Microtext Line Charts

Richard Brath and Ebad Banissi
London South Bank University and Uncharted Software, London South Bank University
{richard.brath@alumni.utoronto.ca, banisse@lsbu.ac.uk}

Abstract

Timeseries line charts are a popular visualization but traditionally do now show many lines. We borrow concepts of tiny microtext and path dependent cartographic text to embed labels and additional text directly into lines on line charts thereby making it easier to identify individual lines in a congested line chart, enabling more lines to be displayed and enabling additional data to be added to the lines as well.

Keywords--- **typography, visualization, line chart**.

1. Introduction

Line charts are an extremely popular type of visualization. In financial charting systems, line charts and variations on line charts dominate over all other types of visualizations of timeseries oriented data (based on conversation at [BCK14]).

A simple line chart with one line doesn't require labeling: the title can unambiguously indicate the line. But with multiple lines, some way to distinguish between lines is required, such as a use of colors which can be decoded by cross-referencing a legend. There is cognitive effort required to refer back and forth between a line in a chart and a descriptive legend [LS87] and furthermore, the simple use of color and/or dash patterns is difficult to scale beyond 10 or so lines (e.g. see [War13]).

Our unique contribution is 1) to move labels directly into the plot area of timeseries charts to allow for many more lines in a chart, and 2) to extend this approach to utilize various typographic attributes to further disambiguate between lines.

2. Background

Type has been used extensively in information visualization for labels. For example, many knowledge visualizations on *scimaps.org* are label-dense: however the labels is simply used to identify the underlying graphical object. For example the Usenet Treemap [SF04] or Skupin's cartographic SOM [SK04] use labels on top of an underlying visualization.

In addition to labels being plain, labels in information visualizations tend towards largish sizes (8-12 point) as opposed to fonts used in print. Historically, information visualization has been associated with interactive video displays, which until the late-2000's were limited to 72-96 pixels per inch (PPI). New high resolution displays are becoming much more common and these displays enable the depiction of very small yet still legible fonts. For example, some mobile phones now exceed 400 pixels per inch and their applications may use fonts as small as 3 physical points.

Note, measurements in points herein use the traditional typographic definition of points, i.e. one point = 1/72 of an inch, or 0.35mm. In some modern programming contexts, a point is defined in relation to the pixels, e.g. in CSS points are relative to pixels, where 1 pixel = .75 points, and when a font is defined in point size it will render at different physical sizes based on the target device's pixel density. For example, Apple iOS guidelines recommend a minimum point size of 13 points, which in turn will render physically at 3.1 points on a 400 PPI iPhone 6plus.

2.1. Small font sizes

In paper-based cartography, small fonts have been used for centuries. Minimum font sizes in cartography are defined as 3 point [RMM95] or 4 point [Hod03], with guidelines recommending 5 point [RMM95] or 6 point [Hod03] as a minimums. Similarly, information graphics in print had very small minimum point sizes, e.g. [Bri39] recommended a minimum size of 4 point, which was the smallest size supported by Monotype machines at the time.

Figure 1. Sample Bell Centennial font, specifically defined for print at small sizes.

Text in print environments can go even smaller: microtext is extremely small sized text and originated as an anti-counterfeiting device for bank-notes and official government documents, as well as archival texts. The quality and legibility of text becomes more difficult to

produce in print as text becomes smaller. For example, Matthew Carter, famously designed the font Bell Centennial to be legible when printed in phone books - i.e. printed at small sizes on high speed printers on poor quality paper - and had special features such as notches in branches of letters to act as ink traps to prevent the spread of ink to form blotches in crotches and holes in letters (fig. 1).

2.2. Type as Texture

Microtext can be clear at extremely small sizes, although the ability to decipher the text depends on the eyesight of the viewer. Microtext can work as a texture due to a unique feature of typography referred to as colour by typographers. A well-designed font has an even distribution of text ink across a sequence of letters regardless if the letters are sparse (e.g. i,v) or dense (e.g. m, e): a blurred paragraph from such a font will appear as an even distribution of intensity. In the example in figure 2, the apparent light brown shading behind the large word CANADA in the center of the banknote, is tiny (one point) microtext when inspected closely.

