Multiple Shape Attributes in Information Visualization: Guidance from Prior Art and Experiments

Richard Brath
Oculus Info Inc.
Toronto, Canada
rbrath@oculusinfo.com

Abstract **- Multiple shape attributes can be used within information visualizations. Prior art from many fields and experiments inform what the attributes of shape are and the potential ways that we may effectively utilize shapes to represent multiple data values within an information visualization.**

Keywords—Shape Attributes, Glyph Design.

I. INTRODUCTION

Data can be mapped to different shape attributes within a visualization. Traditionally, shape has been poorly characterized within the information visualization community, although successive researchers have identified shape attributes that may be effective, based on psychological research [Ber67, Cle85, Hea09, Mac95, Mac06, Maz09, War00, Wol04].

In particular, the use of shape within an information visualization, e.g. within a glyph, to convey more than a single data value, is potentially valuable for increasing information density in visualizations. However, to achieve this, it is necessary to understand what the potential shape attributes are, and how they may be combined together to depict more than one data value.

In the context of this paper, shape attributes refer to independent attributes of shape, such as curvature, terminators, closure, etc; which can be utilized separately or together to convey multiple data attributes within a singular visual marker (e.g. Fig 18); as opposed to other ways of considering shape, such as icons (which are often an abstracted pictographic representation); letters, numbers and common symbols (which may act as a mnemonic label) or compound glyphs (which may utilize only one or two shape attributes in the creation of a shape; and then contain a collection that shape to convey multiple data attributes) as shown in Fig. 1.

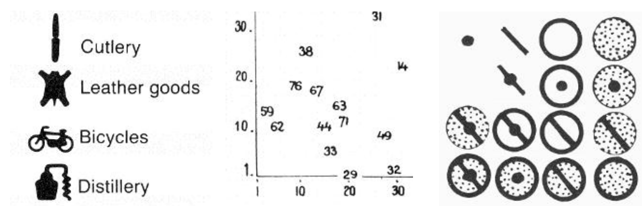


Fig 1. Icons ([Ber67] p. 156), Scatterplot of numbers ([Ber67] p. 249) and Compound Glyphs ([Ber67] p. 338)

Motivation for this exploration was initiated in part by the document "Illuminating the Path, The R&D Agenda for Visual Analytics" [Tho05]. One recommendation is to "Create a science of visual representations". However, in order to utilize shape effectively as a visual representation, we first need to characterize the components of shape that can be utilized within a visual representation.

II. BACKGROUND

There is a wide variety of background information applicable to understanding the attributes of shape and the use of multiple shape attributes together. There are many different potential domains of reference to consider and many more examples that could be considered than covered herein.

A. Information Visualization Research

Bertin [Ber67] originally identified shape as a potentially useful visual variable for representing categorical data, although Bertin did not go into detail regarding additional shape-based attributes. Bertin is somewhat skeptical regarding the use of shape, and comments that it is "tempting to abuse it". Bertin has numerous examples of poor use of shape within a visual display.

Several information visualization researchers have followed on from Bertin's pioneering work, and have also looked to perceptual psychology to create a list of potentially effective visual attributes including several shape-related attributes:

- Termination
- Closure
- Hole
- Curvature
- Added Marks
- Angle
- Intersection

(derived from [Cle85, Hea09, Mac95, Mac06, Maz09, War00, Wol04]). These shape-based attributes could potentially be combined when generating glyphs to represent multiple data variables.

B. Scientific Visualization

The field of scientific visualization has utilized shapes - particularly curvature - for several decades as a means to convey data within a glyph. Most of the techniques have been focused on generating smooth, curved shapes based to represent continuous quantitative data, such as tensor data. For example:

- Superquadrics [Bar81], and similar variants of curvature-based parametric shapes in scientific visualization, have been used in numerous expressive visualizations:

Fig 2. Glyphs using curvature [Bar81]; curvature, hue and thickness [Rop07]; curvature in two dimensions, height & depth [Kin06].

- In Iconic Techniques for Feature Visualization, a variety of compound glyphs are utilized. Some of the glyphs use attributes such as curvature and twist to indicate data attributes.

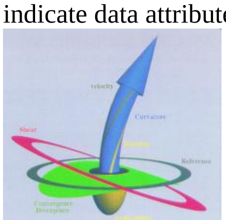


Fig 3. Elements of the glyph use curvature and/or twist [Pos95].

