Eurographics Conference on Visualization (EuroVis) 2013
B. Preim, P. Rheingans, and H. Theisel
(Guest Editors)

Volume 32 (2013), Number 3

Gestaltlines

Ulrik Brandes[1] Bobo Nick[1] Brigitte Rockstroh[2] Astrid Steffen[2]

[1]Department of Computer & Information Science, University of Konstanz
[2]Department of Psychology, University of Konstanz

Abstract

We propose a general technique to visualize multivariate data sequences. It is based on a symbiotic combination of three powerful concepts from information visualization: sparklines, glyphs and gestalt theory. By visualizing several well-known data sets in new ways we first demonstrate how explicit consideration of gestalt principles can be used to leverage visual perception capabilities for the identification of patterns such as trends, periodicities, change points, or outliers. A more detailed case study with complex and noisy data from a psychological experiment then demonstrates how basic design ideas for gestaltlines can be applied in less controlled, and thus more realistic, situations. The case study is complemented with reports on feedback from domain experts and a user study, both indicating that gestaltlines can be a convenient and valid means to explore and communicate patterns in micro-visualizations.

Categories and Subject Descriptors (according to ACM CCS): H.5.m [Information Systems]: Information Interfaces and Presentation—Miscellaneous

1. Introduction

We propose a design concept for graphical representations of sequences of multivariate data, and provide an initial exploration of the idea with examples from the literature, a case study, and its empirical validation.

Our general motivation is the use of visualization to facilitate exploration and communication of patterns in multivariate data sequences. Patterns of interest typically include intervals of relative stability, periodicity, trends, transitions, outliers, and, in the case of sequence collections, shared, related, and discriminating features. These are often difficult to describe quantitatively, even in hindsight, but do possess relatively simple and coherent visual expressions.

A more particular motivation is the intended display of such visualizations on high-resolution, but small scale, media. These allow for complex designs while at the same time facilitating scrutiny within the span of the eye and data narrations within the flow of text.

We advocate the conscious blending of three established concepts that explicitly leverage human pattern recognition capabilities in small space. Our approach is based on the arrangement of multivariate glyphs (see, e.g., [War02] and [War04, Chapter 5]) in sparklines [Tuf06] to evoke gestalts (see, e.g., [Ste08, Chapter 3]) that correspond to patterns in the data. An application of this approach for the special case of time-varying network data has already been proposed in [BN11]. We here focus on the general design principle and its anecdotal, practical, and empirical validation.

The remainder of this article is organized as follows. The general concept is introduced in Section 2. In Section 3 we demonstrate its application with multivariate data sequences from the literature. A case study of our own is described in Section 4, followed by expert feedback and a user study that provides initial evidence of the usefulness and validity of the method. We conclude with a discussion in Section 5.

2. The Concept

Our approach to multivariate data sequence visualization is an extension of sparklines using glyphs that are designed to take advantage of gestalt laws. We review briefly the three constituting elements before proposing a scheme to integrate them.

2.1. Sparklines

Sparklines are “data-intense, design-simple, word-sized graphics” [Tuf06]. The main rationale for these *datawords* is to allow exploratory visual comparison of large amounts of data *within eyespan*. While high resolution is a prerequisite, however, display media need not be large, and often should not. The concept is best thought of as geared toward fine printing on paper.

Because of their size, sparklines have been referred to as the "Tom Thumb of Statistical Graphs" [Yok09]. Small size facilitates the arrangement of multiple aligned statistical graphics as well as their use directly within text. It allows, e.g., to show data elements repeatedly and exactly where they are referred to, thus eliminating the need to go back and forth between a figure and associated statements. Common examples include line, bar and win-loss charts of, say, measurement time series, sports results or, most prominently, stock quotes. Although we cannot quantify this claim, it appears that most applications of sparklines today involve univariate time series data.

The design space for sparklines is almost as huge as it is for any statistical chart. Given their use in small multiples and in chosen locations inside of sentences, however, slightly different goals may be pursued. For example, a tabular arrangement may call for alignment. Moreover, sparklines can be annotated to convey simple statistics such as normal ranges, as well as specific data points of special interest glucose 6.6 [Tuf06, p. 47] which may be different each time the data are referred to in a text passage.

Note, however, that area considerations may also introduce additional constraints. If a line chart is integrated into text, for instance, its height is constrained by font size; if in addition an average slope of 45 degrees [Cle93, AH06, Tuf06] is desired, then the length of the corresponding dataword is implied as well.

2.2. Glyphs

The term *glyph* is used to refer to a class of graphical objects with several degrees of freedom that can be used to represent multidimensional data points by mapping each dimension of the data to a distinct free parameter. Comprehensive introductions and general design guidelines can be found in [War02, War04].