Figure 2. Example banknote. Central stripe appears as a texture, but is can be read as text (BANKOFCANADA2) when inspected closely.

This feature of text as texture can be leveraged in visualizations. As a texture, type has been used in some novel visualizations such as automatic typographic maps [AMJ12], which uses text to completely fill areas on maps with labels, removing need for edges and fills (fig. 3 top). The approach is similar to the micro-text in fig. 2 above - i.e. labels are repeated until the area is filled.

Fatfonts [HNC12] are used to create a weighted texture: this is a specially designed font where numerical glyphs are weighted such that large numbers have proportionally heavier weights (fig. 3 bottom).

Figure 3. Top: Typographic maps, Bottom: FatFonts.

2.3 Text aligned to paths

Some guidelines recommend against text at angles such as axis labels (e.g. [WWP96]). However, there are many cartographic examples with text labels at angles or along paths, for example, aligned to rivers, at angles to coast, or to align with geographic areas, such as the example from Stieler's Atlas [Sti1881] in fig. 4 left (courtesy davidrumsey.com) or Google maps (fig. 4 right).

Figure 4. Angled and bent text labels.

Furthermore, there are many historic precedents of longer phrases or sentences along paths such as early word balloons for example from a late medieval book of hours [Unk1300] or an early renaissance woodcut [Car1524] as shown in figure 5.

Figure 5. Historic word balloons with lines of text flowing along curving scrolls.

More recent examples exist as well. While moveable type set in rectilinear arrays may not have facilitated free form text layouts, some authors experimented with text freed from straight left-to-right lines such as poets. Apollinarie's concrete poems from the early 1900's playfully adjust type layout based on subject matter as

shown in the poem *Visée* [Apo1918] (Figure 6 left). Modern graphic design tools, such as Photoshop and Illustrator provide easy-to-use tools to set text along paths. The programmable graphics format SVG natively provides functionality to place text along paths.

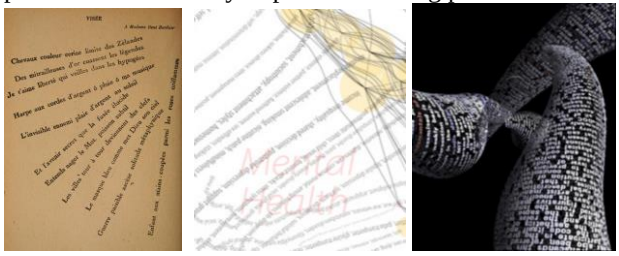

Figure 6. Organic layout of lines of text.

Within the field of information visualization Brad Paley's *Map of Science* [Pal07] has long flowing labels, in this case the nodes represent scientific topics and labels are a long list of topic words: to avoid overlapping labels, the paths for labels bend to avoid collision or cross each other at right angles to better maintain readability (Figure 6 center). Ben Fry's Tendril [Fry00] renders the text of webpages as twisting 3D cylinders with hyperlinks spawning new orthogonal cylinders of text (Fig. 6 right).

3. Timeseries Analysis

In financial services, timeseries charts have been used for more than 200 years, going back to William Playfair's charts and Japanese candlestick charts. By the early 1900's, organizations maintained and updated these physical paper charts, potentially having many charts forming *chart rooms* and *chart libraries* as seen in fig. 7 [Bri14].

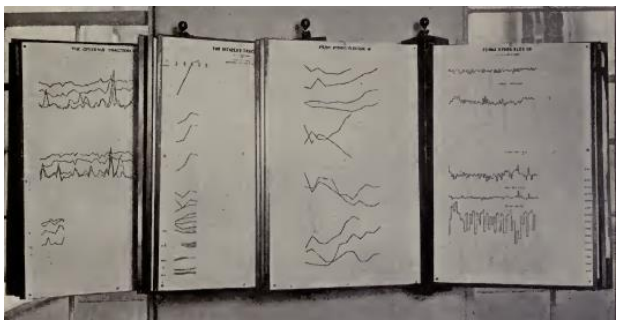

Figure 7. Pivoting boards for organizing many timeseries charts.

