The Connected Scatterplot for Presenting Paired Time Series

Steve Haroz, Robert Kosara, and Steven L. Franconeri

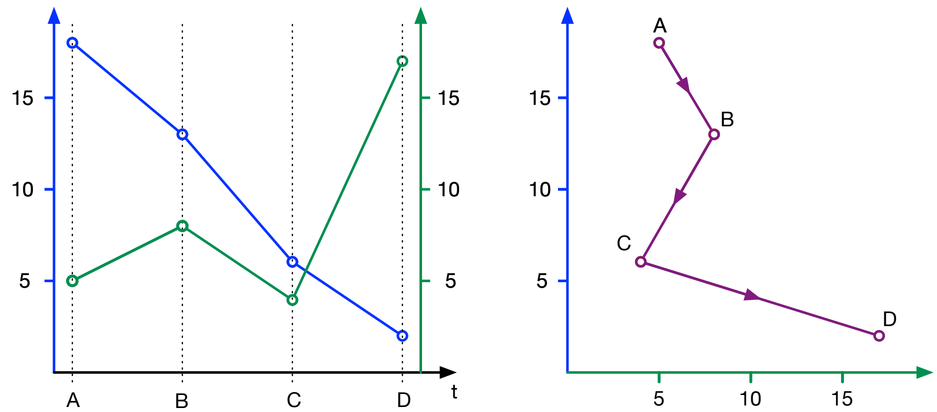


Fig. 1. Two simple time series plotted in a dual-axis line chart (DALC, left) and a connected scatterplot (CS, right). An interactive tool that translates between a DALC and a CS is available at `http://steveharoz.com/research/connected_scatterplot`

Abstract—The *connected scatterplot* visualizes two related time series in a scatterplot and connects the points with a line in temporal sequence. News media are increasingly using this technique to present data under the intuition that it is understandable and engaging. To explore these intuitions, we (1) describe how paired time series relationships appear in a connected scatterplot, (2) qualitatively evaluate how well people understand trends depicted in this format, (3) quantitatively measure the types and frequency of misinterpretations, and (4) empirically evaluate whether viewers will preferentially view graphs in this format over the more traditional format. The results suggest that low-complexity connected scatterplots can be understood with little explanation, and that viewers are biased towards inspecting connected scatterplots over the more traditional format. We also describe misinterpretations of connected scatterplots and propose further research into mitigating these mistakes for viewers unfamiliar with the technique.

1 Introduction

Data visualizations can be used for both exploration and presentation, but journalists are primarily interested in the latter. For presenting paired time series, news media have recently begun using a technique called the *Connected Scatterplot* (CS). One of the first uses of this technique in news graphics was *Oil's Roller Coaster Ride* by Amanda Cox for *The New York Times* in February 2008 [9] (Figure 2). Since that article was published, over a dozen other instances of connected scatterplots have appeared, with the number of uses increasing dramatically in 2013 and 2014. Table 1 lists the majority of these charts that have appeared in the news media.

Although the CS may be new to journalists and their audience, similar charts have been used for hundreds of years to explore time series data – the development of this style of plot even coincided with some of the earliest data graphing by William Playfair. One of the first examples, a physical device called a *steam indicator*, was developed by John Southern in the 1790s (though often credited to James Watt [32]). It drew the cycle of a steam piston over time, graphing piston position against steam pressure to show the timing of the movement, valves opening and closing, and total power output (the area within the curve). In 1958, another connected scatterplot depicted a part of a economic labor model [13], graphing the unemployment rate against the rate of job openings. Known as the Beveridge Curve, Philips curve, or Unemployment-Vacancy rate (UV) curve [40], the shape of this plot can act as an indicator for the state of an economy.

Because the technique has been used for centuries as an analysis tool and has experienced a surge in recent years as a means for communicating data, we were surprised to find minimal testing of the technique's clarity. Although Robertson et al. [39] explored ways of reducing clutter for multiple simultaneous connected scatterplots, we found no experimental evidence that compared comprehension of a connected scatterplot with other static representations. We therefore began with an informal survey of four journalists working for major U.S. news organizations (both daily newspapers and magazines) to learn why they believed the technique to be useful, and these conversations inspired a set of four experiments designed to evaluate their intuitions. Here we explain the construction and idiosyncrasies of the connected scatterplot, discuss four experiments, and present a set of conclusions, guidelines, and open questions.

