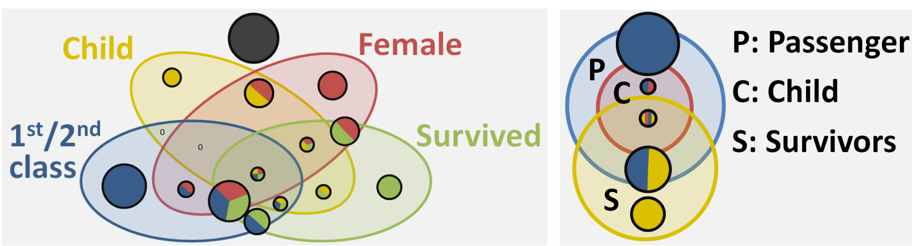

Fig. 7. Colored glyphs use same colors as the sets to aid in identification of set membership.

Glyph with Oriented Whisker: In some set diagrams (e.g. Venn) the placement of the label is typically around the perimeter which can be leveraged by the glyph, specifically by modifying the shape with an added whisker [20] oriented along the same vector from the center of the diagram to the label. To visually decode the membership of any bubble, the viewer can read the orientation of the whisker, similar to decoding the hands of a clock or the spokes of a wind rose.

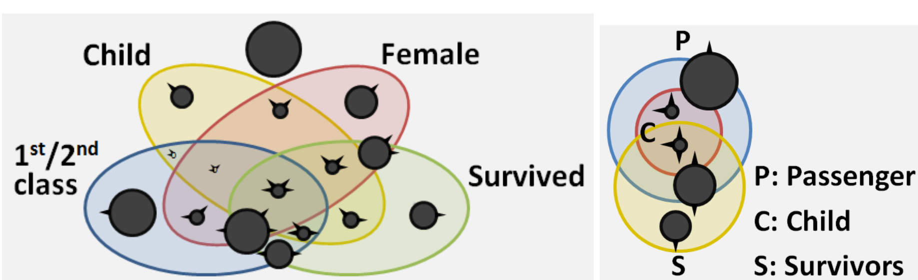

Fig. 8. Glyphs with added whiskers oriented based on set memberships.

The glyph-based approach also allows for additional visual attributes to convey additional data values. For example, the whisker-based glyphs can use:

- Traditional visualization attributes, such as brightness, hue, texture
- Shape-based attributes, such as closure, curvature or edge type
- Each whisker-shape can be independently modified to indicate a data attribute with respect to the set membership, for example whisker length or width
- The internal area of the glyph can be used, in larger glyphs, for example, as a pie chart or with a pictograph.

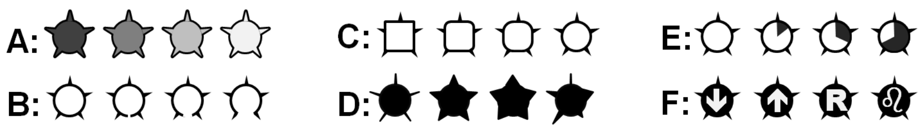

Fig. 9. Example whisker glyphs with additional visual attributes, a) brightness, b) closure, c) curvature, d) whisker width and length, e) internal pie, f) internal pictograph. Negative values could be connotatively conveyed by fill (e.g. red/green hue) or pictograph (up/down arrow).

The whisker-based approach may work well with Venn diagrams, but has may have issues with Euler diagrams and issues where whiskers are potentially occluded. The image below shows a 5-way Venn diagram that has been modified from ellipses to increase the size of the smaller regions to make the technique more workable.