3.1. Non-interactive Charts

Although personal computers in the 1980's made it possible to interactively plot financial timeseries charts, variations of non-interactive charts have persisted through to today. The current notion of a financial *chart book* or *chart library* is a collection of time series charts which exist in a non-interactive format. Chart libraries may consist of hundreds of these charts. These charts may be consumed in many different ways, such as paper output (e.g. a "chart book"), a large PDF file, or used as content embedded into commentary, such as financial research reports. In spite of highly interactive computer systems, these non-interactive charts continue to persist.

The author has had occasion to view, analyze and discuss these chart libraries with users and experts in the capital markets, including a large mutual fund company, a large bank, a market data provider and so forth. Users and experts claim benefits for these non-interactive charts including:

- *Ability to quickly flip through many charts*: Flipping through paper or pages in a PDF file is considered faster than stepping through a software system that dynamically retrieves data and updates charts.
- *Familiarity*: The users have seen these particular charts many times through their careers (e.g. weekly or monthly) and may also return to them for reference as needed. Off-the-shelf financial charting software systems may not store all their preferences, may not maintain consistency over time (e.g. scales change), may not retain user created notes and trend lines, and so forth are some reasons provided for favoring these charts.
- *Interactivity is not required*: The users want to be able to see all the series and the full scope of data: common computer-based visualization interactions such as zoom, filtering, changing scales, and so forth are expressly not desired. For example, they have become familiar with the sizes of items and may even refer to some aspects of such a chart with physical dimensions, e.g. "a quarter inch movement in a line".
- *Annotation*: Pen, pencil and ruler can be used to achieve annotations more flexibly than some interactive annotations.
- *Higher Resolution*: Some practitioners will print these charts out as a physical chart book. The print version can be higher resolution than the screen version. High quality 1200 DPI printers, in theory, provide 15 times more resolution (120m dots) than current state of the art 4k screens (8 megapixels).

Table 1. Number of timeseries per chart in a financial chart library.

NUMBER OF SERIES	NUMBER OF CHARTS
1-5	148
6-10	83
11-15	21
16-20	9
21-25	1

While these charts are not used interactively, they are for the most part created computationally. They may be created with heavily customized off the shelf software (e.g. Excel) or other custom software. These chart libraries are considered a proprietary confidential asset to a financial firm, therefore it is typically difficult to get extended access to many of the charts. The author was able to review all the timeseries charts in one

organization's chart library. This library consisted of 262 timeseries charts. Most of the charts (57%) displayed 5 or fewer timeseries. However, 31 of the charts (12%) displayed 11 or more timeseries, with the current maximum a chart with 23 timeseries, summarized in Table 1.

Note that this firm also had other timeseries charts, for example one with 83 series, however these charts with a higher number of series were consumed only in an interactive format, typically toggling on only a few series at a time for comparison; whereas in the chart library all the charts were always plotted with all the lines.

There are various challenges with these charts when the number of series is more than ten:

- *Differentiation*: Charts typically differentiate between series using hue, however, it can be difficult to scale beyond ten unique hues [War13]. Strategies used included a combination of hues and brightness (e.g. bright red, dark red, bright green, dark green, etc); or hue and dash (e.g. red continuous line, red solid line, green continuous line, green dash line).
- *Labeling*: Labels associated with each series is typically indicated in a separate legend. With many different series, it can become difficult to cross-reference between the line in the chart and the legend entry, for example the hue differentiation may be subtle. Larkin and Simon [LS87], for example, noted one benefit of diagrammatic reasoning is reduced cross-referencing, however this benefit is lost with a separate legend.
- *Other Layouts*: Small multiples, horizon charts and so forth were not acceptable alternatives: lines needed to be superimposed on a common scale for detailed visual comparison.