- Blobs, or more specifically, implicit surfaces based on volume rendering of density fields, provides another algorithmic means for generating smooth, closed, curved shapes based on data. Different areas of the surface (different parts of the curvature) correspond to different data attributes.

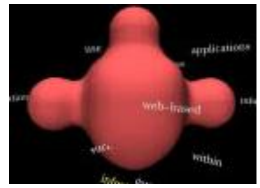


Fig 4. Blobs create amorphous curved surfaces.

C. Information Visualization

Whereas scientific visualization is often based on representing physical real-world phenomena and is therefore constrained to fitting representations within a spatial context, information visualization does not have these constraints. Many information visualizations which utilize shape tend to use shape to represent only a single data attribute, and use simple shapes or icons. Some novel techniques utilizing shape have been explored. For example:

- Chernoff faces, represent multiple features through the use of multi-attribute glyphs that look like familiar objects, i.e. faces. Although the effectiveness and proper use of Chernoff faces is debated, the faces typically utilize shape based attributes to create features such as head eccentricity. Of particular interest, from the point of view of shape attributes, are Chernoff visualizations which utilize multiple shape attributes in a singular visual element, such as the combination of mouth shape and mouth openness, such as this example from Mathematica:

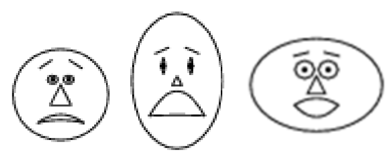


Fig 5. Chernoff faces [Che73]. Note mouth shape and openness in a Chernoff face plot from Mathematica [Wei].

- Use of physical objects as markers on a scatterplot, such as geographic regions or animals, (e.g. [Woo98,Tuf96]), relies on the unique shapes of these objects entities as identifiers. This suggests any of the visual attributes used in the representation of these objects could be used. In the long history of cartography there could be various techniques refined over time that could be leveraged, such as a wide variety of boundaries. These shape features are inherent in the objects, not data-driven attributes, and the effectiveness of objects as glyphs is suspect. (e.g. Wyoming and Colorado are both rectangular, Tufte's mouse and rat are very similar).

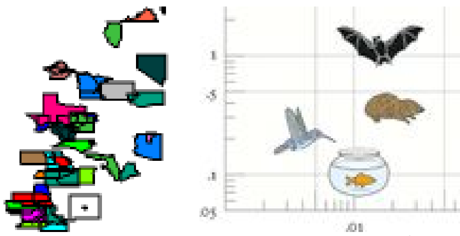


Fig 6. Glyphs as states [Woo09] and animals [Tuf96].

- Star, sticks, and radar plots: Star Coordinates [Kan01] utilizes angle (and length) to convey multiple attributes in a single glyph.

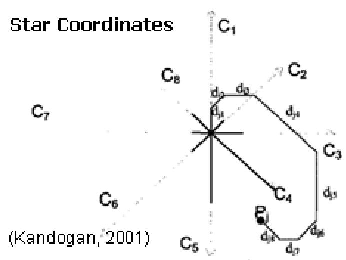


Fig 7. Star Coordinates use of angle and length to create a glyph.

Radar plots offer an interesting opportunity for emergent shapes, although the shapes are based on straight lines and each angle is an artifact of three data values.

Fig 8. Series of radar plots. Shapes emerge.

- Organically inspired "growth" visualizations, such as Fry's Anemone [Fry97] and Dragulescu's Malwarez [Dra08]; utilize generative algorithms and shape-based attributes such as angle, curvature and terminators to generate incredible variety of visual representations.

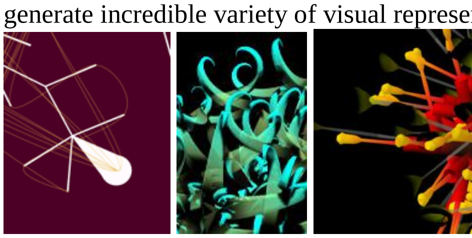


Fig 9. Bulbous terminator on branching curved lines [Fry97] and curls and bulbous terminators [Dra08]

- Other novel shape-based techniques can be found in other organically inspired visualizations, such as furry or hairy shapes in Anymails [Car07] or C.E.B. Reas' Puffs [Rea05].

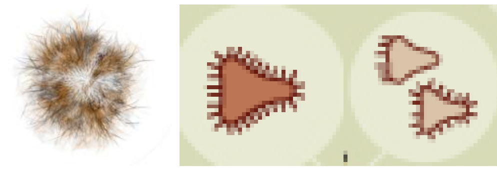


Fig 10. Furry and hairy shapes.

