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.

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.

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.

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 descendents and italics for spouses.

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).
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.

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

- **Notation Systems** such as mathematical formulas (e.g. $\mu(A) = \inf\{\lambda(O) \mid O \in \mathcal{O}, A \subset O\}$), chemical formulas (e.g. \texttt{As(NO$_2$)$_3$}) and markup notation (e.g. \texttt{<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.

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].

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.

In Table 1, font attributes are ordered by importance in visual processing, and then by decreasing font weight. The top five attributes are most likely to be noticed, such as color or case, all of which are considered preattentive. The second five attributes are next in importance, followed by the last five attributes which are the most difficult to notice.

**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 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.
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.

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 Grimm’s 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.

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.
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 **rwu** 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).

**Frère Jacques, frère Jacques, Dor-mez voici, Dor-mez voilà, Son-nëz les ma-të-nëz, Son-nëz les ma-të-nëz! Ding, deng, deng, deng, deng.**

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 notes are expanded, e.g. deng).

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