The main rationale is that uniform depictions of multiple attribute values in a single, complex graphical object are easier to memorize and compare than groups of simpler graphical objects that represent data dimensions separately. A well-known example of this kind are star plots [CCKT83], in which each data dimension is represented along a radial line segment out of a common origin. Bounding the asymmetric stars by filled polygons yields integrated, yet characteristic shapes for each multidimensional data point.

Heterogeneous dimensions and varying measurement scales call for more elaborate designs. Although this is yet another sprawling topic, there are some principled guidelines. As indicated by experimental evidence summarized

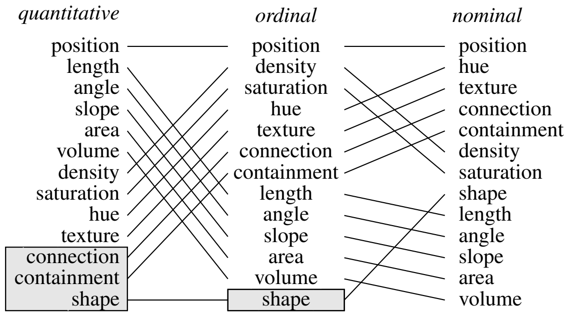


Figure 1: *Ranking of perception accuracy (top-to-bottom) as a guideline for graphical mapping. Boxes indicate variables irrelevant for the corresponding type of data. Redrawn from [Mac86].*

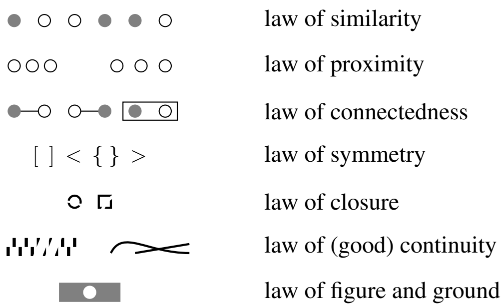


Figure 2: *Seven basic gestalt laws of perception. For example, the law of continuity suggests that we tend to perceive two crossing rather than two touching lines.*

in Figure 1, the various graphical variables yield differential accuracy in elementary perceptual tasks [CM84, Mac86, HB10]. The features of a glyph should therefore be chosen accordingly. These choices may, however, interfere when elements are perceived holistically as in the width and height of a rectangle [War04].

2.3. Gestalt Theory

Wertheimer postulates that the mind organizes disparate visual stimuli into the simplest stable and coherent form [Wer23]. In other words, we are biased toward perceiving wholes, or *gestalts*, rather than collections of individual parts. Based on this so-called *Law of Prägnanz*, gestalt theory consists of qualitative principles such as those listed in Figure 2. A more detailed overview is given in [Ste08].

Previous applications of gestalt theory in visualization design include visual screen design [CDT02], algorithmic animations [EA10], human-computer interaction [FM06], and information dashboard design [Few06].

2.4. Gestaltlines

While the alignment of glyphs in small multiples is a natural extension to depict sequentially or spatially ordered multivariate data (see, e.g., [HBE95, CE97, HPU01, KM01, FCI05]), we posit that two additional design considerations, compactness and gestalt, will bring to bear the real potential of such visualizations. We refer to designs that result from this principle – the arrangement of gestalt theory-informed glyphs in sparklines – as *gestaltlines*.

As a restricted and univariate, but nevertheless inspirational example consider the following illustration of the workings of sorting algorithms [Sed98]. Each element of a data array is represented by a line segment, the slope of which corresponds to the position of that element in sorted order. An unsorted initial array  is thus easily distinguished from a partially sorted intermediate array  and the fully sorted final array . Note the use of slight unimodal length difference to emphasize the visual effect. By visual comparison of intermediate states, the operations of different sorting algorithms are illustrated very graphically.

The main challenge in creating gestaltlines is to identify glyphs and alignment rules from which the presence and specific nature (e.g., location, extent) as well as the absence of certain patterns can be perceived holistically.

Generally important data patterns include clustering and outliers, whereas patterns that are particularly relevant for sequence data include trends, periodicity, disruptions, change points, or phase shifts. Note that in multivariate data such patterns may emerge from combinations of dimensions.

Because of the interdependence of glyph design, arrangements, possibly emerging gestalts and patterns of interest, establishing broadly applicable guidelines for gestaltline design is going to require a major research effort. Note that even such fundamental knowledge as the ranking of graphical variables in Figure 1 may be invalidated when we try to make certain patterns graphic by aiming at specific gestalts.

In the following sections, we will hence start with selected examples of well-known data sets from the literature to illustrate some possible design choices. We then move on to a more elaborate case study of our own and provide empirical evidence for the usefulness and validity of these specific gestaltlines.

3. Examples from the Literature

In this section, we will discuss a few examples from the literature to illustrate the potential of explicit consideration of gestalt principles in the visualization of data sequences.