2 The Connected Scatterplot Technique

We conducted informal interviews with journalists who produce data visualizations. All of the journalists said that the Connected Scatterplot (CS) was novel to them, and most likely to their readers, before they used it. In his book, Alberto Cairo even called it *a most uncommon kind of scatter-plot* [6]. Furthermore, none had heard of the Beveridge Curve, as the the first journalistic use of the technique was inspired by a paper about oil markets [18].

The CS technique depicts two simultaneous time series. A traditional way to plot these datasets would be a dual-axis line graph (DALC), which typically maps the time dimension to the horizontal axis and the series' values onto the vertical axis (Figure 1). The CS,

- *Steve Haroz is with Northwestern University. E-mail: cs@steveharoz.com*
- *Robert Kosara is with Tableau Research. E-mail: rkosara@tableau.com*
- *Steven L. Franconeri is with Northwestern University. E-mail: franconeri@northwestern.edu*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014; date of publication xx xxx 2014; date of current version xx xxx 2014. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

Title (Year published)	L/U-Shapes	Loops	Crossings	Series Pairs	Reference
Oil’s Roller Coaster Ride (2008)	3	1	2	1	[9]
Driving Shifts Into Reverse (2010)	10	1	1	1	[14]
Driving Safety, in Fits and Starts (2012)	9	1	1	1	[15]
The Rise of Long-Term Joblessness (2013)	many	many	many	1	[10]
Helium Supply (2013)	9	2	2	1	[36]
Chart redraw: Troops Vs. Cost (2013)	3	0	0	1	[7]
Janet L. Yellen, on the Economy’s Twists and Turns (2013)	5	many	many	1	[19]
Holdouts Find Cheapest Super Bowl Tickets Late in the Game (2014)	1-3	0	1	5	[46]
The Fed’s Balancing Act (2014)	0-many	0-many	0-many	6	[34]
Il giocattolo si è rotto (2014)	21	0	0	1	[30]
Graduation, marijuana use rates climb in tandem (2014)	2	0	0	1	[27]
Wage Growth Is No Longer as Sensitive to Labor Market… (2014)	many	many	many	1	[51]
In Short-Term Unemployment Data, Good and Bad News (2014)	1	(unclear)	many	1	[17]
Wealth and height in the Netherlands, 1820-2013 (2014)	1	0	0	1	[35]
Obama’s approval versus the economy (2015)	7	(unclear)	(unclear)	4	[5]
What Should We Expect U.S. Wage Growth To Be? (2015)	many	many	many	1	[51]
National Indebtedness (2015)	many	many	many	47	[2]
The M. Night Shyamalan Twist (2015)	4	0	0	1	[25]
How the U.S. and OPEC Drive Oil Prices (2015)	(unclear)	(unclear)	(unclear)	4 (sequential)	[3]

Table 1. Examples of connected scatterplots in news graphics (and two blog postings) that we examined. The examples with ‘many’ loops or L/U-shapes are generally plots of unemployment paired with vacancy rate or inflation, and thus similar to the original Beveridge Curve.

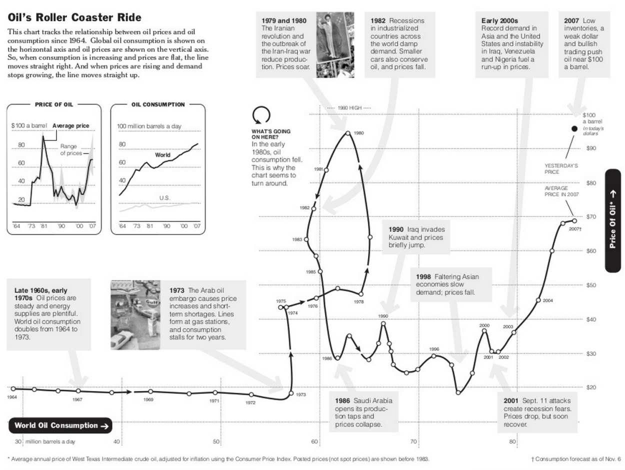


Fig. 2. *Oil’s Roller Coaster Ride* [9] uses the connected scatterplot to show the relationship between oil consumption (horizontal axis) and oil price (vertical axis) over time. The prominent loop draws readers’ attention, annotations point out particular points of interest.

