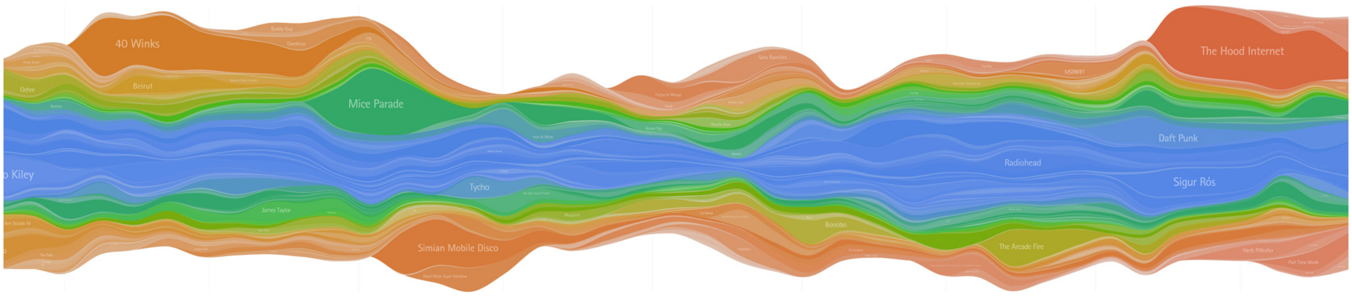
Stacked Graphs – Geometry & Aesthetics

Lee Byron & Martin Wattenberg

Abstract — In February 2008, the New York Times published an unusual chart of box office revenues for 7500 movies over 21 years. The chart was based on a similar visualization, developed by the first author, that displayed trends in music listening. This paper describes the design decisions and algorithms behind these graphics, and discusses the reaction on the Web. We suggest that this type of complex layered graph is effective for displaying large data sets to a mass audience. We provide a mathematical analysis of how this layered graph relates to traditional stacked graphs and to techniques such as ThemeRiver, showing how each method is optimizing a different "energy function". Finally, we discuss techniques for coloring and ordering the layers of such graphs. Throughout the paper, we emphasize the interplay between considerations of aesthetics and legibility.

Index Terms — Streamgraph, ThemeRiver, listening history, last.fm, aesthetics, communication-minded visualization, time series.



1 Introduction

In February 2008, The New York Times stirred up a debate. The famous newspaper is no stranger to controversy, but this time the issue was not political bias or anonymous sources—it was an unusual graph of movie ticket sales. On information design blogs, opinions of the chart ranged from "fantastic" to "unsavory." Meanwhile, on other online forums and blogs, hundreds of people posted insights and questions spurred by the visualization.

The story of the design process and algorithms behind this engaging (and polarizing) graphic makes an illuminating case study in the role of aesthetics in visualization design. Our goal in this paper is to tell this story, while documenting and analyzing the specific geometric algorithms used in creating the visualization. We believe that both the design process and the algorithms may be of use in other contexts.

The visualization method behind the Times graphic was originally developed by the first author to visualize trends in personal music listening. Data for that visualization came from last.fm [11], a social music service that tracks the listening histories of its members. These histories, one time series per artist representing the number of "listens" per week, were shown on last.fm only via bar charts of the activity over the last week and overall top artist rankings.

Since this data was of obvious personal significance, finding a better way to display it was a natural challenge. One conventional method is a stacked graph, with each layer representing an artist's time series. For histories with a large number of artists, however, legibility of the individual layers became a problem. Equally troublesome, however, was the sense that this type of graph was too "statistical" and did not visually embody the rich emotional connection that listeners have with their music.

To solve these problems, the first author created a new form of stacked graph, called a Streamgraph (*see frontispiece*). A Streamgraph layout emphasizes legibility of individual layers, arranging the layers in a distinctively organic form. Applied to last.fm data as part of an academic project called Listening Histories, the Streamgraph design received strong popular response online by both information visualization enthusiasts and music lovers. It eventually drew attention by the New York Times, where it was used to create a printed graphic and accompanying online interactive visualization of the box office revenue for 7500 movies over a 21-year period.

In this paper we first provide a case study of the New York Times and last.fm visualizations. We pay special attention to the response on the web and the role of aesthetics in the appeal of visualizations. Second, we perform a detailed analysis of the algorithms that define these graphs. A key theme is the role of aesthetics in visualization design, and the process and trade-offs necessary to create engaging information graphics.

2 Related Work

Visualizations of multiple time series date back centuries. Scholars have long recognized that despite their simplicity, time series graphs involve many subtle tradeoffs. Bertin [1] and Cleveland [5] both noted that the aspect ratio of a graph has a significant effect on readability of slopes. Bertin also pointed out that for seeing shapes at different levels of detail, different aspect ratios might be optimal. Heer and Agrawala [9] introduced the "multi-scale banking" technique for automatically handling these compromises. In some systems interactivity has been a concern, with tools such as TimeSearcher [10] introducing elegant ways to filter many series. While not directly applicable to stacked graphs, these efforts to handle conflicting criteria for different levels of detail presage much of our work.

Graphs that show time series by using stacked layers date back at least to Playfair's work [15]. Only recently, however, have versions been created that can scale to larger number of time series. The inspirational ThemeRiver system of Havre et al [17] may be the first advance to exploit computation to enhance to power of stacked graphs. In this system, the layers represented the frequency of occurrence of certain terms or "themes" in a historical news feed.

Among the innovations in ThemeRiver were a novel technique for creating a smooth interpolation from discrete data, and a layout method in which layers were not stacked starting on the x-axis, but rather in a symmetrical shape with the x-axis at the center.

In [19] the second author introduced a highly interactive layered graph, the NameVoyager, which enabled rapid exploration of more than 6,000 data sets at once. While the layout method of the NameVoyager was not novel—it used a standard stacked graph layout, with some level-of-detail calculations—the popular response to the applet suggested that stacked graphs have the ability to engage mass audiences.

A follow-up design to the NameVoyager, described in [20], showed hierarchical time series. That is, it used interactivity and color to display time series that were arranged into categories and subcategories. In the Many Eyes system [17], this technique was made broadly available on the web.