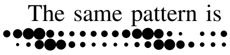
Please observe that the purpose of this discussion is to pinpoint aspects for consideration, rather than the design of the most appropriate diagrams. The examples are meant to demonstrate that using glyphs in sparklines may be only a slight generalization, if any, but that explicit consideration of gestalt principles does make a difference.

3.1. Phase Shifts in Population Dynamics

The dynamics of predator-prey populations are examples of bivariate data sequences. We here use a classic data set in which fur trade records indicate the population size of Canadian lynx and snowshoe hares between 1900 and 1921 (see, e.g., [Odu71]).

Such pairs of sequence data can be depicted straightforwardly in sparklines using superimposed line  or bar  charts. The data are represented with high accuracy and comparison of the size of the two populations in any given year is easy.

The dominant and well-known pattern in this data is one of periodic peaking of both populations, with predators lagging behind prey.

The same pattern is also visible in an alternative design  using pairs of dots with areas proportional to population sizes. Here, the law of similarity suggests that diagonal grouping of large dots is more immediate than vertical grouping to time. The slope of perceived diagonals is an indicator of the lag between population surges. Since the perceived diagonals are approximately parallel, the lag is roughly the same between both pairs of peaks and phase changes stick out.

So far this re-iterates what is also obvious from the line or bar chart representations, and we even paid a price because relative areas are perceived less accurately than relative positions or lengths (recall the ranking in Figure 1). In a longer sequence involving more peaks, however, the law of similarity also applies on another level. If there was a period in which the lag differs, this is more easily recognized as an outlier among the otherwise similar diagonals. In the extreme case that the sign of the slope is reversed, the outlier is detected pre-attentively. A population of predators surging before the prey would be a very interesting pattern. Due to the strong separation of dimensions in line or bar charts of color coded populations such a pattern is more easily overlooked than in the case of an emergent atomic feature, as is the slope of a perceived diagonal.

3.2. Streaks in Sports Results

Tufte uses the win-loss charts for baseball teams to demonstrate that "Sparklines can simultaneously accommodate several variables" [Tuf06, p. 55]. These charts contain a tick for each game  and the tick's location above or below an imaginary center

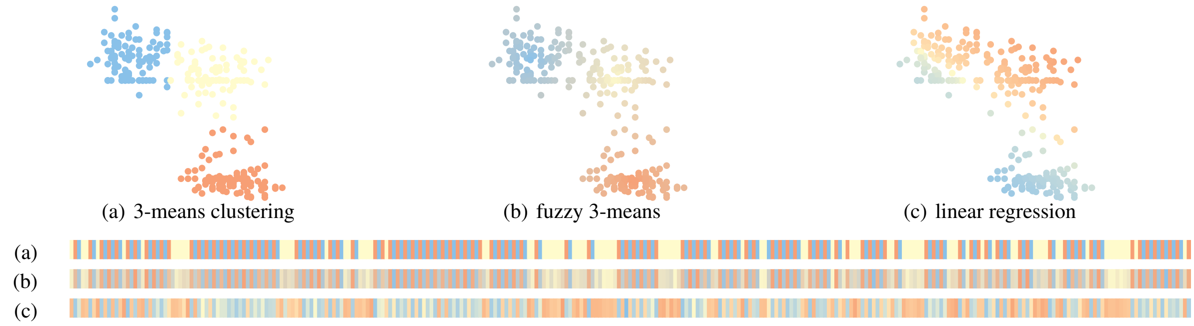
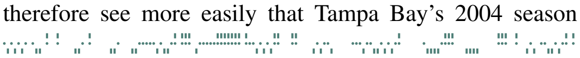


Figure 3: *Scatterplots of geyser eruption patterns showing waiting times (x-axis) versus subsequent eruption duration (y-axis) with colors indicating membership in cluster models (a),(b) and estimation error in regression model (c). The gestaltlines below show the actual sequence of eruptions using the same colors.*

line indicates whether the game was won or lost. In addition, the centerline is actually drawn when the game was played at home.

By the law of proximity, streaks of wins and losses are grouped, and by the law of connectedness, partially contradicting stretches of home games are perceived as units. Since home games are statistically more likely to be won, however, large parts of both groupings may be induced by the season schedule and not indicative of interesting variation in performance.

To eliminate some of the redundancy, and to focus more on performance variation, we could omit expected results  and show only home losses and away wins so that grouping occurs for the performance-induced and relatively surprising results. In this way, the two data dimensions, home-away and win-loss, are no longer represented separately in presence-absence of a horizontal line and above-below center placement of a vertical line.

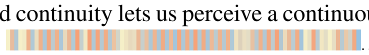
Observe that horizontal merging of home-game lines is visually dominant. Since home-away schedules in baseball are streaky by design, this may not be the most interesting piece of information, however. In a variant gestaltline, we only place dots on the center line to indicate a win. Using above-below ticks as before, every single outcome can still be uniquely decoded, but groups and gaps on the center line now indicate streaks of wins or losses. We therefore see more easily that Tampa Bay's 2004 season  started out ordinarily with a few too many home losses. They had a long winning streak midseason, including a series of away wins, and after roughly two-thirds of the season there are two particularly poor stretches of home losses.