however, maps two values onto a 2D Cartesian plane, with one time series being represented on the horizontal axis, the other on the vertical (visible as different colors in Figure 1). A line is drawn to connect the points in temporal order. Note that the common time sampling, and the line that represents its progression, could in theory be replaced with any other strictly monotonically increasing dimension.

A useful metaphor for thinking about the connected scatterplot is the *Etch-A-Sketch*. Two knobs on the front surface of this popular American children’s toy control the vertical and horizontal direction, respectively, of a stylus that draws on a glass screen. In this metaphor, each time series controls one of the knobs, as time is incremented from the start to the end time. The result is a two-dimensional image that reflects the changes in those values on each axis. It also explains why there is no change in the connected scatterplot when there is no change in the values within each time series across time steps, because the knobs are not moved.

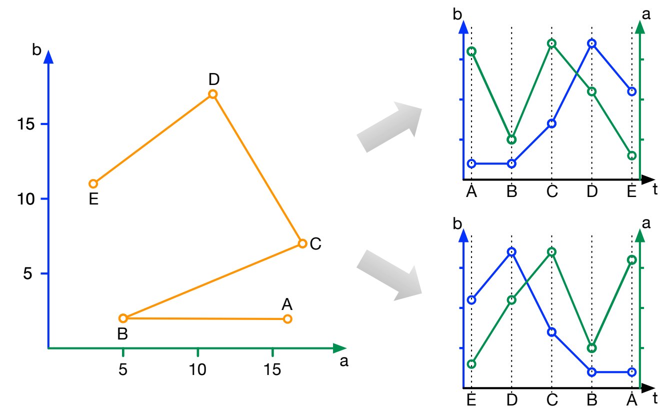


Fig. 3. Indicators for reading direction like arrows are key to correct interpretation of the connected scatterplot. The example above has two potential interpretations (A-E or E-A), depending on which direction it is read in (notice the time axis).

2.1 Components

While a simple technique in principle, the components of the connected scatterplot have very specific functions.

Points are typically shown as dots or circles. They help users see when the values were sampled. Since time is not directly represented in the plot, the points are an important indicator of time steps, as their spacing indicates the rate of change. In contrast to the DALC, the connected scatterplot (CS) also generally requires that the points for each time series are sampled at the same times (see also Section 2.4 below).

One exception which has irregularly spaced samples is a graphic in Wired Italy [30], which plots inequality vs. GDP in Italy over 150 years. It draws a line for each prime minister, whose office terms vary considerably (from one year to over a decade).

Lines connect consecutive points, allowing the observer to see temporal connections, as well as giving the data a shape. Without the lines, the chart simply reverts to a traditional scatterplot, with no indication of sequence.

Arrows Without an indication of the direction of time, a connected scatterplot can be drastically misinterpreted (Figure 3). There are other