A final related work is the Revisionist [7] visualization of changes in source code over time. While not technically a stacked graph, the geometry is related since each line of code is represented by a curved stripe. Revisionist minimizes visual distortion by having a curved baseline that allows the visualization to roughly align identical lines of code between releases.

3 Last.fm and The New York Times

3.1 Listening History - Last.fm

Listening History was created by the first author for a class project at Carnegie Mellon University. The six-week assignment was to collect and display a data set in an interesting and novel way. As described in the introduction, Listening History [4] visualizes trends in an individual's music listening, as derived from data in the last.fm service. The x-axis represents time and each stripe represents an artist. The thickness of a stripe shows the number of times that songs from the artist were listened to in a given week. The color, as detailed in section 5, encodes two dimensions: the saturation is determined by the overall number of times an artist is listened, and the hue is related to the earliest date at which one of the artist's songs were heard.

A critical design goal for this visualization was to create a graphic that did not look scientific or mathematical, but rather felt organic and emotionally pleasing. In section 5 we will see that, ironically, achieving this goal relied on significant computation. A side effect of the algorithm is the signature asymmetry between the top and bottom curves which form the organic shape and, as discussed later, minimizes internal distortion.

At the end of the course, a few large-scale posters, some over 12 feet long, were printed as holiday gifts. The reaction of the recipients provides evidence, if anecdotal, that the graphic succeeded in eliciting strong emotional reactions when people saw their own listening history. People often remarked at the ability to see critical life events reflected in their music listening habits.

One pointed to the beginning and end of three separate relationships, and how his listening trends changed dramatically. Another noted the moment when her dog had died, and the resulting impact on the next month of listening. A third pointed out his dramatic differences between summer and winter listening trends. As in the Themail system of Viégas *et al*. [18], the visualization of historical and personal data seemed effective at eliciting reflective storytelling.

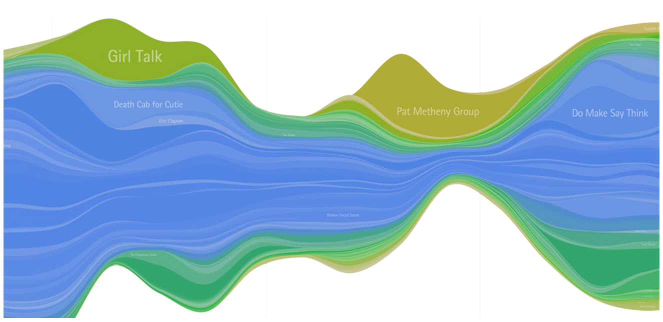


fig 1 – section from Listening History of primary author

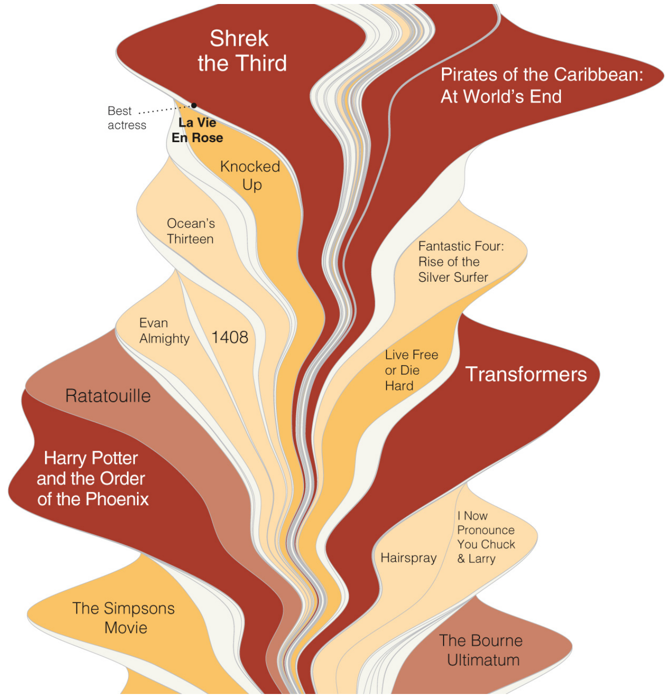


fig 2 – films from the summer of 2007

After Listening History was made public, there was high demand for personalized versions of these graphics by other last.fm members. In fact this demand was so strong that a number of imitators emerged, including Maya's Extra Stats [12] and Godwin's Last Graph [13] Interestingly, these services and other imitators use the simpler ThemeRiver layout and a simpler color scheme.

The popularity of these imitators (Last Graph has created visualizations for more than 24,000 users) suggests the hypothesis that stacked graphs have an ability to communicate large amounts of data to the general public in an intriguing and satisfactory way.

3.2 New York Times - Box Office Revenue

The Box Office Revenue graph, created by the first author and the graphics department of the Times [2,6] highlighted the dichotomy between box office hits and Oscar nominations, discussed in the original article. The printed graphic ran vertically to best use the available space, time running top to bottom; the online version ran left to right. To allow a quick reading of the graph, coloring was much simpler than in Listening History: a discrete palette signified ranges of overall revenue. Furthermore, stroke lines were added because of issues with print registration.

The online response to these graphics was intense and rapid. Many blogs and social websites featured long lists of comments discussing data-points shown in the graph. As with the NameVoyager, blog posters and their commenters engaged in a social style of data analysis and critique of the new visual form. What follows are anecdotes discussing these visualizations, which provide a rough sense of the breadth and intensity of the online response.

Individual bloggers often found particular discoveries and pointed them out to their readers. For example, one said:

C1: *note the double spike on 'Harry Potter an the Order of the Phoenix'. And the long hump on 'Alvin and the Chipmunks'. 'Juno' also has an interesting curve as it did almost nothing for a month before popping out later in it's run. Though that may be because it was released in just enough theaters to become Oscars* *eligible before going into wide release.*

Others looked at broader trends.