In comparison to the original sparkline, we have transformed the data to be relative to a baseline (wins at home and losses away) to reduce visual complexity, and determined glyph parameters from combined data dimensions to place more emphasis on the most important aspect (wins).

3.3. Periodicity in Geyser Eruptions

An example of complex repetitive patterns are the eruption sequences of Old Faithful. The geyser, which faithfully erupts about twenty times a day, is a major tourist attraction in Yellowstone National Park, Wyoming, USA. Understanding and predicting the geyser's behavior has been subject to various scientific studies. Among the most recognized ones is the investigation of Azzalini and Bowman [AB90] which has been based on 299 successive observation pairs of waiting time between the starts of eruptions (43 to 108 minutes) versus duration of the following eruption (50 to 327 seconds). Both waiting time and duration are bimodal distributed, and the scatterplot in Figure 3 reveals three distinct eruption patterns that can be recovered from 3-means clustering or simple thresholding just the same.

A gestaltline ... with adjacent stripes colored according to a 3-means clustering model reveals a known periodicity: stretches of an alternating sequence of short waits for long eruptions and long waits for short eruptions are interrupted by shorter stretches of long waits for long eruptions.

Replacing the partition into groups by fuzzy memberships values, however, yields a gestaltline in which the law of good continuity lets us perceive a continuous alternating pattern ... that is only subdued during what seemed to be interrupts. To the best of our knowledge, this observation is even a new finding not yet reported in the literature.

In comparison to the clustering, the regression model ... appears to yield much less systematic outcomes.

3.4. Exceptions in People Flow

Ihler, Hutchins, and Smyth recorded the number of people going in and out of a building on the University of California at Irvine (UCI) campus over a 15-week period in 2005 [IHS06]. Their interest was in modeling relatively stable multilevel (daily, weekly, seasonal) behavioral patterns to detect unusual events.

In their original publication, these data are depicted in standard scatterplots of occupancy or line charts of entry numbers [IHS06, Figures 1 and 3]. Provided an explanatory variable and a period can be fixed in advance, outliers and periodicity are fairly easy to recognize in such diagrams.

Now consider the gestaltlines in Figure 4 which demonstrates how the recorded data could be shown in full on a single page. Arranging each day in a row of its own supports the detection of weekly patterns such as low building occupancy during weekends, and the vertical alignment of daytime supports the detection of daily patterns such as lunch break times. Unusually late arrivals on Wednesday of the first week (Figure 4(a)) and a Monday holiday (Figure 4(b)) stick out by breaking perceived groups of similar dots.

So far, we have made similar use of glyph parameters determined from combined data dimensions to direct attention to information derived from the data rather than to the raw data itself.

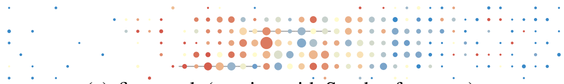
Some apparent outliers in this data can be related to known special events taking place in the building. An additional horizontal line segment in the background connects the dots inside the time interval corresponding to such events. At least this is how we perceive what might actually be short line segments between neighboring dots. This is because of the law of closure (the segments are aligned and of equal appearance) and the law of figure and ground (consistent gray color). By the law of connectedness, the entire occasion is perceived as a whole and discounted for when eyeing for groups in the remaining data.

4. Case Study

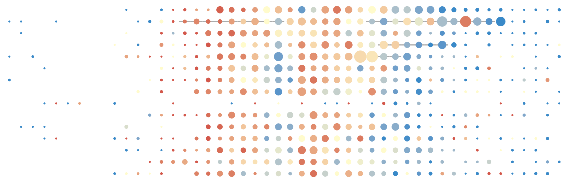
The examples of the previous section are well-behaved in the sense that a relatively simple data set exhibits relatively clear patterns from known categories.

We now turn to a data set that is more complex in the number of dimensions and relations between them, and for which no prior knowledge about the presence of a pattern exists. In fact, these data motivated the research reported here. Our goal is to demonstrate how the design ideas of the previous section can be applied in less controlled, and thus more realistic, situations.

After describing the origin and nature of the data, we elaborate on an exemplary gestaltline design. We conclude the case study with expert feedback on the improved interpre-



(a) first week (starting with Sunday from top)



(b) all Mondays

Figure 4: *Net flow of people entering and leaving a building [IHS06, FA10]. Each row represents a day, each column a 30 minute interval. The area of a dot is proportional to the maximum of the number of people entering and leaving the building within the corresponding half hour, whereas the color indicates the ratio of in- and out-flow on a color scale from red (in) via yellow (balanced) to blue (out). Horizontal background lines indicate known exceptional events taking place in the building.*

tation of their data, and a user study that provides formal evidence on the usefulness and validity of gestaltlines.

