

Using Type to Add Data to Data Visualizations

Richard Brath and Ebad Banissi

Abstract— Type attributes, such as bold and italic, can be used to represent data in visualizations. A review across domains shows various uses of type attributes and these can itemized and assessed for different ways to encoded data. Numerous examples show how these attributes can be applied to new kinds of visualizations..

Index Terms—Font attributes. Text visualization. Alphanumeric glyphs. Quantitative typography.

INTRODUCTION

Data visualization is the process of turning data into visual attributes, such as the size of a bar on a bar chart, the color of a county on an election map, or the location of a dot on a scatterplot. Unfortunately, most visualization designers today do not leverage typography to express data using type attributes such as bold, italic, condensed and so on. Visualization textbooks identify visual attributes such as size, hue, intensity and texture but relegate text to a single entry (e.g. [1,2,3,4,5]). Type is an afterthought and many visualizations are typographically mute.

For example, figure 1 shows two highly popular data visualizations. Top is a treemap visualization: this treemap is a visualization of the stock market where each box represents a company with the size and color indicating the company's stock market size and the change in its stock price. Boxes are organized by sector with simple labels added on top.

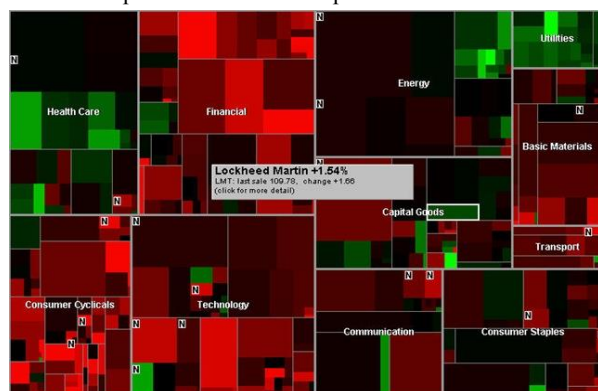


Fig. 1. Popular data visualizations. Top: Treemap visualization of the stock market where boxes indicate companies [6]. Bottom: Tag cloud visualization where word size indicates word frequency from text *Alice in Wonderland* [7].

The lower image in fig. 1 is a tag cloud. Type is important in this visualization - the size of the word is related to the frequency of the word in the original text. However font color and typeface are arbitrary and are not associated with any data.

Instead, the design space of visualization could be much richer harnessing typography to express data encoded directly into text. There are many possibilities that can result from design exploration.

There are three objectives for this research:

1. *Explore* type history for examples where differences in typographic attributes convey data.
2. *Catalogue* those attributes and how they could be used.
3. *Invent* new kinds of typographic visualizations.

1 EXPLORE BACKGROUND

There are many domains that have historically used type to express data. The most obvious may be cartography (e.g. [8,9]), for example, road maps that use all caps to indicate state capitals. The map in figure 2 uses typographic attributes such as reverse italics to indicate rivers, underline style to indicate different administrative levels of cities, as well as all caps, spacing and typeface.



Fig. 2. Map with font size, caps, spacing, various underline styles, and italics/ reverse italics indicating data [10].

In typography, there are many examples. The family tree in figure 3 uses bold to indicate major branches, all caps to indicate regions, small caps to represent rulers, roman text for descendants and italics for spouses.

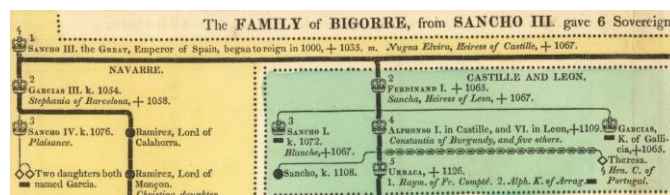


Fig. 3. Genealogical diagram from 1820 with bold, all caps, small caps, roman and italics indicating data [11].

The typographic table of contents in Chambers' *Cyclopaedia* from 1728 uses italics for major branches of science; small caps for specific topics, and superscripts for a chapter number (figure 4).

- Richard Brath and Ebad Banissi are with London South Bank University.
E-mail: richard.brath@alumni.utoronto.ca.
- Manuscript submitted to TypeCon 2015, annual conference of SOTA, August 12-16, 2015, Denver, CO.