C2: *As time goes on, movies open bigger but don't last nearly as long in the theater as they used to. There are also more movies to choose from in 2007 than in 1986.*

One of the most extensive analytical conversations took place on the popular community news site Digg, where the discussion around the graph consisted of 156 comments at the time of writing. Here the talking points were briefer. Many times people posed questions:

C3: *Why would Top Gun have such a long tail?*

C4: *You'd think they'd figure out to take advantage of the lull in March and April.*

and in some cases people pointed out implications for other issues:

C5: *Looks like they are doing just fine. F**k the MPAA.*

In addition to comments on the data, many people expressed opinions on the technique itself. Negative comments often expressed confusion about the scale:

C6: *It's a little confusing at first because the vertical scale is basically irrelevant*

C7: *Uselessly complex, endlessly confusing, and the stuff below and above...what?*

On the other hand, many commenters were extremely enthusiastic, calling the graph "amazing," "fantastic," and "brilliant." More detailed positive comments seemed to focus on the richness of the data:

C8: *As an avid fan of Tufte's Visual Display of Quantitative Information… I relish content-rich graphics such as the one you cite.*

C9: *The graph is SO well done. Really good data and easy to read.*

C10: *This has to be one of the most intuitive data visualizations I've seen in a long time.*

Finally, many readers found the shape and colors of the graph suggestive. As the popular media blog Gawker put it:

C11: *New York Times Makes Box Office Analysis Seem Cool, Vaguely Erotic*

Other comments echoed this theme. One of the printable ones came from a prominent academic who stepped into the blog space to speak of its suggestions:

C12: *This was a great graphic, visually appealing and even sensuous (in a Georgia O'Keefe way).*

Others noted different connotations; some viewers liked these, some didn't, and some felt both ways at once:

C13: *anyone else find that chart either disgustingly gross or strangely delicious? it's either the trail left behind a big brown slug, or someone spilled a bucket of dulce de leche pudding. awesome chart none the less ;)*

A rigorous content analysis of these comments is beyond the scope of this paper. The various types of comments, however, suggest hypotheses for future study. Comments C1-C5 indicate there might have occurred the same kind of social data analysis seen around the NameVoyager. Opinion on the legibility of the graph was mixed. Comments C6 and C7 indicate the asymmetric top and bottom curves caused some people trouble. Despite this, many people claimed to read overall trends: C2, C4, and C5. At the same time C1, C2, C3 suggest that viewers are able to extract details of individual movie sales effectively; C8-C10 seem to refer to this fact implicitly. As discussed below, these comments indicate there may be subtle tradeoffs in readability of various aspects of the graph. Finally, C11- C13, point to the idea that the general look of the graph effectively caught and held many people's attention.

4 Considerations in Stacked Graph Design

The popular reactions to the NameVoyager, Listening History and its imitators, and the Box Office Revenue graph suggest that this type of visualization is capable of conveying a large amount of data in a manner that engages mass audiences. The ThemeRiver system's success indicates that these graphs are helpful for expert analysts as well. At the same time, all these systems have subtly different variations on the techniques for defining the geometry, layout, color, and interaction of the graph. We believe it is worthwhile to produce a unified treatment of the issues involved with these graphs, partly as a guide for designers, and partly as a way of pointing out some unresolved issues that are ripe for future research.

The design considerations for stacked graphs fall into two categories. First, as with any information graphic, legibility of the data is critical. Indeed, one of the main polarizing aspects of the visualization related to its overall legibility. Second, as the reactions to the Listening History and New York Times visualizations show, aesthetics seem to play an important role in the popularity of this type of graphic.

We present these perceived design issues so we may later refer to them as rationality for making particular design decsions. To ease reference, we have marked each issue with a letter, i.e. (A).

4.1 Legibility

The main idea behind a stacked graph follows Tufte's macro/micro principle [16]: the twin goals are to show many individual time series, while also conveying their sum. Since the heights of the individual layers add up to the height of the overall graph, it is possible to satisfy both goals at once. At the same time, this involves certain trade-offs. There can be no spaces between the layers, since this would distort their sum. As a consequence of having no spaces between layers, changes in a middle layer will necessarily cause wiggles in all other surrounding layers, wiggles which have nothing to do with the underlying data of those affected time series (A).

Reading and comparing the thicknesses of the various layers can be problematic for this reason, but for other reasons as well. Two layers of the same vertical height but with different slopes may appear to have radically different thicknesses (B). A related issue is related to Cleveland's principle of banking to 45 degrees [5]: with typical data, there may be a trade-off between having individual layers being too "flat" versus the overall graph being too "spiky" (C) (*fig 3*).

A second issue for both individual layers and the overall graph shape is the difficulty in comparing the vertical thickness of two curves with different slopes. This is why, of course, traditional

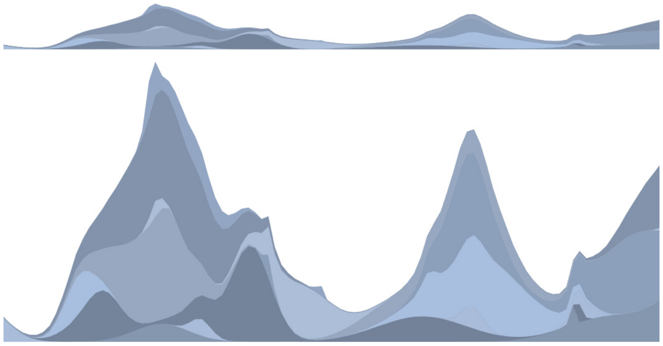


fig 3 – a comparison of angles of layers too "flat" and too "spiky"

statistical stacked graphs have their bottom at the x-axis: it makes the overall height at each point easy to estimate. The trade-off in the case of traditional stacked graphs with many layers, however, is that this may cause the individual layers to be harder to read (D).

A third issue is the ability of a reader to distinguish effectively the many layers of a stacked graph (E). In several stacked graphs with thousands of time series (the NameVoyager, the Listening History graph) color serves a twofold purpose: to convey a particular dimension of data and to distinguish layers without using heavy borders. In the case of the NameVoyager, stripes are colored according to the gender of a name and the most recent value of the time series. The Listening History graphic uses the time of onset and relative personal popularity of musician. In the Many Eyes implementation of a "stack graph with categories," the colors are chosen to convey a sense of the overall hierarchy.