4.1. Background and Data

The data originate from a psychological experiment that was conducted as part of a larger study on the influence of early-life stress on human reward processing and decision making. Further details are reported in [Ste10].

Thirty subjects participated in the experiment. Eighteen of them were psychiatric patients who experienced early life stress (10 female, average age 39.1 ± 12.6), half-and-half to a low and high degree. The other twelve subjects form a healthy control group (7 female, average age 43.4 ± 17.2).

In a repeated measurement design, each participant was subjected to a sequence of 240 computerized gambling trials while being measured for brain activity. As within-subject factors, 10 (euro) cents or 50 cents were at stake in each trial and the announced chance of winning was 10%, 50%, or 90%. Each of the six variants was played 40 times, with the entire sequence in individually randomized order. Subjects had to decide whether they wanted to pass or play. Hence, there are four possible outcomes for each trial: passed, played and won, played and lost, or no decision until a time-out of 2 sec. Each outcome was presented to the subject before the next trial.

With a theoretical maximum gain or loss of $120 \cdot 50\,\text{cents} + 120 \cdot 10\,\text{cents} = 72\,\text{EUR}$, purely rational decisions (play at 90% chance of winning, pass at 10%, and any strategy at 50%) yield an expected gain of $40 \cdot [0.9 \cdot (50\,\text{cents} + 10\,\text{cents}) + 0.1 \cdot (-50\,\text{cents} - 10\,\text{cents})] = 19.2\,\text{EUR}$. Subjects were given a starting budget of 10 EUR and received an actual payment between 0 EUR and 20 EUR determined from 20 randomly chosen trial outcomes.

The experiment thus generated 30 sequences of 240 four-dimensional data points (stake, chance, decision, outcome) with 24 possible values. Analysis on the aggregate level confirmed some expected differences between patients and control subjects in cortical activation patterns and also in the number of irrational decisions (play at 10% chance of winning, pass at 90%).

Interestingly, however, five of the seven subjects beating the expectation were patients, three of them even with a high stress level. A closer look at the detailed data is supposed to test for systematic effects such as strategy learning or the onset of boredom that do not show on the aggregate level.

4.2. Gestaltline Design

Similar to the examples from Section 3, we display each subject within a separate gestaltline and represent all 240 four-dimensional data points in sequence order on the horizontal axis left-to-right, i.e. time progresses in Western text reading direction.

To support the domain experts in noticing systematic effects in decision making, we choose glyphs such that different gambling strategies induce holistic forms. Given the noise produced by randomization of game settings, we present the data with regard to a baseline (Section 3.2), instead of ranking data dimensions by importance and mapping them separately to an equal number of the most accurately perceived graphical features in the same order (Figure 1).

There are two data dimensions of utmost importance. On the one hand, the deliberate decision to play, pass, or wait until timeout, is the only data dimension that is non-randomized and directly attributed to the subject. On the other hand, the announced chance of winning is the pivotal determinant of the randomized gambling design, because this information alone determines the sign of the expected gain and should hence figure prominently in any decision. Consequently, we propose a baseline – rationality – that combines these two data dimensions.

Clearly, rationality of a subject's decision is defined relative to announced chance and subsequent decision – rational decisions: play at 90%, pass at 10%; irrational decisions: pass at 90%, play at 10%. However, rationality can not be assessed for decisions at 50% chance of winning, since the expected gain/loss is zero. Instead, we can score such trials as active (play at 50%) or passive (pass at 50%) decisions. Both concepts, rationality and joy of playing, overlap in a sense that extremely active (play at 10%) or passive (pass at 90%) decisions become irrational. Thus, we can order decisions from one extreme (irrational-passive) to the other (irrational-active) – pass at 90%, pass at 50%, pass at 10%, play at 90%, play at 50%, play at 10% – relative to a baseline (rational decisions; pass at 10%, play at 90%).

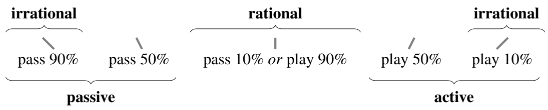


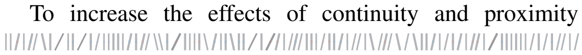
Figure 5: *Visualization and **interpretation** of participant decisions relative to a rational baseline.*

A corresponding graphical feature to map this ordering can be borrowed from Sedgewick's illustrations (Section 2.4). That is, each trial is represented by a line segment, the slope of which corresponds to the interpretation of that decision with regard to our baseline; cf. Figure 5. We found this mapping intuitive, since the leaning of line segments suggests a correspondence to the subjects' postures during the experiment, such as sitting straight during rational decisions and leaning forward while gambling.