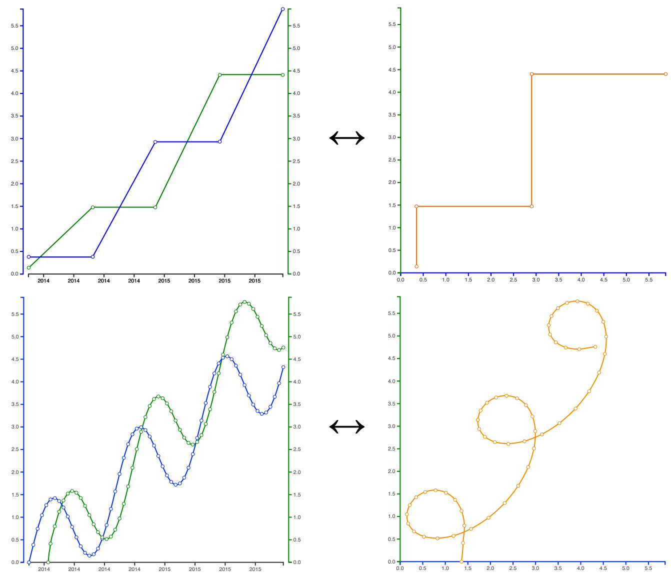


Fig. 4. (Top) A right angle indicates a sharp change in the rate of change of one series or a swap of the rates of change between series. (Bottom) Loops in the connected scatterplot are the result of similar patterns that are shifted by up to a quarter of the periodicity of the pattern (Section 2.2).

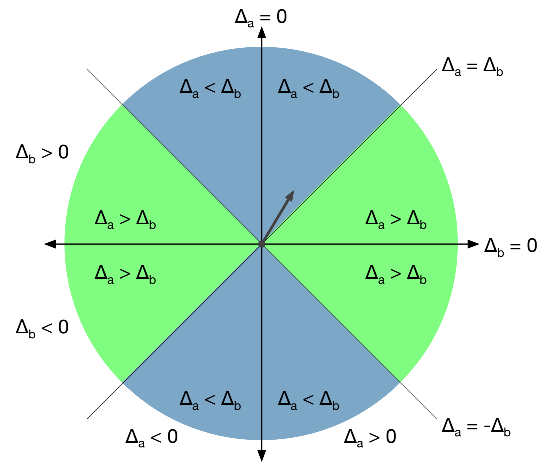


Fig. 5. Line direction as a function of the difference in value between the two time steps in each series. Δa indicates the difference on the horizontal axis, Δb the difference on the vertical axis.

ways of indicating direction, such as lines with varying thickness, gradients, or even animation, but the majority use arrows. These arrows can be omitted when the direction is explained separately (e.g., with symbols indicating the start and end of the line), when the points are labeled, or when there is an obvious direction (usually left to right) explained in the text. However, this alternative makes the chart less self-contained, requiring the reader to seek critical information.

2.2 Distinctive Shapes: Ls and Loops

Connected scatterplots often contain two particularly interesting features: L-shapes and loops (Figure 4). Both are visually salient features and unusual (L-shapes), if not impossible (loops), in line charts.

L-shaped features, where the line changes direction at close to 90°, are visually salient and potentially reflect important patterns in the data. They represent sudden changes in the relationship between the two time series, for example if one variable remains constant while the other is changing, an L appears when this pattern suddenly reverses (Figure 4, top).

Loops often indicate a temporal shift between the series. For each local maximum and minimum pair (a peak and a valley), which occurs at different times in each series, a loop will appear (if the series is truncated, the first or last loop may be incomplete). The offset between series can be as short as a single time interval or be up to half the length of the series. The number of time intervals in the loop indicates the size of the temporal offset. Meanwhile, the direction of the loop indicates in which time series the pattern occurs first, which might suggest a causal relationship. A clockwise loop means that the series on the vertical axis starts the pattern first, and a counter-clockwise loop indicate that the pattern appears first on the horizontal axis. See Figures 2 and 4, bottom for examples.

One property of loops is that the line crosses over itself. In a CS, these intersections have a clear meaning, that both series have returned to a value from a the same previous point in time. DALCs also have intersections, but they are only meaningful if the units and scales of the two vertical axes are the same.

2.3 Features Between Pairs of Points

While many users are familiar with similar patterns in line charts, few have been taught or have explored the same patterns in the connected scatterplot. The following patterns demonstrate a number of conditions that a particular time segment may show:

No change. Individual points are mapped purely by the values of the two time series at a given point in time. This has the consequence that consecutive time points with the same values coincide on the CS (Figure 6A).

Only one series changes. When one series does not change, distance and direction of the line between consecutive points is entirely determined by the other series. The result is a line that is parallel to the axis that changes (Figure 6B), going up or right if the value increases, and down or left if it decreases.

Correlation. When both time series increase and decrease together, they are positively correlated, and the resulting line in the CS is parallel to the bottom-left to top-right diagonal (Figure 6C). When they are negatively correlated and move in opposite directions, the CS follows the opposite diagonal (top-left to bottom-right, Figure 6D).