4.2 Aesthetics

Legibility is not the only consideration, however. Just as important in the case of stacked graphs may be their aesthetic quality. Many of the comments on the Box Office Revenue graphic support the idea that the visual appearance of the graph drew people in or kept them looking at the graphic (F).

We speculate that some of the aesthetically pleasing—or at least engaging—qualities may be in conflict with the need for legibility. The fact that the New York Times graph does not look like a standard statistical graphic may well be part of its appeal. If this is true, it is unclear how much weight to put on this fact in creating designs. If the graphic becomes commonly used, could its appeal actually diminish? On the other hand, if the appeal is more enduring, perhaps this is a good example of how a pleasing look may entice readers to dig deep into a set of data.

The relative priorities of aesthetic and utilitarian considerations in a visualization clearly depend on context. In a situation with a fixed captive audience, there may be no need to compromise legibility in order to get the attention of a viewer. In other situations, it may be worth prioritizing aesthetics to broade the appeal of a graphic. Through example, we display specific decisions based on this tradeoff. Exploring this balance, and studying when and how to compromise, may be an important area for future research.

5 Algorithms for Stacked Graph Design

There are four main ingredients that determine a generalized stacked graph. The shape of the overall silhouette is the first ingredient; this shape is critical since it determines the overall slopes and curvature of the individual layers. The second important parameter is the ordering of the layers, which may be chosen to conform to different aesthetic criteria. As with any visualization, labels are important as well—but the organic forms of a Streamgraph mean that labels require additional attention. Finally, color choice is critical, enabling viewers to distinguish different layers and potentially conveying additional data dimensions. In this section we describe algorithms that address each of these four ingredients with respect to the design issues of leigbility and aesthetics.

5.1 A Unified Approach to Stacked Graph Geometry

The geometry of a stacked graph consists of a set of layers, corresponding to the time series. To conform to the "macro/micro" principle there can be no space between layers, so that the thickness of the overall stack reflects the sum of the individual time series. Given this constraint, the overall geometry of the stacked graph is determined by two factors: the shape of the "baseline," or bottom of the lowest layer, and the order of the layers. In this section we discuss the effect of the baseline on the overall geometry of the graph, and in the next section discuss layer ordering.

To describe the geometry precisely, we use the following notation. We model our time series as a set of n real-valued non-negative functions, $f_1, \ldots, f_n$. In what follows, for simplicity, we will assume these are differentiable and defined on the interval [0,1]. One might also consider functions taking values at a discrete set of points, but the notation becomes more cumbersome and in any case it is easy to move from the discrete case to the differentiable case through interpolation.

We refer to the baseline function that defines the bottom of the stacked graph as g_0. The top of the layer corresponding to the i^{th} time series f_i is therefore given by the function g_i, where

$$g_i = g_0 + \sum_{j=1}^{i} f_j$$

Which is illustrated for these definitions for n = 2 by *fig 4*.

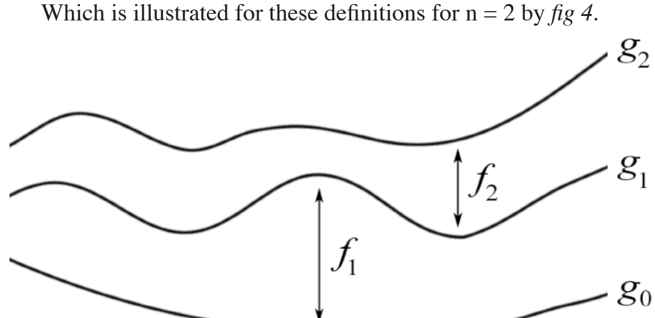


fig 4 – a visual description of stacked graph functions f_i and g_i for n=2 as used in this section

How should the baseline function g_0 be chosen? There are a variety of possible criteria. The simplest is the traditional stacked graph, which has

$$g_0 = 0$$

This has the effect of making the graph of the sum of all the series into a traditional graph, based at zero (*fig 5*). (In this and the following figures, we use a synthetic data set with randomly assigned

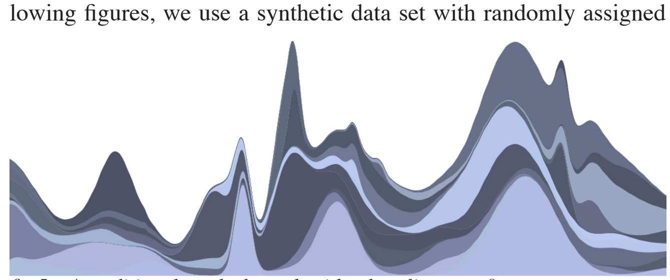


fig 5 – A traditional stacked graph with a baseline g_0 = 0 colors to distinguish layers.) In this layout the size of the sum of the series is easy to read, potentially at the expense of the legibility of the individual layers as mentioned in design issue (D).

A simple alternative layout was suggested by Havre et al in the ThemeRiver system. They used a layout symmetric around the x-axis. Mathematically, this can be expressed as:

$$g_0 + g_n = 0$$

or, from the definition of g_n,

$$2g_0 + \sum_{i=1}^{n} f_i = 0$$

which yields the ThemeRiver (*fig 6*) solution for g_0:

$$g_0 = -\frac{1}{2}\sum_{i=1}^{n} f_i$$

This is a simple enough definition, but there is another way to look at this formula which suggests some generalizations. What is special about the symmetric layout? Aside from a certain aesthetic quality, this layout has the effect of minimizing some important quantities. In particular, at each point, the silhouette is as close as possible to the x-axis, and in addition the slopes of the top and bottom of the silhouette are in a sense as small as possible (in the sense of total sums of squares). This directly addresses design issue (C) by making the overall graph much less "spiky" thus greatly reducing the horizontal space needed to satisfy Cleveland's principal. To see this, recall the following fact about averages. For any set of real numbers $\{a_1,\ldots, a_n\}$, the value of x that minimizes

$$\sum_{i=1}^{n} (x + a_i)^2$$

is

$$x = -\frac{1}{n}\sum_{i=1}^{n} a_i$$

From this fact it follows that the value of g_0 that gives the symmetric ThemeRiver layout minimizes the sum of squares of the top and bottom of the silhouette of the graph (at each point in [0,1]):

$$silhouette(g_0) = g_0^2 + g_n^2$$

since

$$g_0^2 + g_n^2 = g_0^2 + (g_0 + \sum_{i=1}^{n} f_i)^2$$

A similar calculation shows that the ThemeRiver layout also minimizes the sum of squares of the slopes of g_0 and g_n at each point.