3.2. Line Labelling Alternatives

There are some known alternatives to legends for these many line time series charts.

Labels at the line start or end. Rather than a legend, labels can be depicted at the start or end of the line. This can be challenging if there are many lines starting or ending at similar values.

Labels in the plot area. Rather than rely on the legend, some users add textual annotations in the plot area of the chart, placing a straight line of text near the target line, potentially with a leader line to visually connect the label with the line. This has the benefit of reducing the cross-reference to the legend (e.g. Figure 8 left [Bri39]). However, it does clutter the plot area and some users are adamant against this. This is consistent with the arguments that labels can make it more difficult to see patterns formed by data, e.g. [Cle94, Few06].

Labels aligned to paths. In historic hand-drawn examples of timeseries charts - such as Playfair, or pre-computer generated charts - the authors' hand-letter charts and can thus easily align text to shape of the line (Figure 8 right). This has the benefit of reducing clutter and more directly associating the label with a line as opposed to the straight line of text.

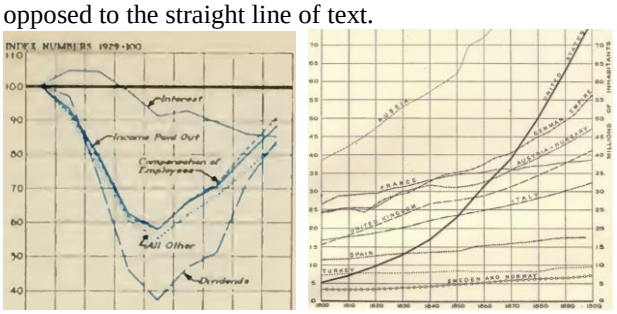

Figure 8. Labels in the plot area.

Labels aligned to graphical shapes are common in on-line maps (e.g. Google street map), however, they are uncommon in current time series visualizations. Directly labeling lines has challenges with occlusion (labels obscuring each other or the lines).

4. Microtext Financial Timeseries Charts

Microtext can be used directly label lines. Using local labels means that cross-referencing a legend or visually following a line back to an end label is not required - the label is local to the area of inspection. There are alternatives as to how the labels can be utilized:

Labels aligned to paths. Similar to river labels on maps, labels can be aligned to lines. Within a congested line chart, some care must be taken to reduce overlap. In Figure 9, a line chart showing unemployment rate for 37 countries is shown, with labels both at the end of the series and a few labels along the length of the times series (in 7 point font). A simple collision detection algorithm has been used to push the labels apart to minimize overlapping labels.

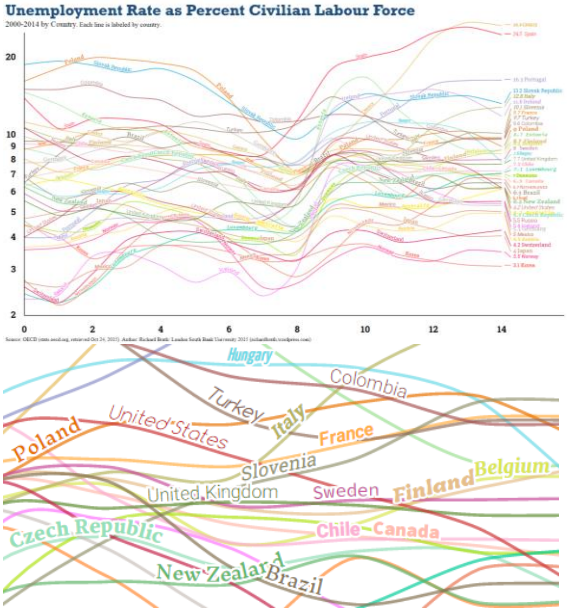

Figure 9. Congested timeseries chart with labels along lines. Overview top, closeup below.