Directions. Which time series is changing more quickly, and the sign of those changes, determines the angle of the line segment. Eight regions around the origin (illustrated in Figure 5) correspond to the various relationships between the time series.

2.4 Limitations

A connected scatterplot of two time series can only be drawn when the points in time at which they are sampled are the same, or largely the same. This is similar to the scatterplot, which can only be drawn for data sets that share a criterion that identifies values on different dimensions as belonging to the same point. This does not present a problem in most journalism scenarios, because their data is typically reported at certain fixed intervals (monthly, quarterly, etc.), such that points always coincide in time (though there may be gaps). If the series don't have the same sampling, the data can be interpolated and resampled before being visualized.

Another limitation is that the CS can create extremely complex shapes that can be impossible to read. This has been an issue in some studies that have looked at similar techniques (such as Robertson et al.'s work on gapminder [39] and Rind et al.'s TimeRider [38]). Since we are interested in presentation, we find this to be less problematic. A journalist will try the technique on data and then decide if the resulting graph is sufficiently interesting, readable, etc. for publication. If not, other options are available, such as the dual-axis line chart or small multiples.

3 Study 1A: Qualitatively Understanding the CS

Due to the lack of familiarity with the technique among the general population, the journalists that we interviewed presumed that readers

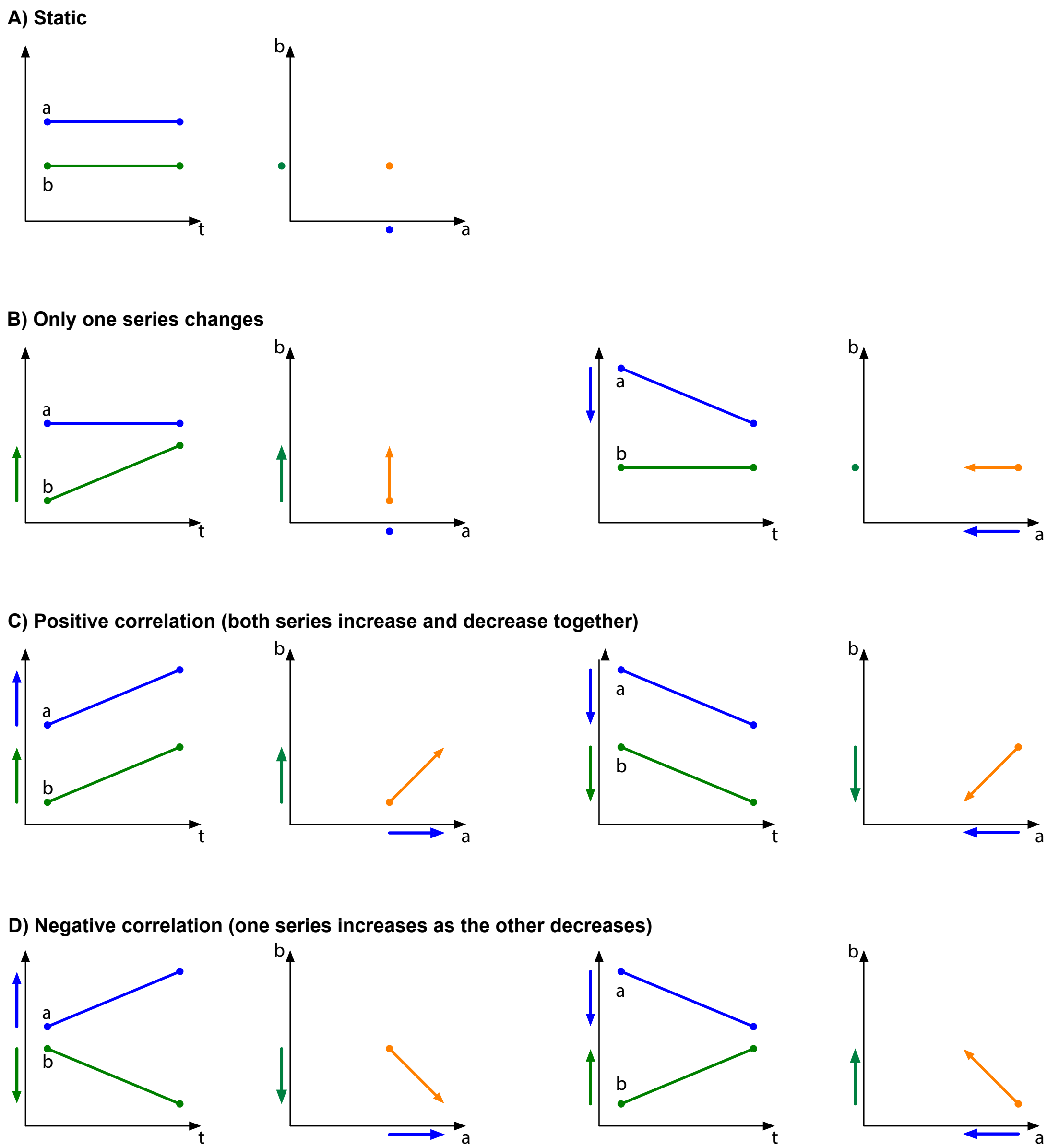


Fig. 6. A sampling of cases showing the same values in dual-axis line charts and connected scatterplots.

would look closer at the charts, however none expected people to be able to immediately understand them. Connected Scatterplots violate many of the usual charting conventions people are accustomed to, such as shifting the representation of time onto the line connecting the dots and using the x-axis to represent one of the variables. Nevertheless, the journalists believed that despite initial difficulties, connected scatterplots should be understandable with only a small amount of instruction, which could be aided by annotations.