Seen in this light, the ThemeRiver layout does not just produce a pretty symmetry, but is optimal in the sense of minimizing certain mathematical measures of distortion, a hint at addressing minimizing wiggles, design issue (A).

Since a stacked graph depends on the readability of the individual layers, as noted in design issue (D), it is natural to ask about applying the same optimization criteria to each layer overall. For example, one might ask for a layout that reduced the overall sum of squares of the distance from the x-axis of all layer edges. In other words, we might like to minimize a *deviation* measure at each value of x, defined by:

$$deviation(g_0) = \sum_{i=0}^{n} g_i^2 = \sum_{i=0}^{n} (g_0 + \sum_{j=1}^{i} f_j)^2$$

(Note: In this and future formulas, we adopt a special, convenient subscript notation. When a summation's top subscript is less than the bottom, as in the case *i=0*, we take the sum to be empty and equal to zero.)

We might also be interested in minimizing the sum of squares of the slopes at each value of x:

$$wiggle(g_0) = \sum_{i=0}^{n} g_i'^2 = \sum_{i=0}^{n} (g_0' + \sum_{j=1}^{i} f_j')^2$$

By the same logic, the deviation quantity is minimized when

$$g_0 = -\frac{1}{n+1}\sum_{i=0}^{n}\sum_{j=1}^{i} f_j = -\frac{1}{n+1}\sum_{i=1}^{n} (n - i + 1) f_i$$

Moreover, differentiating this yields:

$$g_0' = -\frac{1}{n+1}\sum_{i=0}^{n}\sum_{j=1}^{i} f_j'$$

which minimizes the *wiggle* measure as well. In other words, this choice of g_0 has the effect of simultaneously minimizing both distance from the x-axis and variation in slope. Moreover, the formula is extremely efficient to calculate. As *fig 7* shows, it creates a more "even" layout compared to the ThemeRiver graph. This minimization directly addresses design issue (A) by attempting to reduce the effect of middle "wiggles" on the surrounding layers.

Nonetheless, the layout can be improved further. In particular, it is potentially problematic that very thin layers receive the same treatment as very thick ones. After all, the thick layers are visually more important. Thus we might want instead to optimize the following quantity, which represents an average of the squares of slopes between the midpoints of each layer, weighted by layer thickness:

$$weighted\_wiggle(g_0) =$$
$$\sum_{i=1}^{n} (\frac{1}{2}(g_i' + g_{i-1}'))^2 f_i = \sum_{i=1}^{n} (g_0' + \frac{1}{2} f_i' + \sum_{j=1}^{i-1} f_j')^2 f_i$$

By the properties of weighted averages, this is minimized at each point in [0,1] when

$$g_0' = -\frac{1}{\sum f_i}\sum_{i=0}^{n} (\frac{1}{2} f_i' + \sum_{j=1}^{i-1} f_j') f_i$$

which can be integrated numerically to yield a solution for g_0. Indeed, it turns out to be equivalent to the algorithm used in the Streamgraph method, as portrayed in Listening History and Box Office Revenue graph. While this formula, which minimizes *weighted_wiggle*, is not as computationally efficient as the formula for optimizing wiggle, it seems to have better visual properties (*fig 8*). Focusing on minimizing wiggle per layer in this way attempts to solve design issue (B) by making the height of a layer at any point easier to read at the cost of causing the silhouette of the whole graph to be slightly more difficult to read in this same way. In the appropriate context, such as the examples given in sections 2 and 3, this layout method also approaches design issue (F) by creating a subjectively beautiful aesthetic form.

We believe that framing layout choices in terms of the quantities they minimize provides a useful way to organize the design space of stacked graphs. In a sense, the various distortion functions serve the same role as conventional "aesthetic criteria" in graph drawing, namely a translation of design judgment into quantitative form. Furthermore, this method points to natural extensions: it is easy to imagine "distortion functions" besides *silhouette*, *deviation*, *wiggle*, and *weighted_wiggle*. One could also write formulas that optimized

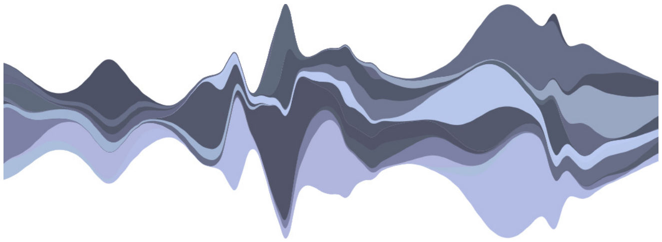


fig 6 – the same data set using the ThemeRiver layout algorithm

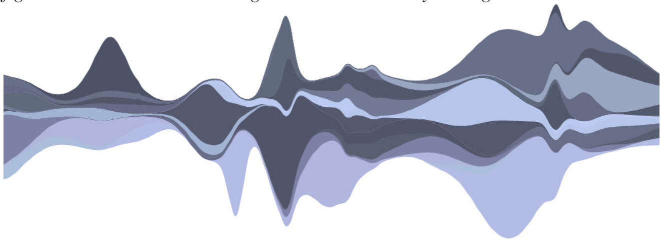


fig 7 – the same data set optimized to reduce the "wiggle" function, or overall variation in slope

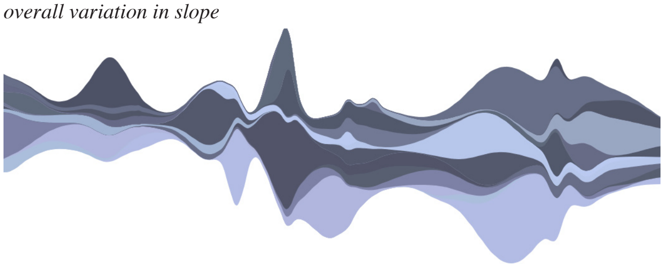


fig 8 – the same data set optimized to reduce the "weighted_wiggle," the algorithm used in Streamgraph

a weighted average of several criteria. Of course, more complicated distortion functions may not admit the simple point wise optimizations above, possibly requiring results from the calculus of variations or numerical optimization.