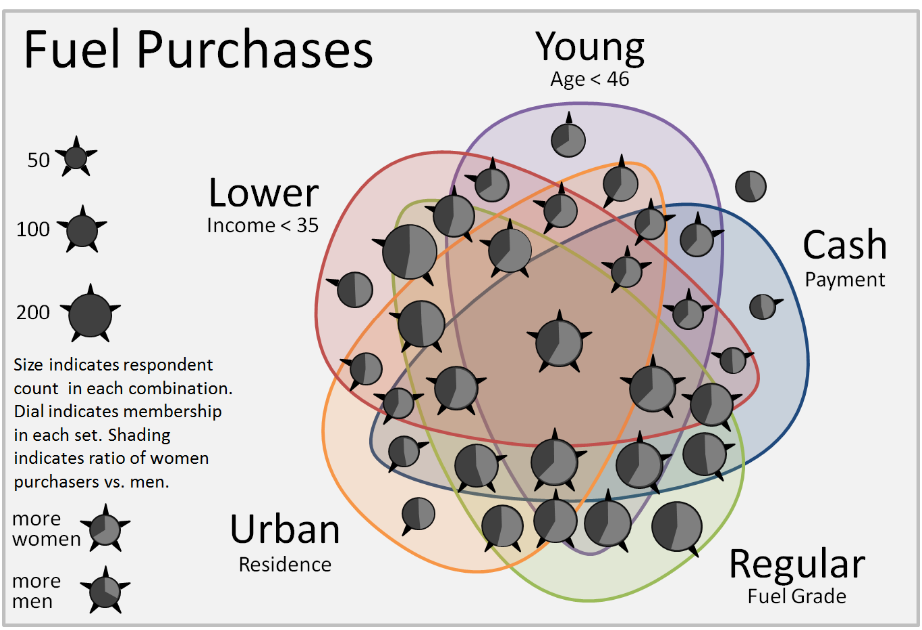

Fig. 10. Whisker glyphs on a 5 way Venn diagram showing data from a survey of 5000 people purchasing fuel, in sets by age, payment type, fuel grade, residence and income. Size of glyph indicates number of respondents in a given component, whiskers indicate set membership, and angular shading indicates the ratio of female to male purchasers of a component.

3 Discussion and Next Steps

Our contribution shows that glyphs can be used to separate the representation of data such as quantities from the representation of sets. Glyphs can:

- Indicate data layered over set diagrams, either as scalable glyphs or as pictographs, and simple size is preferred for fast visual estimation as opposed to irregularly shaped areas [16].
- Also indicate additional data attributes, such as set membership or other data attributes, using visual attributes such as color or orientation of sub-shapes.

While current research in visual comparison indicates this approach may work, evaluation is required to validate. The examples provided indicate various limitations:

- **Glyph size** needs to be carefully managed. Glyphs too big can result in occlusion or require an offset and leader lines. Glyphs too small can be difficult to add additional visual attributes, e.g. for visual decoding.
- **Glyph colors** can effectively represent set membership except color can be difficult to discern when used internally on small glyphs.
- **Glyph whiskers** can effectively represent set membership when set labels are organized around the perimeter, such as in Venn diagrams; but is problematic when sets are distributed throughout the plane.

Implementation on a wide variety of data sets and testing with users requires further effort. Other work could include 3D and interaction techniques, for example, on interaction, extend whiskers to set labels to aid interpretation of whiskers.

References

1. Chow, S., Ruskey, F.: Drawing Area-Proportional Venn and Euler Diagrams. Lecture Notes in Computer Science 2912:466–77 (2004)
2. Edwards, A.W.F., Edwards, J.H.: Metrical Venn diagrams, Annals of Human Genetics 56: 71-75 (1992)
3. Micallef, L., Rodgers, P.: eulerAPE: Drawing Area-Proportional Euler and Venn Diagrams using Ellipses. EMEA Google Scholars Retreat 2011 www.cs.kent.ac.uk/pubs/2011/3119 (2011)
4. Chow, S., Rodgers, P.: Constructing Area-Proportional Venn and Euler Diagrams with Three Circles. Presented at Euler Diagrams Workshop 2005, Paris (2005)
5. Pirooznia, M., Nagarajan, V., Deng, Y.: GeneVenn – a web application for comparing gene lists using Venn diagrams. Bioinformation, 1:420–422 (2007)
6. Wilkinson, L.: Venn and Euler Data Diagrams. Science (2), Citeseer (2010)
7. Healey, C., Booth, K., Enns, J.: High-speed visual estimation using preattentive processing. ACM TOCHI (1996)
8. Jock Mackinlay. Automating the Design of Graphical Presentations, ACM Transactions on Graphics, 5(2) (1986)

9. MacEachren, A.: How Maps Work: Representation, Visualization and Design (2004)
10. Spoerri, A.: InfoCrystal: A Visual Tool For Information Retrieval And Management. In: CIKM '93 Proceedings of the second international conference on Information and knowledge management (1993)
11. Clark, J.: TwitterVenn. www.neoformix.com/Projects/TwitterVenn/view.php
12. Brase, G.L.: Pictorial representations in statistical reasoning. Applied Cognitive Psychology, 23(3), 369-381 (2009)
13. Neurath, O.: International picture language. London: Kegan Paul (1936)
14. Rose, S.: Social Stratification in the United States: The American Profile Poster. The New Press (2007)
15. Jonker, D., Wright, W., Schroh, D., Proulx, P., Cort, B.: Information Triage with TRIST. 2005 Intelligence Analysis Conference (2005)
16. Ware, C.: Visual Thinking for Design. Ch. Structuring Two-Dimensional Space, pp. 43-65, Morgan Kaufmann (2008)
17. Byelas, H., Telea, A.: Texture-based Visualization of Metrics on Software Architectures. Software Visualization (2008)
18. Simonetto, P., Auber, D., Archambault, D.: Fully Automatic Visualisation of Overlapping Sets. Eurographcs/IEEE-VGTC Symposium on Visualization (2009)
19. Ware, C.: Information Visualization: Perception for Design. Ch. Glyphs And Multivariate Discrete Data, pp. 176-184 (2004)
20. Brath, R.: The Many Dimensions of Shape. IV'09 Opening Keynote, www.oculusinfo.com/expertise.html (2009)