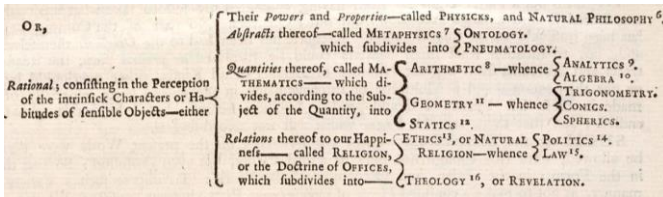


Fig. 4. Portion from Chambers' *Cyclopaedia* 1728 using roman, italics, small caps and superscript [12].

Haeckl's diagram of the *Pedigree of Mammals* uses various typefaces to differentiate between common names, scientific names and so forth for various branches of mammals.



Fig. 5. Portion Haeckl's *Pedigree of Mammals* using different typefaces for different domains [13].

Other domains also use typographic attributes to encode data within text (see [14]) including:

- *Notation Systems* such as mathematical formulas (e.g. $\mu_e(A) = \inf\{\lambda \mid O \mid O \in \partial, A \subset O\}$), chemical formulas (e.g. $[As@Ni_{12}As_{20}]^3$), and markup notation (e.g. `<div class="body">Text </div>`) use different type elements to emphasize, delineate or otherwise add information to text.
- *Alphanumeric Charts* from statistics and finance may use attributes such as color, underlines and bold.
- *Data Visualization* sometimes does use text with typographic attributes. E.g. Skupin visualizes knowledge domains using cartographic labelling techniques such as font size, orientation and spacing [15].
- *User Interface Design* historically recommended against using font attributes (because of low resolution displays) although current web design references now often recommend using font attributes and modern code editors use type attributes such as bold, underline and italics to highlight programming syntax.

2 CATALOGUE OF TYPE ATTRIBUTES

Based on an analysis of type use across domains and typography texts (e.g. [16,17,18,19]), potential font attributes can be identified as outlined in table 1. The second column lists the font attributes with the first column a grouping. The final attributes at the bottom - x-height, contrast, stress angle, serif length, etc - are font attributes intrinsic to a typeface's design and more speculative, as examples of these are not found in historic uses.

Group	Font Attribute	Best for encoding: Q: quantitative O: ordered C: categoric G: grouping/relationship L: literal	Example
Glyphs	Alphanumeric Text Glyph Symbols	L, O	ape bat cat dog 123 456 ! ? # @ % comment \$var
Font Family Attributes	Font weight	C (2-9 levels)	1.0 2.0 3.0 5.0 8.0
	Oblique / Italic	C, Q using slope angle	-2.0 -1.0 0.0 1.0 2.0
	Case inc small caps	C, possible O (2-3 levels)	BIG AVG. Small tiny
	Typeface	C (2-6 levels)	Swiss French German Italian
Sequence	Underline	C, O, Q (using length)	plain dash single double
	Condensed	Q, O (2-4 levels)	1200 2000 3000
	Squished	Q	average thin plain wide fat
	Spacing	Q, O	tall grand venti
Font Design?	Super/subscript	C (2 levels)	Normal ^{High} (but) *and* <or>
	Delimiters	G	ick ick ick
	X-height	O, Q (few levels)	LOW MED. HIGH
Font Design?	Contrast / Stress angle	O (few levels)	
	Serif length / Bracket size	O, Q (too small to see?)	

Table 1. Font attributes and types of encoding.

The third column indicates how these attributes can encode data as identified by information visualization researchers and cartographers. For example, quantities, can be encoded using font weight or obliques; while typeface is best used to encode different categories. The final column shows some examples of each attribute.

Some visual attributes visually pop-out from their surrounding - being perceived almost immediately regardless of the number of other items (figure 6) [20].