5.2 Color Choice

Coloring stack graphs with many layers is challenging. Although color is an opportunity to communicate additional data, strong or jarring colors become visually distracting and make the graphic hard to read. At the same time, there must be enough local contrast between layers in order to differentiate each layer, a particularly important issue as raised in design issue (E). The design problem is complicated by the need to balance these perceptual considerations with aesthetic ones which stem from design issue (F): does the final graphic look good? Are its emotional connotations consistent with the nature of the data?

These complex trade-offs mean that choosing a color scheme is highly dependent on the underlying data as well as the context in which it will be presented. In this section we describe the decisions behind the Listening History and the Box Office Revenue graph. In both cases, the darkness and saturation of the color for a particular time series reflected the sum total of the series; this drew attention to series with larger sums—favorite musicians and blockbuster movies—which tended to be more significant.

In the Listening History graphic, the color of each time series also encoded the time of onset. In this version of the visualization, onset time is shown using a visual gradient from cool colors to warm colors. This significance of onset time for Listening History is due to the particular form of the last.fm data. Musicians may peak once when discovered, but then often experience many later resurgences. Box office films do not experience this same resurgence trend. Due to the short duration of box office films and their lack of resurgence, there is no need for differentiation based on time of onset since it is almost always apparent by placement alone.

For Listening History, the colors chosen to represent the spectrum of old to new, of most significant, or most listened, to least significant construct a two dimensional gradient (*fig 9*). The gradient across the x axis is a detailed movement through hue representing the "cold core" of known musicians versus the "hot new" discoveries of recent musicians, while the gradient down the y axis represents significant to not significant, consistently decreasing in saturation and increasing slightly in brightness.

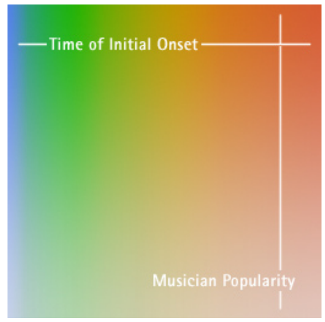


fig 9 – the 2d color palate used in Listening History

The colors used in Listening History are not strictly computer generated, and not a pure transition through hue. Instead they are assembled to be visually expressive from the designer's subject point of view, contextually addressing design issue (F). The colors are chosen from highly saturated images of nature. The blue is from a clear sky, the green from a tree leaf, and the reds, oranges and yellows from images of flame. These colors are then formed into a gradient using Photoshop, giving specific care to the interpolation of color between these core points, compensating for the differences between numerical and perceptual consistency. Notably these colors appear natural and pleasing, and are not over-saturated.

The range of color should not be confused with a round-trip hue "rainbow map" [3]. Firstly, this range represents only half of the available hues, marking a clear difference between extremes of the dataset. Rather, the range of color is chosen to represent a complementary color scheme between old and new layers while also representing analogous color schemes amongst individual layers.

"Core" musicians, which appear early in the data set, tend to comprise much more area than recently discovered musicians. In order to avoid a heavily blue colored graphic, this color gradient is biased towards the warmer colors. This counteracts the common area bias towards earlier onset time series and gives the resulting graphic a balance between warm and cool colors.

The result of this handcrafted expressive palette is a wide range of colors to represent each layer of data which work together in an analogous way to please the eye but are also different enough to create local contrast between layers to ease differentiation as per design issue (E).

Related to these two examples is the color scheme of the second author's Name Voyager [14]. Name Voyager creates a simultaneous complementary color scheme between male and female names and analogous color scheme within each gender in order to create local contrast. This approach is very similar to that taken by Listening History. This coloring scheme is effective in communicating broad information as well as differentiating a large number of time series.

5.3 Layer Labeling

Design issue (E) considers distinguishing individual layers. Solving this, however, is pointless if it is not clear what each individual layer represents. Stacked graphs with a small number of layers do not necessarily need labels for the layers, since a legend and color coding scheme may be appropriate. Unfortunately, such a simple solution is not possible for a graph with hundreds or thousands of time series. A critical aspect of the design of stacked graphs is therefore the placement of layer labels. Ideally a label is visually associated with the data it represents, will not overlap other labels or layers,

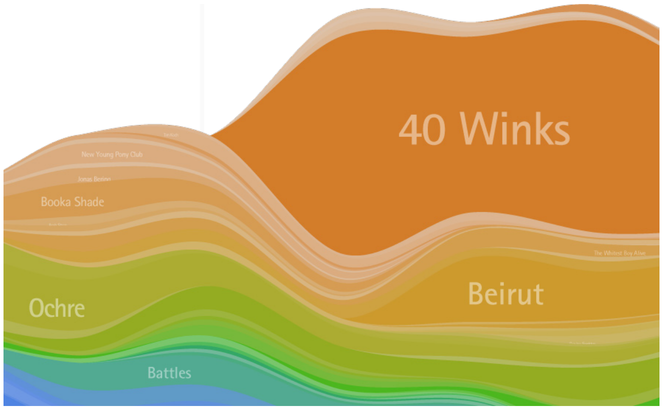


fig 10 – a detailed look at the labeling strategy of Listening History

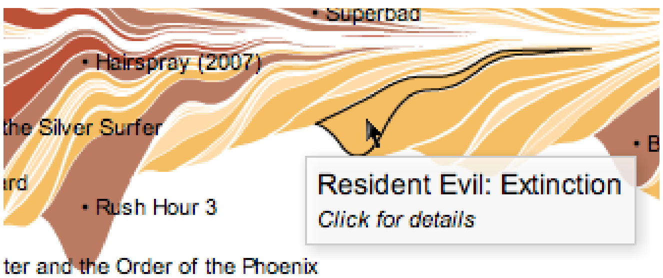


fig 11 – detail of a roll-over label in the New York Times graphic

and will not distract from the rest of the graphic.