Integrating the remaining data attributes into these basic glyphs is straightforward. Note that we use the same vertical line segment for both rational decision, since we do not interpret them as being active or passive. To distinguish these decisions, we exploit that outcome of pass at 10% (no change in a player's budget) is different from play at 90% (change in budget). Concretely, outcomes are represented only by small colored dots (blue for profit, red for loss) in the center rather than coloring the entire line segment to promote strategy information over outcome. For the same reason, irrational decisions are slightly highlighted with a stronger glyph hue.

Finally, line width and dot size are determined by the amount of money at stake and the occurrence of a timeout is distinguished from authentic pass decisions by the use of white dots. In this way the colored dot areas correspond directly to the change of budget. Using less important graphical features for less important experimental conditions and special outcomes facilitates macro-micro reading [Tuf90] in which the participant strategy is represented as the visually dominant information, and details about the individual trials can be scrutinized if desired.

To increase the effects of continuity and proximity



we decided against a monospace arrangement. That is, the (horizontal) space of a single glyph is proportional to the deviations in our baseline (Figure 5), and we use kerning to reduce space in between consecutive glyphs.

These design decisions take advantage of several gestalt laws: Local trends in decision making go along with similar leaned line segments /// (law of similarity), and little whitespace in between (law of proximity). Still, slight variations in the leaning of line segments resulting from the randomized experiment do not hamper the detection of stable decision patterns (law of continuity). Moreover, strategy changes are promoted by the law of symmetry, and irrational decisions are highlighted with prominent line segments (law of figure and ground). Finally, placing small dots in the center of line segments does not influence the perception of leanings (law of closure).

4.3. Expert Feedback

The visualization of the data as it was presented to the domain experts is reproduced in Figure 6. They were excited to "*see*" their data in its entirety and commented several times on the gestaltlines' elegance and compactness. In the following we report on some of their interpretations.

The first thing to note is that participants with a net gain around the expected 19.2 EUR played rationally for the most part, which was to be expected. Secondly, there are some obvious differences in strategy. Participant C_1, for instance, employs an aggressive, risk-taking strategy start to finish; whereas subject H_3 starts out aggressively active and acts increasingly errant toward the end, where many irrational decisions go together with the occurrence of a timeout — possibly indicating the onset of boredom. As another example, player L_8 employs a risk-avoiding strategy that occasionally results in irrational decisions. A rare case of learning appears to be participant L_2 for whom the chart calms considerably from the first quarter of games to last quarter of games.

Seeing the entire data in this way, the domain experts found no indication that the analysis on the aggregate level was distorted by unlucky patients or superior learning of control subjects. This was reassuring for the earlier analysis. On the other hand, the domain experts confidently hypothesized that there is more within-class variation among patients than among controls, and that this variation is larger with respect to strategy than performance. While this suggests further substantive work beyond the scope of this paper, it should be noted that the strategy hypothesis would have been difficult to arrive at analytically.

4.4. User Study

Feedback from the domain experts supported our driving hypothesis that sparklines made from gestalt theory-informed glyphs facilitate holistic recognition of patterns, trends, and outliers in multivariate sequences. However, the gestaltlines have been designed to resonate with their understanding of the substance. To gather evidence on the general validity of

	active	passive	rational	irrational
/	92.3%			92.3%
/	94.2%		21.2%	1.9%
|			98.1%	
\		92.3%	21.2%	1.9%
\		90.4%		94.2%

Figure 7: *Results for introductory question of user study (Section 4.4); **Q1** "Please mark for each row the appropriate fields". Unexpected answers highlighted in red.*

our design, we tested whether it appeals to a broader audience. Concretely, we conducted an unannounced study with a group of 52 second-year undergraduate students in computer science who had no prior background in information visualization.

Subjects were handed a questionnaire with 12 items on the gestaltlines from Section 4.2 together with a two-sided A4 handout explaining their design and its basic interpretation. Both handout and questionnaire had been refined following several pretests. The complete user study was limited to 35 minutes, and the fastest subjects needed 17 minutes to read the handout and answer the questionnaire.

Initially we tested the subjects' understanding of decisions being represented relative to a baseline defined by expected outcome (Figure 5). Subjects were asked to classify single decisions (**Q1**) as well as small sequences of glyphs (**Q2**). As summarized in Figure 7, the vast majority of subjects understood the basic design. In particular, not a single decision was unexpectedly classified as passive or active, and there was only one outlier concerning rationality. We were mildly surprised, however, to find that more than one fifth of the subjects interpret every decision at 50% chance as being rational. On the other hand, subjects were able to match small sequences of rational (100%), passive (98.1%), active (98.1%), and irrational (96.2%) strategies with the expected label.