Text as lines. With high resolution displays, line can instead be drawn directly as continuous micro-text, similar to the continuous text in Automatic Typographic Maps. In Figure 10, lines are replaced with five point microtext and appear perceptually similar to a dashed line.

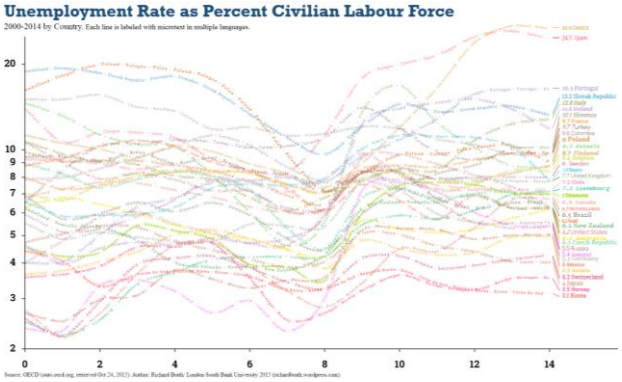

Figure 10. Congested timeseries chart with labels for lines. Closeups in Fig. 11.

Since lines cross and overlap, some consideration must be given to text legibility at intersections.

Halo: In geographic maps, labels may be placed over top various other graphics. A halo may be added around the text to make the text more legible (e.g. TileMill *text-halo-fill*). In a congested line chart, however, there may be significant areas of overlap between lines. Adding a halo is feasible, but many other microtext lines may be partially occluded reducing line visibility and reducing legibility of partially obscured text. See fig. 11 top.

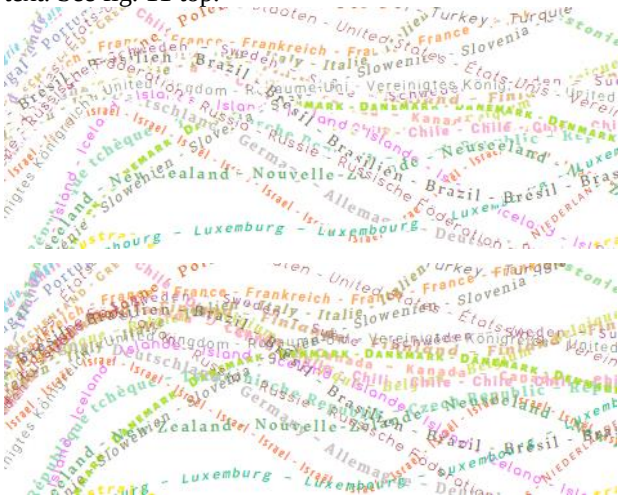

Figure 11. Top: text with halos to improve legibility. Bottom: overplotted text.

No halo: In the lower image of fig. 11, no halo is provided around text. At holes in letters and gaps between letters, the color of text beneath appears through the higher text, and provides a better indication of line density than halos. Assuming different colors of text, one can visually trace other microtext lines and maintain some degree readability through crossings, although not through areas of high congestion. For example in fig. 11 compare Israel (red, near bottom) in both the top image (with halo) and lower image (without halo). The path of line is more easily discerned through the many crossings in the lower, non-halo representation. Or, Finland, in the middle image (brown) is complete obscured, while in the lower image, although often mixed with other text, can be somewhat discerned with some effort.

To further enhance differentiation between the microtext lines, each line is set with different font attributes: each line varies in font weight, italics, case and font family, in addition to colour.

Both microtext along a path or microtext as lines can be combined with other labeling techniques, such as labels at the ends of lines, or a legend. The viewer can choose to ignore the microtext and use these other aids to decode the lines.

Another benefit is that the labels can vary: rather than repeat the same text, the microtext can be multilingual, extending the usefulness of the chart across nationalities - as seen in fig. 12.

Figure 12. Multilingual microtext lines.