Study 1 tested how this less familiar format might affect the ability of a set of participants (college students) to understand and interpret the underlying data. It consisted of two related parts, one that asked participants to explain what they were seeing, and one that had them predict what either a CS or DALC would show given certain patterns. We describe the two parts of this study separately in this section and the next.

3.1 Materials and Procedure

We presented 14 participants with a series of questions about two datasets extracted from a news story, *Driving Safety* [15], as well as from a chart redesign on a blog, *Army* [7] (Figure 7). Since this was a qualitative study that relies on informal interviews with participants, we conducted it in a lab setting using printed pages. Participants were undergraduate students at a research university.

Each participant saw both of these datasets, with the *Driving Safety* example first. Seven participants saw the first example as a dual-axis line chart and the second as a connected scatterplot, and seven participants saw the reverse pattern. The analysis below collapses across the ordering differences between these datasets, focusing on the contrast in responses depending on the graphical format.

For each dataset, participants were first presented with a set of 'qualitative' questions. The first six questions were open-ended, asking participants to describe their initial understanding of the graph, any visual patterns they noticed, the general 'shape' of the graph, the total change in Y1 (Y1 refers to *Auto Fatalities Per 100K People* for the first dataset and *Army Budget* for the second), the total change in Y2 (Y2 refers to *Miles Driven Per Capita* for the first dataset and *Number of Troops* for the second), and the relationship between Y1 and Y2.

These questions were followed by a set of seven iterations of the question "Describe the relationship between Y1 and Y2 in the highlighted region", with periods of 2-10 years highlighted with a yellow