D. Other Fields

Other fields have explored use of multiple shape attributes to convey multiple data attributes. In chemical notation, use of line thickness, line dash style, curvature, angle, terminator/thickness and parallel lines are used.

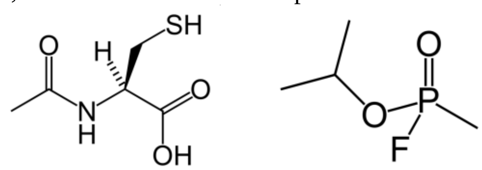


Fig 11. Sample molecular notations utilizing angle, line styles, etc. (from Wikipedia)

Biology extensively utilizes shape in classification, with examples from bird, fish, tree and leaf identification showing multiple different shape attributes used for identification purposes. For example, leaves are identified by a global form (e.g. rhomboid, ovate, deltoid); edge (e.g. spiny, serrate, undulate); and venation (e.g. rotate, longitudinal, parallel):

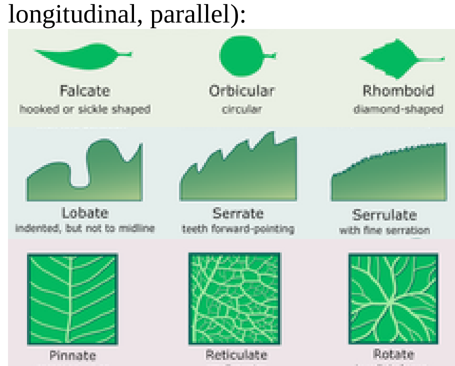


Fig 12. Samples of three shape-based attributes used in leaf identification (from Wikipedia)

Similar to biology, heraldry also utilizes boundaries and has a long tradition formalizing many different types of boundaries with rules for combining them:

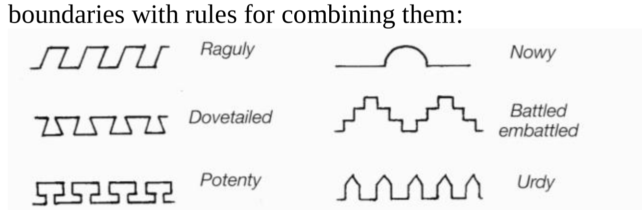


Fig 13. Sample lines from heraldry [Fea08]

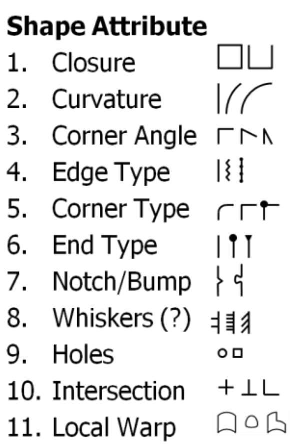
Graphic design is not bound by a set of formal rules but potential shape-based attributes may be outlined in various textbooks and guides or may be inferred from large samples of design. Based on a review of icon categories at Logo Lounge (logolounge.com) and graphic design texts (e.g. [Dre72,Kra04]), the following attributes may also be useful for the design of visualization glyphs, in addition to the shape-based feature list in section 2.1:

- Edge type
- Corner type
- Warp
- Notch

Fig 14. Logos categorized as "particle fields" by Logo Lounge, all displaying warping; and sample corners from [Kra04]

E. Background Summary

This background provides a basis for different shape attributes as used in different fields, and potentially applicable to information visualization. As outlined at IV09's keynote [Bra09], a working list of shape attributes may be considered as follows:



III. Experiments and Lessons

In our work we have been cautious regarding the use of shape. Using more than one shape attribute to convey more than one data variable within a singular glyph has largely been limited to experimental visualizations. These experiments have been done to "see what's possible", have not been rigorously tested, evaluated simply by review by experts within our firm, and acknowledged that some of these experiments are not particularly effective. At this point, the criteria for understanding which combinations may or may not work well together are uncertain; and it would be challenging to test all the various permutations of combinations.

A. What Are the Attributes of Shape?

Which shape attributes to include on the working list of shape attributes is debatable. For example, early on, we considered concavity to be a shape attribute. An "Email Visualization" used basic shapes and then concave variations of those shapes used to indicate a second data dimension. However, within a visual field of many of these shapes, the more complex shapes (i.e. concave shapes) did not seem to visually pop-out. This led to a reconsideration of concavity: is it just a special case of a notch; or did the design of this particular visualization with partially overlapping glyphs interfere with the perception of concavities and thus affect how we were able to visually identify concavity?