Listening History places labels within the layers themselves, rather than using a call-out line, not allowing the label to overlap the boundary of the layer (*fig 10*). The font size of the labels is adjusted to fit each layer, and the labels are placed in an optimal spot along the graph where the font size could be the largest. (A simple brute-force approach to finding this best position had acceptable performance for offline processing.) When printed at a high resolution, even the smallest layers have readable labels. To minimize unnecessary contrast, and to visually connect the labels to the layers, labels are drawn white and slightly transparent to take on a portion of its layer's color. This also creates the higher local contrast on more saturated—significant—layers.

For the graphic printed in the New York Times, the resolution of the printing process was not reliable; to compensate, labels were added by hand. The online interactive piece does not use this proposed label placement strategy because of the poor real-time performance of the brute-force algorithm. Rather it allows for roll-over details for smaller layers (*fig 11*).

5.4 Layer Ordering

A final choice in the design of a stacked graph is the order of the layers. In some cases there may be a particular intrinsic sequence. For example, in the NameVoyager the ordering is alphabetical, since the emphasis of the visualization is on the initial letters of the name.

In other cases, however, the ordering can be chosen to enhance legibility or to make a better-looking graph. approaching the design issues (A-C) and (F). In the remainder of this section we describe how these choices were made for Listening History and the Box Office Revenue graph. These examples illustrate the interplay between aesthetic and communicative concerns (F), as well as how particular statistical characteristics of data sets may affect the geometry (A-C). We end with a general discussion of other ways to use ordering.

A particular type of burstiness characterizes both the Listening History and Box Office Revenue data sets. A typical time series in each set begins at zero—a musician is unknown or a movie not yet released—and remains zero for a while, and then suddenly "bursts" to a maximum—a musician is discovered, a movie released to great fanfare—followed by decay in value—a musician becomes boring, a movie fades from public view. This pattern presents a challenge for the stacked graph layout, since bursts can cause disruptive wiggly artifacts in the geometry (*fig 12*). By the same token, the "onset time" of a time series—i.e., the moment when it is first nonzero—becomes a variable that users may want to see highlighted.

One might consider sorting the data set by "onset time". If the "new" layers are always added along the top, the graph takes on a distracting downward diagonal stripe pattern in addition to an upward angle to the overall silhouette due to the layout algorithm's effort to keep the sum of slopes low (*fig 13*).

To prevent this, layers are given a "inside-out" ordering, in which early-onset time series are placed at the middle, with later-onset series at the top and bottom. This has three benefits in addition to avoiding the diagonal-stripe effect. First, it places the biggest bursts in the layers—the first non-zero value—at the outside the graph, where they will disrupt the layout of other layers the least, drastically improving legibility, design issues (A-C). Second, we speculate that the top and bottom regions of the graph tend to be most prominent areas, since they occur near the high-contrast silhouette. The central "core" of the graph, the middle, may be read secondarily. Since the bursts are the most "interesting" part of the data in many cases, the inside-out layout places them in the potentially prominent position (*fig 14*). Third, it prevents a drift of the layout away from the x-axis, an artifact that can be seen dramatically in *fig 13*.

The particular inside-out ordering is defined as follows. Note that one easy method would be simply to sort the layers by onset time, and then add layers alternately to the beginning and end of a layer list. Unfortunately, this simple method could potentially lead to a highly asymmetric graph if the layers that end up at the beginning of the list represent much larger values than the ones at the end.

To prevent this asymmetry, we use the following algorithm in

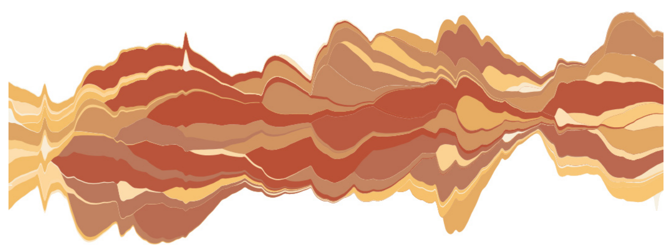


fig 12 – an unsorted data set, exhibiting the type of "burstiness" apparent in last.fm and box office data sets

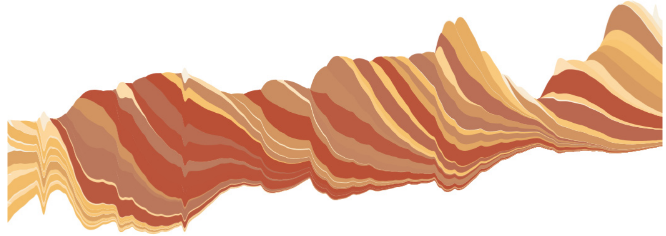


fig 13 – the same data set, naively sorted in order of "onset time" exhibiting the distracting diagonal striping effect

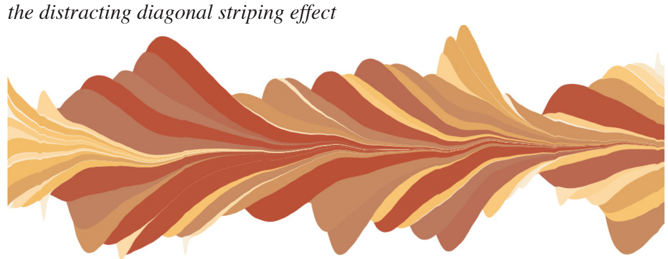


fig 14 – the same data set sorted using the weighted "inside out" strategy to highlight the initial onset of each time series

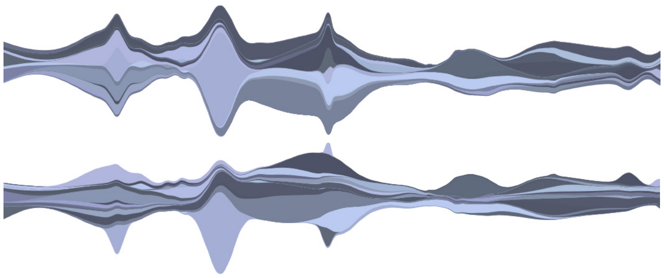


fig 15 – before and after applying an "inside out" sort using a "measure of volatility" in place of "onset time"

ordering the layers. First, we define the "weight" of a time series as the sum of all its values. Then after sorting by onset time, we add time series to the list one by one, attempting to "even out" the weight between the top and bottom half: more precisely, if the sum of the weight of the first half of the current list is greater than half the total weight, we add the series the end; otherwise, we add to the beginning.