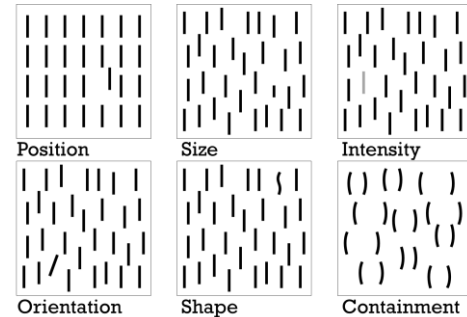


Fig. 6. Preattentive visual attributes. A variation in these attributes can visually pop-out from surrounding items. The bottom left shows pairs of symmetric shapes, where an asymmetric pair can pop-out.

In psychology, these visual channels are called pre-attentive and some of these attributes have a stronger pop-out effect than others [21] (in typography, this effect is referred to as color, where the objective is to reduce any pop-out effect in a well-designed font.) In turn, each font attribute can be related back to these preattentive attributes, to get a sense of which may be more effective at visually standing out (table 2). For example, font weight, using visual channel of intensity and size is likely to be more effective at visually standing out than case. Although size is preattentive, serif size or bracket size may be too small to be noticeable, and further small changes may be subject to change blindness.

Group	Font Attribute	Visual Channel					
		Position	Length/Size	Intensity	Orientation	Shape	Containment
Glyphs	Alphanumeric Text Glyph Symbols						D
	Font weight		•	♦			HP
Font Family Attributes	Oblique / Italic				♦		HP
	Case inc small caps		•			♦	P
	Typeface					♦	P
	Underline	♦	•				HP
Sequence	Condensed		♦	•			HP
	Squished		♦	•			HP
	Spacing		♦	•			HP
	Super/subscript	♦	•				HP
Font Design?	Delimiters						D
	X-height		♦				HP
	Contrast / Stress angle					♦	P
	Serif length / Bracket size		♦				HP

♦ / • indicates primary / secondary visual channel for font attribute
† HP: Highly probable, P: probable, D: doubtful

Table 2. Font attributes and the relation to visual channels. Final column indicates probability of preattention [21].

3 NEW VISUALIZATIONS

Given a set of font attributes, the next question is how they might be used to create new kinds of visualizations.

3.1 Skimming Text

Text skimming is a reading technique of rapid eye movement across a large body of text to get the main ideas and content overview. At a low level, the strategy requires the reader to dip into the text looking for words such as proper nouns, unusual words, enumerations, etc. To make uncommon words pop-out, word frequency analysis can be used to identify the least common words and then make these have the heaviest weight. Figure 7 shows the opening paragraph of *The Wizard of Oz* formatted for skimming.

Italics are also used for less important parts of speech (e.g. articles, pronouns) to create greater differentiation from the heaviest weight text and enhance figure-ground separation.

Dorothy lived in the midst of the great Kansas prairies, with Uncle Henry, who was a farmer, and Aunt Em, who was the farmer's wife. Their house was small, for the lumber to build it had to be carried by wagon many miles. There were four walls, a floor and a roof, which made one room; and this room contained a rusty looking cookstove, a cupboard for the dishes, a table, three or four chairs, and the beds. Uncle Henry and Aunt Em had a big bed in one corner, and Dorothy a little bed in another corner. There was no garret at all, and no cellar except a small hole dug in the ground, called a cyclone cellar, where the family could go in case one of those great whirlwinds arose, mighty enough to crush any building in its path. It was reached by a trap door in the middle of the floor, from which a ladder led down into the small, dark hole.

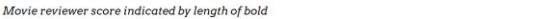
Fig. 7. First paragraph from *The Wizard of Oz* formatted for skimming.

3.2 Reviewing Opinions

The movie review website, *Rotten Tomatoes*, has potentially hundreds of reviews per movie. Attempting to get an overall sense of the reviews across some different movies is difficult. Instead, figure 8 shows a list of movie reviewer's quotes for two movies. The amount of bold along each line indicates the overall reviewer's score - a short amount of bold is a poor score while full bold indicates the highest score. For a single movie, one can see what the range of reviews are - e.g. how strong is the difference of opinion between the best and worst reviews; and also facilitates comparison across reviews - for example a review that has more total bold is a higher scoring movie overall.