5. Tasks, Observations and Discussion

The microtext timeseries were evaluated by six different experts in capital markets who use non-interactive chart books and chart libraries. All the subjects have at least 10 years expertise in capital markets specifically with financial charts and visualization; and have affiliations or certifications in professional financial analysis. All use financial charts in their daily work. All work in a capacity where they may be sharing their charts and observations with other people, such as clients.

The youngest participant was 35, the oldest 65, with an average 50. All were Caucasian men - as were most of their clients. All had limited time availability: questions and tasks were kept simple. No formal testing apparatus was set up: these were meetings at the subject's office.

5.1. Tasks and Questions

1. Initial Question: How many lines do you have at a maximum in your charts? Why?
2. Then each of three *Unemployment Rate as Per-cent of Civilian Labour* Force charts were presented in sequence: first standard line chart with only labels at the ends of lines; followed by one of the two micro-text versions the followed with the other (i.e. fig. 9, 10). For each variant the following two questions were asked:
   a. Which country was top (or bottom) in year x?
   b. How did country x fare relative to its peers

through the 2008 recession?

3. Follow-up Question: Are these techniques rele-ant to the kinds of analysis that your team does? Can you identify use cases where you would consider using this kind of chart?

5.2. Observations

With respect to the first question, the number of lines on a chart, only one subject out of six claimed that he did not need to plot more than 15-20 lines. Another claimed that while most charts used a low number of lines, there were cases where they could go much higher - beyond 50. The most insightful answer came from an analyst: the community was constrained by their tools and it was very difficult to make an effective chart with more than 10 lines so people have been conditioned to keep their charts simple - however there are actually many cases where one would want to view more lines if feasible, assuming the tools and representation were appropriate. This user pointed out that many peers and subordinates used Excel, which was poorly suited for working with many simultaneous lines.

With regards to the tasks and responses to the three chart variations: When presented with the first task (highest/lowest country in year x with just a line chart), the response was either a dismissive comment (e.g. that's not going to work) occurring for 4/6 subjects or taken as a challenge (i.e. attempting to trace the line to find the right answer, requiring more than ten seconds to do so). If the initial answer was dismissive, the second task country x through recession) was not asked.

For the second chart, the subjects responded almost instantaneously with a visceral response, such as "Wow, this is exciting", or "I really love this." One interesting aspect was 4/6 subjects physically traced their response using their finger, e.g. for the top in year x question, sliding across the x-axis to the year, moving up to the target line, moving across to the first local label to the answer (requiring no more than 4 seconds).

The second task (how did country x fare through the recession) was done typically by tracing vertically near the beginning of the recession until the target was identified, tracing across through the recession to a matching label on the opposite side to confirm they were on the same line, and then an observation relative to the rest of the peers, e.g. "I'd say that Switzerland did quite well" or "I don't think things were good for Portugal."

The third chart progressed much like the second, often eliciting a second visceral response equal to the first. Two subjects did not complete the task for the third variant. For the remaining 4/6 there was not a significant difference in time.

Most subjects felt compelled to compare and contrast the two techniques. Opinion was divided:

- The font sizes were slightly smaller for the text as lines variant (5 point) compared to the labels on path (7 point). For the oldest participant, the slightly smaller size was borderline legible and he had some difficulty reading the labels whereas he could read the slightly larger path labels. He preferred the larger labels but speculated that he might prefer the smaller labels if size varied depending on the space available.
- One participant found the microtext as line variant extremely compelling and referred to it as a very clean layout. He specifically noticed and called out the font variation (i.e. weight, case, typeface) as an effective means for creating differentiation.
- One subject had a strong preference to the path labels. The approach represented the best readability for the entire chart as the labels tend to push out to the less dense regions on the chart while the most congested areas of the chart retained high legibility by having only thin, accurate lines through these areas. He wanted labels at the end removed to make the chart even cleaner. The microtext as line approach was not preferred as the labels were less legible when overlapping and multi-lingual labels were a distraction to the core content.
- One noticed the multilingual labels and hypothesized that the approach could make the charts more broadly accessible to his global audience.
- One participant hypothesized that the labels on a path may be appealing because of its familiarity from cartography.
- The text as line was identified by one viewer as more stimulating than the other two: any local anomaly or pattern could draw you into the chart by having a meaningful label immediately visible thereby engaging the viewer to spend more time with the chart.
- Overall, 3/6 preferred the microtext only approach, 2/6 preferred river labels and 1/6 did not express a preference.