For data sets with different statistical properties, other layouts are possible. Indeed, this is a promising area for future research. For example, in order to further improve design issue (A), one might calculate a volatility measure for each time series, which would then be used in place of time of onset for the algorithm described above (*fig 15*). This approach would lend to minimizing the energy function of layer distortion by placing the individual series with the least amount of change in the center of the graph and the series with the most amount of change along the edges. Another possibility would be to find an order that minimized the aesthetic criteria of section 5.1, e.g., by making sure that the "wiggles" of neighboring layers cancelled out to the greatest extent possible.

6 Conclusion

In this paper we have described a new kind of stacked graph, the Streamgraph. We began with a case study of the design of the method and a general description of the popular reaction to two instances of this visualization. Using this description as a jumping-off point, we then turned to a detailed discussion of the compromises involved in specifying a streamgraph. In doing so, we specified a unified approach to defining stacked graph layouts with respect to a set of primary design criteria and provided a mathematical treatment of the various possible geometric algorithms. The unifying theme behind our treatment is that each layout optimizes a certain quantity, such as the overall slope of the layers. By treating each layout method as the solution to an optimization problem, we are able to connect the various layout options to aesthetic criteria.

An important purpose of this paper is to spotlight stacked graphs as an interesting object of study. There are many unresolved questions in their design and evaluation. From a design perspective, one might ask for new algorithms for stacked graph layouts that optimize the aesthetic criteria discussed in this paper. For instance, we only touched briefly above on the question of reordering layers; this may be a fruitful area for future research. A second interesting design question is how best to show hierarchical information. The Many Eyes method of using color to show tree structure has clear limitations, and it would be helpful to find alternative methods.

A second issue for stacked graphs is assessment. Here the issue is as much finding the right questions as answering them. From a traditional point of view, the varying baseline of the Streamgraph layout should make the overall graph much harder to read. One can ask: does it indeed make it more difficult to read? And if the answer is yes (as may be expected), then the real question may be: does this matter? For instance, are there gains in legibility of individual layers that can be shown to outweigh the problems in reading the overall shape? More radically, one might ask whether the engaging nature of a more fluid view actually outweighs the loss of legibility, and how the context of delivery affects the answers to all these questions. We believe these questions are interesting in themselves, but also cut to the heart of issues with visualization as it is used in the mass media.

Acknowledgements

The authors wish to thank Ben Fry for his support and guidance during the Listening History project. We also thank Amanda Cox, Mathew Bloch, and Shan Carter from the New York Times for their insight and efforts in the creation of the Box Office Revenue visualizations. We would also like to thank Last.fm for their support of the project and the many individuals in the blogging world for all of their collective constructive feedback and interest in both projects. Finally, we thank Fernanda Viégas and the anonymous referees for helpful comments on this paper.

References

[1] Bertin, J. (1983) Semiology of Graphics (translated by A. Berg). University of Wisconsin Press.

[2] Bloch, M., Byron, L. Carter S., and Cox, A., (2008) The Ebb and Flow of Movies: Box Office Receipts 1986-2007. http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html

[3] Borland, D and Taylor, R.M. (2007) Rainbow Map (Still) Considered Harmful. *IEEE Comp. Graphics and Applications*. 27(2).

[4] Byron, L. *Listening History*. http://megamu.com/lastfm/.

[5] Cleveland, W. (1993) *Visualizing Information*. Hobart Press.

[6] Cox, A. and Byron, L., (2008) "The Ebb and Flow of Box Office Sales, 1986-2007," *The New York Times*, February 23, 2008.

[7] Fry, B. (2003) Revisionist. http://acg.media.mit.edu/people/fry/revisionist/

[8] Havre, S., Hetzler, B., Nowell, L. (2000) ThemeRiver: Visualizing Theme Changes over Time. *Proceedings of the IEEE Symposium on Information Visualization*.

[9] Heer, J. and Agrawala, M. (2006) Multi-Scale Banking to 45 Degrees. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(5). Sep/Oct 2006.

[10] Hochheiser, H., Shneiderman, B., (2004) Dynamic Query Tools for Time Series Data Sets, Timebox Widgets for Interactive Exploration, *Information Visualization 3, 1*.

[11] Last.fm, http://www.last.fm

[12] Last.fm Extra Stats, http://www.lastfm.de/user/C26000/journal/2006/07/30/195693/

[13] *Last Graph*, http://lastgraph.aeracode.org/

[14] *NameVoyager* (2005) http://babynamewizard.com/namevoyager/

[15] Playfair, W. (1786) *Commercial and Political Atlas and Statistical Breviary*.

[16] Tufte, E. (1986) *The Visual Display of Quantitative Information*. Graphics Press.

[17] Viégas, F.B., Wattenberg, M., van Ham, F., Kriss, J., & McKeon, M. *Many Eyes: A Site for Visualization at Internet Scale*. In Proc. of IEEE InfoVis 2007.

[18] Viégas, F.B., Golder, S., and Donath, J. Visualizing Email Content: Portraying Relationships from Conversational Histories, *Proceedings of ACM CHI, 2006*.

[19] Wattenberg, M. (2005) Baby Names, Visualization, and Social Data Analysis. *Proceedings of the IEEE Symposium on Information Visualization*.

[20] Wattenberg, M. and Kriss, J. (2006) Designing for Social Data Analysis. *IEEE Trans. Vis. And Comp. Graphics*. 12(4).