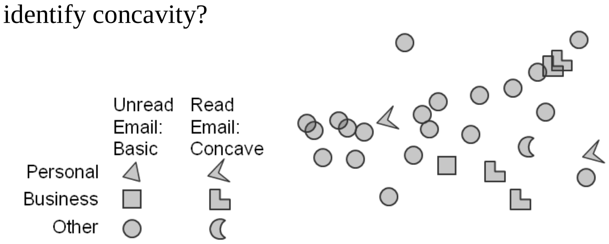


Fig 15. Email visualization relying on concavity.

Furthermore, for any shape attribute, there may be multiple factors available to define that attribute. A terminator at the end of a line or on a corner may have many possible options, including length, height, fill, closure, etc. There is the potential to apply recursive principles to shape attributes, with limits determined by the size of the object and the quality of the resolution – e.g. there are not many options for terminators if the space available is only 2 x 2 pixels; but with interactive zooming techniques, lower priority data variables could be encoded in a "serif" that become visible when zoomed in.

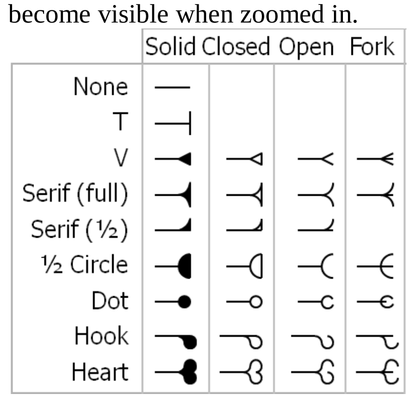


Fig 16. Sample of possible sub-attriubtes for terminators, including type, closure, fill.

B. Like Interferes with Like

A number of experiments utilize the same type of shape attribute to convey more than one variable. For example, a "Chappe Telegraph" visualization uses two angles to convey two data attributes; and a morphological approach uses two curves to convey two data attributes in a "Gas Survey" visualization.

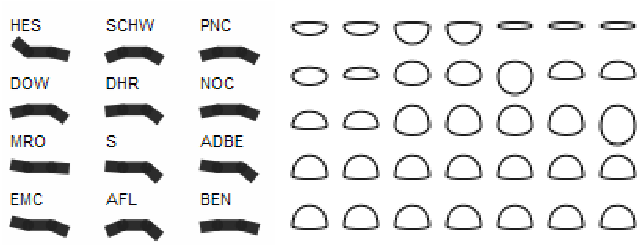


Fig 17. Chappe Telegraph (using angle) and Gas Survey (using curvature)

Other visualization experiments map different data attributes onto different shape attributes. For example, a variant of Gas Survey (4var) maps four different data variables onto four different shape variables: curvature, edge type, terminator and angle. Similarly, "World Demographics" visualization maps three data variables onto three different shape attributes: curvature, angle and terminator.

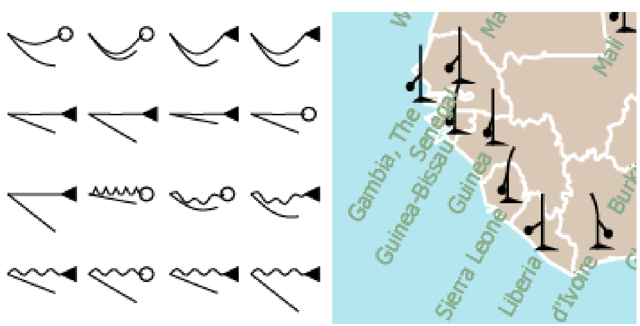


Fig 18. Gas Survey (4vars) and World Demographics (3vars) use different shape attributes for different data.

In general, it seems that it is easier to perceive and understand mappings more easily when different variables are mapped to different shape attributes. This should follow from the general rule "like interferes with like" [War08] or "use different visual dimensions differently" [Bra97] and therefore using different shape attributes can help maximizes distinctness.

However, this general rule must be applied with care. For example, scientific tensor visualization utilizes curvature to convey more than one dimension and seems to be effective. Similarly, a "Stock Correlation Visualization" we created depicts 5 pairs of variables as arms of a star, with twist being used consistently to represent one variable, and bulge being used to represent the second variable.