The final question provided responses that ranged across user types, user locations, and data types. Users identified various timeseries datasets that could benefit from using these techniques such as index analysis, peer analysis, economic analysis and state analysis. There was some discussion as to the nature of the data - the example dataset had low volatility (i.e. the lines did not zig zag up and down) - and how would the approach fare under different datasets. Various enhancements were suggested, such as:

- Labels as lines could indicate other data, such as values at high points and low points, rate of change, or political leader during time periods.
- Bold could be used to indicate additional information part way along a path.
- Interactivity could combine both the benefits of the textual static view with selection for ability to easily focus on any subset.

In addition, one insightful comment was that the application area was much broader than capital markets - understanding trends across peer groups are applicable in many policy areas. In most of these cases, the reports will be published and distributed and not interactive. Any decision maker would benefit from these techniques.

5.3. Discussion

The application area of timeseries analysis with many lines seems like a rich area for further investigation, particularly given a much broader application area outside of financial markets. However, a broader evaluation should be done for this broader audience - unlike financial market chart users, this broader community may have less familiarity with charts and visualization literacy.

Minimal font size is also an area for additional investigation. While one older subject found 5 point text borderline legible another subject had undergone cataract surgery one year earlier and had no difficultly reading the small 5 point text. Lighting, glare, paper quality and other factors can also affect legibility and this was not controlled. Cartographic guidelines imply smaller sizes could be used and thus data densities increased.

Experts suggested that lines could contain more information than simple labels. Figure 13 is one possibility of a line chart with more data. This figure plots the number of retweets over time for some popular Donald Trump tweets originating in late August 2015. Lines show the actual content of the tweets, including the author of the tweet in bold, the tweet content in plain text, and the lines differentiated by color.

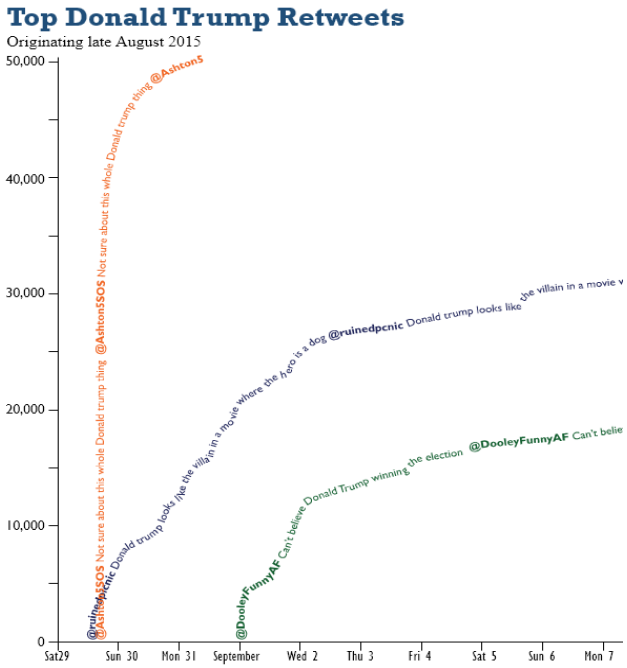

Figure 13 Number of retweets over time for popular tweets about Donald Trump in mid-2015.

6. Conclusion

Microtext is a promising extension to information visualization. Microtext goes beyond previous type-based approaches by:

1) using text smaller than most visualization labels, but staying large enough to remain readable without interaction;

2) uses lines to embed additional information beyond simply labeling the area or line, e.g. items counted, multilingual labels, or text content of the item measured.