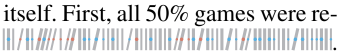
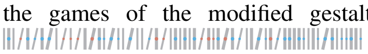
Subsequently, similar results were obtained for classifying a complex sequence of rational-passive decisions with noise (**Q4**; 88.5% rational, 78.9% passive), and identifying most (ir)rational strategies among various gestaltlines (**Q7**; 96.2% expected answers). Likewise, subjects were able to find the most prominent strategy change across two gestaltlines (**Q8**; only 3 subjects selected unexpected gestaltline; 7 subjects selected unexpected breakpoint). About 80% of the subjects were able to mark the longest uninterrupted subsequence of rational decisions by player H_2 (**Q9**; Figure 6).

In **Q3**, we asked subjects to mark all (irrational) outliers within a complex sequence of otherwise rational decisions. Notably, each outlier instance has been marked by at least 86.5% (and up to 94.2%) of the subjects. Somewhat disappointingly one third of the subjects also marked a pass-at-50%-chance decision that was surrounded by play-at-50%-chance decisions. While this constellation indeed results

Figure 6: *Sequence data for all 30 subjects with 240 trials each using the gestaltlines as defined in Section 4.2. Subjects are partitioned into groups and ordered by net gain/loss (expectation for a rational strategy is 19.2). Two general observations are larger gains for rational subjects and few apparent strategy changes over the course of the experiment. More detailed interpretation in the main text.*

Figure 8: *Results for **Q11** "Please divide the sequence into sub-sequences and label them with a brief description" (irrational, rational, **outliers**, passive, active). Answers split into irrational vs. rational* (left) *and passive vs. active* (right) *with **outliers** highlighted. Top two rows indicate our predictions, others the actual answers (one row per subject).*

in perceiving increased leanings of glyphs [TGH12], well-trained readers would have noticed that irrational decisions are slightly highlighted with a stronger glyph hue. Still, subjects found an interesting subsequence (after losing two 50% games, the next one is passed) and potentially have interpreted this temporal change in decision making to be irrationally inconsistent.

Since an occasional confusion of rationality and consistency was already observed in our pretests, additional questions served to disentangle observations from interpretations: We asked subjects to compare a sequence of rational decisions  with two modified versions of itself. First, all 50% games were redefined as being played . 88.2% of subjects (expectedly) evaluated the modified sequence to be more consistent (**Q10**), but 51.9% of subjects also (unexpectedly) evaluated this sequence to be more rational (**Q5**; 38.5% equally rational). Second, the games of the modified gestaltline were permuted  to match the length of the original gestaltline. As a result, the majority (54.9%) of subjects evaluated both gestaltlines to be equally rational (**Q12**); only about half of the subjects did realize that the length of gestaltlines is influenced by rationality of decisions, consistency of decisions, and permutation of games.

Another impression from the pretests was reproduced in **Q6**: We asked "Did the following subject go lucky?" with regard to a sequence of games that were primarily won, but did involve unfavorable outcomes alike. While the correct answer was "No" (almost all 50% games were lost), about half of the answers were wrong ("Yes"), with explanations such as "more blue dots" / "big wins". That is, subjects did observe correctly the relevant information that needs to be scrutinized (colored dots), but about half of their interpretations did not succeed in disentangling (expected) profit from (unexpected) luck.

Finally, subjects were asked to annotate a short sequence of complex patterns in decision making with strategies, breaks, and outliers (**Q11**). The reassuring results are summarized as gestaltlines in Figure 8.

Our user study thus demonstrates that a short briefing is sufficient for untrained readers to reliably find holistic patterns, outliers and breaks within the proposed visualizations.

5. Conclusions

We proposed a conceptual design approach aiming for compact graphical representations of multivariate data sequences. It consists of the arrangement of especially designed glyphs in sparklines such that patterns of interest yield gestalts which can be perceived holistically and pre-attentively.

We feel that our initial explorations with illustrative examples, a detailed case study, and its internal validity test show that, as a design principle, the concept of gestaltlines is viable. We see a wide-open space for creative research into glyph design and alignment based on this principle. But two examples for exciting future research topics are the search for alternatives to existing representations such as separation plots [GWS11] and the extension to hierarchical patterns.

Given the vastness of the potential design space and usage scenarios, however, comprehensive design guidelines are far beyond the scope of this paper. Much detailed research will be needed to assess with confidence the effectiveness of gestaltlines for specific data patterns in specific applications. Note that studies such as [FFM*13] are likely to yield different results when considering gestalt-informed glyphs and arrangements. It will also be important to understand the relative reading accuracy and efficiency of variant gestaltlines and alternative graphical designs. In addition to task and design-related factors, comparative studies will have to take contextual factors such as the available media into account.

Like every form of visualization, gestaltline design is limited by constraints such as resolution, number of discernible colors, or shape complexity. Where these boundaries are is yet another question that we cannot answer today.

Acknowledgments. Raúl Soriano-Hontanilla participated in the exploratory phase of this research (supported by the ERASMUS student exchange programme) and Tobias Döbele and Volker Mühlberg have been instrumental in conducting the user study. Four very considerate anonymous reviews helped improve the presentation. We gratefully acknowledge financial support from the Stiftung der Deutschen Wirtschaft (stipend to Astrid Steffen) and the University of Konstanz (grants FP 626/08 and FP 665/10).