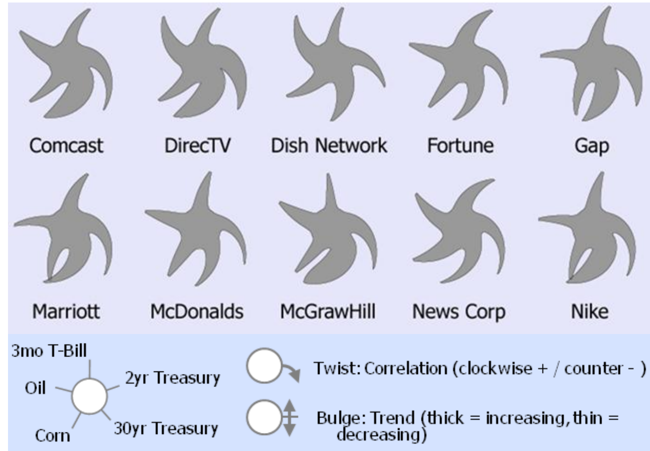


Fig 19. Stock correlations to commodities using twist and bulge to indicate 2 variables of correlation to each commodity.

The corollary to the rule may be that like shapes should be used for like variables, with clear separation, e.g. by use of a transformation.

C. Integral vs. Separable Shape Attributes

The issue of integral vs. separable dimensions [War00, Wil05] is related to the like vs. like issue. Integral dimensions are perceived holistically not independently. Early experiments creating shapes based on morphologies of curved vs. angular corners did not result in glyphs understood as having four separate attributes, but were rather understood as simply being different (fig 20). With the Gas Survey (4 vars) visualization, some of the visual entities are very separable, such as the terminator or angle which can be readily understood as separate from the rest of the glyph. The top line has both curvature and line style which are separable, but presumably not as perceptually separable as the terminator:

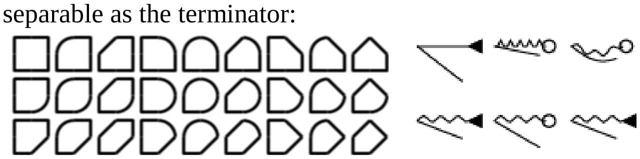


Fig 20. The four separate attributes defining the shapes on left (i.e. corners) are not as easily distinguishable as the four separate attributes defining the shapes on the right.

The issue regarding concavity discussed earlier may be an issue of separable vs. integral, and the particular use of concavity in the experiment is integral, therefore making it difficult to visually parse concavity separately.

D. Shape as a Frame of Reference

Rather than using shape attributes to form a glyph, shape attributes could be used to form a common reference upon which data is displayed, much like Tufte's Small Multiples [Tuf90]: the common reference in each frame provides a basis to locate and compare visual elements across each frame. In this case, the reference shape could represent data structures such as trees, graphs, baselines, etc. This use of shape will require a different set of criteria to understand and evaluate. An example of an experiment is a visualization of the Fortune 500 companies by sector and state:

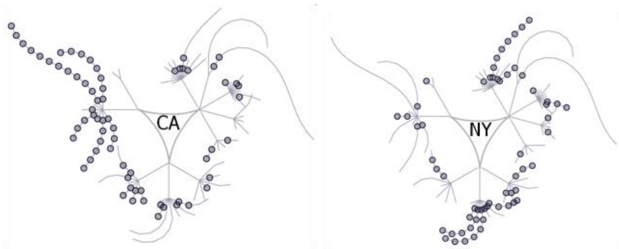


Fig 21. Fortune 500 by sector and state. The background (light grey) tree indicates sector hierarchy, repeated once per each state. Nodes along branches indicate companies.

IV. Conclusions

The use of multiple shape attributes increases the expressive range and the information density of visualizations. The experiments show the potential to convey 1-10 or more different data attributes within a glyph based on shape attributes. Experiments have shown potential value as well as problems regarding effective combination of shape attributes.

There is much future work to be done. At this point in time, we are becoming more comfortable with the working list of shape attributes, although there is uncertainty as to where this list should end, as well as the sub-attributes within any attribute (e..g. a terminator may have both a length and width and therefore represent two data attributes).

We are still unsure of which shape combinations work well together. There is the opportunity for much more research in this area, the results of which would reduce the amount of trial and error in glyph design.