Despicable Me 2

This is a sequel that's even less necessary than *Monsters University*; often times it feels like *Despicable Me*, the animated supervillain comedy from 2010, was an average flick with Given the outlandish premise, you'll wish the film twinkled with a more savvy sense of cynicism and detractors, it may at times feel like this sequel exists for nothing more than Cute family fun, but lacks the pop of the original. Gru has gone from despicable to do its hyperactive vibrancy is universally boredom-proof. Not a great movie for sure, but if your kids want to see this there is enough humor to Gru still has charm and kids will adore the Minions. Steve Carell's Slavic inflections as Gru do the trick, as before. Wiig's clever hesitator Once again, there's nothing here that's particularly original or memorable, but the ch The film easily surpasses the original, while leaving room for further sequels. An animated sequel that, despite not achieving the inspired lunacy of the first movie Parts James Bond flick, "Get Smart" episode and Pixar-esque family adventure, "Des Ranks as one of the best animated sequels of all time. Though jammed-up with too much pointless plot, *Despicable Me 2* remains one of the Not as consistently funny as the original, *Despicable Me 2* still proves itself a quite- The pratfalls, gizmos and Looney Tunes 'violence' will elicit giggles from kids while



How to Train your Dragon

Everything from the angle of the shot to the speed of the editing projects an end visu Here, Viking life is grim, hostile and heavy with social pressure -- kind of like Gossip The visuals are striking, the script sharp and well paced and it all wraps up with a bre Full of wonder, charm and dragons not doubling as stand-up comics. It's a brisk, amusing piece that doesn't have the weary sarcasm that besets a Shark Following a slow, overly verbal start, this dragon tale takes flight. Baruchel, Ferguson and Butler supply a contagious sense of eccentricity that spread Beautifully animated and superbly written, this is a hugely entertaining, frequently fu Magical storytelling that makes perfect family entertainment for the Easter holidays It's not only better than I thought it was going to be, it's a lot better than I thought it v Beautifully crafted and effortlessly entertaining, this is an unexpected triumph. It respects its audience enough without sticking to cheap shots and bad jokes. The dragon designs are wonderful, the action is exciting and the anti-wormongering How to Train Your Dragon is a visual marvel, and not just because it's in 3-D. Knowledge is power in this film, and I always love a good pro-intellect story, even be Though the 3-D effects are awesome, this movie also succeeds in two-dimensions Undoubtedly Dreamworks' best film yet, and quite probably the best dragon movie e

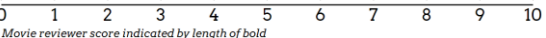


Fig. 8. Movie reviews from *Rotten Tomatoes*. The length of bold indicates the reviewer's score.

3.3 Describing Topics

While tag clouds (fig. 1) can possibly show topics in a text, they don't provide any qualities associated with the topic. In figure 9, characters from *Grimms' Fairy Tales* are listed along the left side and adjectives that occur within +/- three words are listed to the right. Adjective frequency is indicated by font weight: one can see that kings are mostly old, princesses are beautiful and girls are little.

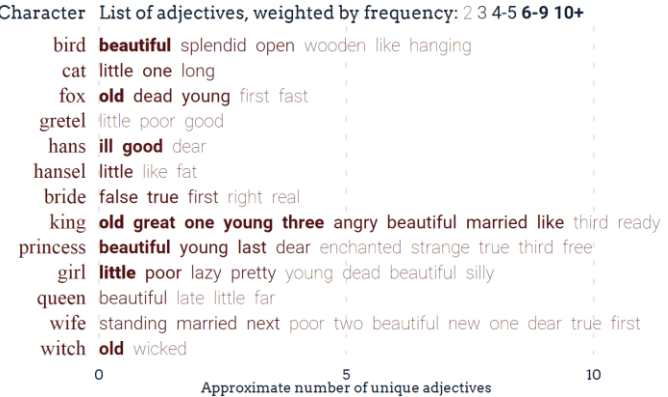


Fig. 9. Adjectives associated with characters from *Grimms' Fairy Tales*. Heavier weights are more frequent.

3.4 Comparing Countries

A choropleth map is a popular visualization technique that colors each country based on a data value associated with that country. Choropleth maps have problems, such as small countries (e.g. Singapore, Luxembourg) are not visible (fig. 10 top) and only one variable is shown at a time.