3) the approaches can be combined with other visualization approaches (e.g. color)

4) the approach can be extended to use font attributes to differentiate between categories (e.g. bold).

Furthermore, the approach is highly applicable to the analysis of many timeseries, particularly in environments with no or low interaction, which may be very broad.

References

[AMJ12] S. Afzal, R. Maciejewski, Y. Jang, N. Elmqvist, & D. Ebert. "Spatial text visualization using automatic typographic maps." *IEEE Trans. on Visualization & Computer Graphics*, (12). 2012.

[Apo1918] Guillaume Apollinaire, « Visée », in Calligrammes. Poèmes de la paix et de la guerre (1913-1916), Paris, Mercure de France, 1918. https://archive.org/details/calligrammespo00apol (pg. 88).

[BCK14] R. Brath, L. Comrie, D. Keller, E. Knuth, and E. Sorenson. "Challenges in Financial Visualization Panel." at *VisWeek 2014*. Paris. http://bit.ly/2lFIm4H

[Bri14] W.C. Brinton, *Graphic Methods for Presenting Facts*. The Engineering Magazine Company. New York. 1914

[Bri39] W.C. Brinton, *Graphic Presentation*. Brinton Associates. New York. 1939.

[Bru1524] Bruno Carthusianus. *Brunonis Carthusianorum Patriarche sanctissimi, theologi Parisiensis Scholae doctissimi: & Remensis ecclesiae canonici moratissimi: Opera & vita post indicem serie literaria indicanda. Venundatur* Jodoco Badio Ascensio, Paris, 1524 https://archive.org/details/bub_gb_3FEGaBmWN0UC (page 999). Accessed Aug 24, 2016.

[Cle94] W. Cleveland, *The Elements of Graphing Data*. Hobart Press, Summit NJ, 1994.

[Few06] S. Few. *Beautiful Evidence: A Journey through the Mind of Edward Tufte*. 2006. www.perceptualedge.com/articles/b-eye/beautiful_evidence.pdf.

[Fry00] B. Fry. *Tendril*. 2000. benfry.com/tendril/index.html

[Hod03] E.R. Hodges: *The Guild Handbook of Scientific Illustration*. John Wiley, 2003.

[LS87] J. H. Larkin and H. A. Simon, "Why a diagram is (sometimes) worth ten thousand words," in *Cognitive Science* 11, no. 1 (Wiley: 1987): 65–100.

[NHC12] M. Nacenta, U. Hinrichs & S. Carpendale. "FatFonts: combining the symbolic and visual aspects of numbers." In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM. 2012.

[Pal07] B. Paley. "Relationships Among Scientific Paradigms". *SEED Magazine*. 2007.

[RMM95] A. Robinson, J. Morrison, P. Muehrcke, A. Kimerling, S. Guptil: *Elements of Cartography*. Wiley, New York, NY, 1995. 3, 4.

[SF04] M.A. Smith and D. Fisher. 2004. Treemap View of 2004 Usenet Returnees. http://www.scimaps.org/detailMap/index/treemap_ view_of_2004_57. Accessed 12//18/ 2015.

[Sku04] Skupin, A. “The World of Geography: Visualizing a Knowledge Domain with Cartographic Means.” *PNAS 101* (Suppl. 1): 5274-5278. 2004.

[Sti1881] A. Stieler. *Stielers Hand Atlas of Modern Geography*. Justus Perthes, Gotha. 1881. www.davidrumsey.com/luna/servlet/s/425okp

[Unk1300] Author Unknown. Book of Hours (Sarum Use). Early 1300s. East Anglia. http://cudl.lib.cam.ac.uk/view/MS-DD-00004-00017/16 Accessed Aug 23, 2016.

[War13] C. Ware. *Information Visualization: Perception for Design*. Elsevier. 2013.

[WWP96] A. Wallgren, B. Wallgren, R. Persson, U. Jorner, J. Haaland. *Graphing Statistics & Data: Creating Better Charts.* Sage Publications. 1996.