References

[Bar81] A. Barr. Superquadrics and Angle-Preserving Transformations. In *IEEE Computer Graphics and Applications, 1*. p 11-23. 1981.

[Ber67] J. Bertin. Sémiologie Graphique. Paris. Gauthier-Villars. 1967.

[Bra97] R. Brath. 3D Interactive Information Visualization: Guidelines from experience and analysis of applications. In *Design of Computing Systems: Social and Ergonomic Considerations.* Elsevier Science. Amsterdam. 1997.

[Bra09] R. Brath. The Many Dimensions of Shape. IV09. url: www.oculusinfo.com/expertise.html

[Car07] C.Horn, F. Jennett. Anymails. url: carohorn.de/anymails

[Cle85] W. Cleveland. Elements of Graphing Data. Hobart Press. Summit, NJ. 1985.

[Dra08] A. Dragulescu. Malwarez. url: www.sq.ro/malwarez.php

[Dre72] H. Dreyfuss. Symbol Sourcebook. John Wiley & Sons. New York. 1972.

[Ebe99] D. Ebert, R. Rohrer, C. Shaw, P. Panda, J. Kukla, A. Roberts. Procedural Shape Generation for Multi-Dimensional Data Visualization. url: cobweb.ecn.purdue.edu/~ebertd/papers/vissym99.pdf

[Fea06] J. Fearn. Discovering Heraldry. Shire Publications. Princes Risborough, UK. 2006.

[Fry97] B. Fry. Organic Information Design. url: benfry.com/organic/thesis-0522d.pdf

[Kan01] R. Kandogan. Visualizing Multi-Dimensional Clusters Trends and Outliers using Start Coordinates. In *Proceedings ACM SigKDD 01.* ACM. 107-116. 2001.

[Kin06] G. Kindlmann, C. Westin. Diffusion Tensor Visualization with Glyph Packing. url: lmi.bwh.harvard.edu/~gk/vis06

[Kra04] J. Krause. Design Basics Index. HOW Design Books. Cincinnati. 2004.

[Maz09] R. Mazza. Introduction to Information Visualization London. Springer-Verlag. 2009.

[Mac95] A. MacEachren. How Maps Work. New York, NY. Guildford Press. 1995.

[Mac86] J. MacKinlay. Automating the Design of Graphical Representations. In *ACM Transactions on Graphics Vol 5, No.2.* 2006.

[Mac06] J. MacKinlay. url: www.tableausoftware.com/files/designing-great-visualizations.pdf

[Pos95] F. Post, F. Post, T. Walsum, D. Silver. Iconic Techniques for Feature Visualization, in *IEEE Visualization '95*. IEEE. 1995.

[Rea05] C.E.B. Reas. Puff 1-12. url: reas.com/iperimage.php?section=works&view=&work=p6_images3_p&id=0

[Rop07] T. Ropinski, M. Specht, J. Meyer-Spradow, K. Hinrichs, B. Preim. Surface Glyphs for Visualizing Multimodal Volume Data. url: viscg.uni-muenster.de/publications/2007/RSMHP07/vmv07-glyphs_500dpi.pdf.

[Tho05] J.J. Thomas, K. A. Cook. Illuminating the Path: The Research and Development Agenda for Visual Analytics. url: nvac.pnl.gov/docs/RD_Agenda_VisualAnalytics.pdf

[Tuf90] E.R. Tufte. Envisioning Information. Cheshire, CT, Graphics Press. 1990.

[Tuf90] E.R. Tufte. Beautiful Evidence. Cheshire, CT, Graphics Press. 1996.

[War00] C. Ware. Information Visualization: Perception for Design. San Francisco. Morgan Kaufmann. 2000.

[War08] C. Ware. Visual Thinking for Design. Morgan Kaufmann. Burlington, MA. 2008.

[Wei] E. Weisstein. Chernoff Face. from *Math World – A Wolfram Web Resource*. Mathworld.wolfram.com/ChernoffFace.html

[Wil05] L. Wilkinson: The Grammar of Graphics 2nd ed. Berlin, Springer Verlag. 2005

[Wol04] Visual Search. Table 1: Classification of Guiding Attributes(derived but modified from Wolfe & Horowitz) url: www.scholarpedia.org/article/visual_search

[Woo98] A. Woodruff, J. Landay, M. Stonebreaker. Constant Density Visualizations of Non-uniform Distributions of Data. in *Proceedings of the 11th annual ACM symposium on User Interface software and Technology*. 1998.