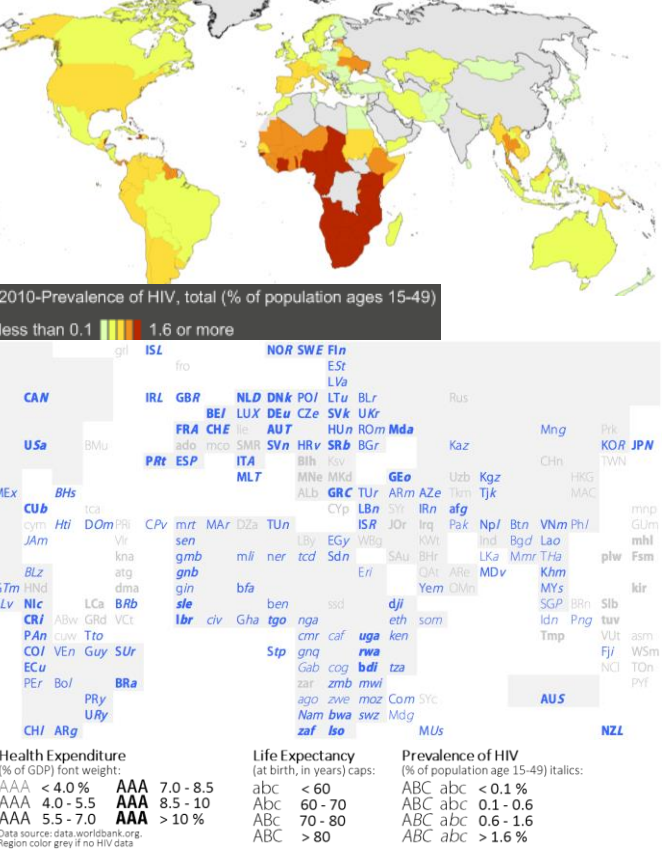


Fig. 10. Top: Choropleth map shows HIV prevalence by country color. Bottom: Font weight, caps and italics indicate health expenditures, life expectancy and HIV prevalence respectively.

Instead, the map using three letter country ISO codes can represent multiple variables using font weight, case and italics (figure 10 bottom). Small countries are clearly visible. Complex questions can be answered, such as countries with high spending and short lives (yes, e.g. Rwanda *rwa*) or low spending and long lives (yes, e.g. Luxembourg, LUX).

3.5 Songs in Text

In written prose, singing may be differentiated from surrounding text, for example, by being set in italics. However, this does not convey any of the song qualities such as the note pitch and note duration. While traditional music notation could be used, this would interrupt the flow of the text and require a lot of space. Instead, syllables could use x-height to indicate note pitch and compressed/expanded fonts to indicate note duration (fig. 11).

Frè-re Jac-ques, frè-re Jac-ques,
Dor-mez VOUS? Dor-mez VOUS?
Son-nez les ma-ti-nes! Son-nez les ma-ti-nes!
Ding, dang, dong. Ding, dang, dong.

Fig. 11. Song text with note pitch indicated by x-height (i.e. high notes have a tall x-height, e.g. vous) and note duration indicated by font width (i.e. short notes are narrow, e.g. son-nez; and long nodes are expanded, e.g. dong).

4 CONCLUSION

The examples here suggest that there is a wide range of possibilities for unique visualizations using font attributes. Examples shown cross many domains. Many of the examples use font weight and other font attributes are under-explored. A few additional examples can be found in [22,23,24].

Some interesting implications arise from this work that may be of interest for further investigation by type designers. One example uses variable x-height: traditionally the x-height is an inherent property of a typeface. Is it feasible to have a type family with multiple x-heights?

Many of the examples use multiple font weights. These are largely restricted to sans serif fonts which currently provide the greatest variation in weights. Can a serif typeface be provided in a broad variety of weights? Is it feasible to design a lightweight blackletter or a black cursive?

Many of the examples use multiple font attributes within a single visualization, for example combining variations across weight, italics, case; or combining variations in x-height and font widths. Can type families be created that provide even more variation across attributes, for example, a typeface with multiple weights and multiple oblique angles? Given potentially a resulting very large type family - how is it installed and licensed?