References

[AB90] Azzalini A., Bowman A.: A look at some data on the old faithful geyser. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 39*, 3 (1990), 357–365. 4

[AH06] Agrawala M., Heer J.: Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 701–708. 2

[BN11] Brandes U., Nick B.: Asymmetric relations in longitudinal social networks. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2283–2290. 1

[CCKT83] Chambers J. M., Cleveland W. S., Kleiner B., Tukey P. A.: *Graphical Methods for Data Analysis*. Chapman and Hall, 1983. 2

[CDT02] Chang D., Dooley L., Tuovinen J. E.: Gestalt theory in visual screen design: a new look at an old subject. In *Proceedings of the 7th World Conference on Computers in Education* (2002), CRPIT '02, pp. 5–12. 2

[CE97] Chuah M. C., Eick S. G.: Glyphs for software visualization. In *Proceedings of the 5th International Workshop on Program Comprehension* (1997), pp. 183–191. 3

[Cle93] Cleveland W. S.: *Visualizing Data*. Hobart Press, 1993. 2

[CM84] Cleveland W. S., McGill R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79*, 387 (1984), 531–554. 2

[EA10] Esponda-Argüero M.: Techniques for visualizing data structures in algorithmic animations. *Information Visualization 9*, 1 (2010), 31–46. 2

[FA10] Frank A., Asuncion A.: UCI machine learning repository, 2010. URL: http://archive.ics.uci.edu/ml. 5

[FCI05] Fanea E., Carpendale S., Isenberg T.: An interactive 3d integration of parallel coordinates and star glyphs. In *Proceedings of the 2005 IEEE Symposium on Information Visualization* (2005), pp. 149–156. 3

[Few06] Few S.: *Information Dashboard Design : The Effective Visual Communication of Data*. O'Reilly, 2006. 2

[FFM*13] Fuchs J., Fischer F., Mansmann F., Bertini E., Isenberg P.: Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), CHI '13. 9

[FM06] Flieder K., Mödritscher F.: Foundations of a pattern language based on gestalt principles. In *Extended Abstracts on Human Factors in Computing Systems* (2006), CHI EA '06, pp. 773–778. 2

[GWS11] Greenhill B. D., Ward M. J., Sacks A. E.: The separation plot: A new visual method for evaluating the fit of binary models. *American Journal of Political Science 55*, 4 (2011), 991–1002. 9

[HB10] Heer J., Bostock M.: Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), CHI '10, pp. 203–212. 2

[HBE95] Healey C. G., Booth K. S., Enns J. T.: Visualizing real-time multivariate data using preattentive processing. *ACM Transactions on Modeling and Computer Simulation 5*, 3 (1995), 190–221. 3

[HPU01] Horn W., Popow C., Unterasinger L.: Support for fast comprehension of ICU data: visualization using metaphor graphics. *Methods of Information in Medicine 40*, 5 (2001), 421–424. 3

[IHS06] Ihler A., Hutchins J., Smyth P.: Adaptive event detection with time-varying poisson processes. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2006), KDD '06, pp. 207–216. 5

[KM01] Kosara R., Miksch S.: Metaphors of movement: a visualization and user interface for time-oriented, skeletal plans. *Artificial Intelligence in Medicine 22*, 2 (2001), 111–131. 3

[Mac86] Mackinlay J.: Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics 5*, 2 (1986), 110–141. 2

[Odu71] Odum E.: *Fundamentals of Ecology*, 3rd ed. Saunders, 1971. 3

[Sed98] Sedgewick R.: *Algorithms in C*. Addison-Wesley, 1998. 3

[Ste08] Sternberg R. J.: *Cognitive Psychology*, 5th ed. Wadsworth Publishing, 2008. 1, 2

[Ste10] Steffen A.: Decision-making in the (stressed) brain. Dissertation. Available at KOPS., 2010. URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-126600. 5

[TGH12] Talbot J., Gerth J., Hanrahan P.: An empirical model of slope ratio comparisons. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2613–2620. 9

[Tuf90] Tufte E. R.: *Envisioning Information*. Graphics Press, 1990. 6

[Tuf06] Tufte E.: *Beautiful Evidence*. Graphics Press, 2006. 1, 2, 3

[War02] Ward M. O.: A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization 1*, 3/4 (2002), 194–210. 1, 2

[War04] Ware C.: *Information Visualization: Perception for Design*, 2nd ed. Morgan Kaufmann, 2004. 1, 2

[Wer23] Wertheimer M.: Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung 4* (1923), 301–350. 2

[Yok09] Yokum T.: Sparklines: The Tom Thumb of Statistical Graphs. *Foresight: The International Journal of Applied Forecasting 14* (2009), 48–50. 2