Attempting to mix and match different typefaces to encode different categories can be difficult. For example using a serif, sans serif and a slab serif is difficult when it is desired to have matching weights, oblique angles and so forth.

From a data visualization perspective, while there are many examples shown here - these are all speculative applications and have not been tested with real users. How effective these techniques are has yet to be established. However, the breadth of examples suggest rich possibilities for future work.

REFERENCES

- [1] J. Bertin. *Sémiologie Graphique*. Gauthier-Villars, Paris, 1967.
- [2] M. Chen and L. Floridi. An analysis of information in visualization. *Synthese*, 2013.
- [3] R. Mazza. *Introduction to Information Visualization*. Springer-Verlag, 2009.
- [4] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2015.

- [5] C. Ware. *Information Visualization: Perception for Design*. Springer-Verlag, 2000.
- [6] M. Wattenberg. "Visualizing the stock market." In *CHI'99 extended abstracts on Human factors in computing systems*, pp. 188-189. ACM, 1999.
- [7] J. Feinberg. "Wordle". In *Beautiful Visualization: Looking at Data through the Eyes of Experts* (J. Steele, N. Illinsky eds.) O'Reilly Media, 2010. www.wordle.net. Accessed 08/08/2015.
- [8] J. Krygier. *Making Maps: A Visual Guide to Map Design for GIS*. Guildford Press, New York, 2005.
- [9] K. Börner. *Atlas of Knowledge: Anyone Can Map*. MIT Press, Cambridge, Mass, 2015.
- [10] A. Stieler and H. Haack. *Stieler's Atlas of Modern Geography*. Justus Perthes' Geogr. Institute, 1925. via www.davidrumsey.com 06/06/2015
- [11] M. Carey and M. Lavoisne. *A Complete Genealogical, Historical, Chronological, And Geographical Atlas; Being A General Guide To History, Both Ancient And Modern*. Philadelphia: M. Carey And Son, 1820. via davidrumsey.com 06/06/2015.
- [12] E. Chambers. *Cyclopædia*. London. 1728. via [digital.library.wisc.edu/1711.dl/HistSciTech.Cyclopaedia](http://digital.library.wisc.edu/1711.dl/HistSciTech/Cyclopaedia) accessed 07/30/2015.
- [13] E. Haeckel. *The Evolution of Man*. Watts & Co., London, 1906.
- [14] R. Brath and E. Banissi. "The design space of typeface." *Visualization and Computer Graphics, IEEE Transactions on*, 2014.
- [15] A. Skupin. "The World of Geography: Visualizing a Knowledge Domain with Cartographic Means." *PNAS* 101 (Suppl. 1): 5274-5278.
- [16] K. Cheng. *Designing Type*. Laurence King Publishing, 2004.
- [17] J. Craig, I. K. Scala, and W. Bevington. *Designing with type: the essential guide to typography* - 5th Ed. Watson-Guptill, New York, NY, 2006.
- [18] V. Squire, H. P. Willberg, and F. Forssman. *Getting it Right with Type*. Laurence King Publishing, 2006.
- [19] E. Lupton. *Thinking with Type: A Critical Guide for Designers, Writers, Editors & Students* - 2nd ed. Princeton Architectural Press, 2010.
- [20] C. G. Healey and J. T. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18:1170-1188, 2011.
- [21] J. Wolfe and T. S. Horowitz (2008) *Visual search*. Scholarpedia, 3(7):3325.
- [22] R. Brath and E. Banissi. "Using Font Attributes in Knowledge Maps and Information Retrieval." *Proceedings of Knowledge Maps and Information Retrieval (KMIR) at Digital Libraries* (2014).<http://ceur-ws.org/Vol-1311/paper3.pdf>
- [23] R. Brath and E. Banissi. "Using text in visualizations for micro/macro readings". *IUI Workshop on Visual Text Analytics*, 2015. <http://vialab.science.uoit.ca/textvis2015/papers/Brath-textvis2015.pdf>
- [24] R. Brath and E. Banissi. "Font Attributes Enrich Knowledge Maps and Information Retrieval." *International Journal on Digital Libraries*, Springer. forthcoming.