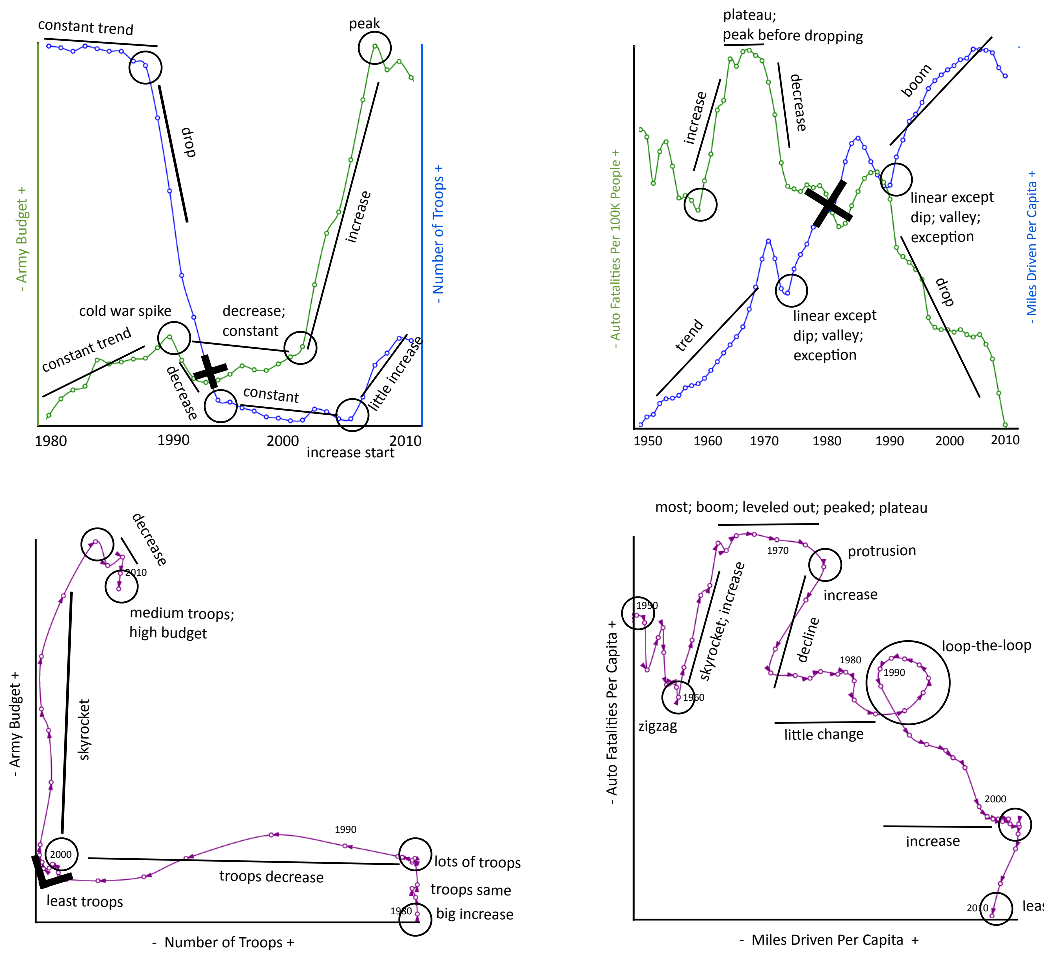


Fig. 7. Features described by participants in the understanding study (Section 3). DALCs in the top row, CSs on the bottom. The two charts on the left show the *Army* dataset, while the charts on the right show *Driving Safety*. Note the different features pointed out in the DALCs (e.g., intersections denoted by X) vs. the CSs (e.g., right angles and loops) .

translucent rectangle (see Figure 11 for a similar form of highlighting). These periods were chosen to reflect a diversity of possible trends in the data, such as positive relationships, negative relationships, and change in one variable but not the other. For the last of these questions in the *Driving Safety* example, and the last two from the *Army* example, the question was replaced with a more contextualized version, e.g., "American cars get bigger, faster, and more deadly, all the while becoming more popular. Does the highlighted region reflect this statement?" The participant should respond "Yes" to this if the chosen region contained both an increase in fatalities and an increase in miles driven per capita. One contextual question in the *Driving Safety* dataset was later determined to have multiple possible answers, and was removed from the analysis.

The complete experiment (including both this part and the one described in the next section) lasted approximately 35 minutes, and the responses of each participant were transcribed by the experimenter or the participant. Subjects were paid $10 for their participation.

3.2 Results

Questions. Despite the novelty of the connected scatterplot format, our set of college students performed at near-ceiling accuracy in their open-ended descriptions for both chart types. When asked to describe their understanding of the graph, note visual patterns, total changes in one measure, or the relation between the two measures, participant responses reflected a high ability to determine relative increases and decreases separately for each measure. While the level of detail and added context (e.g., relevant historical events) varied greatly, a typical response to the question of "what is your initial understanding?" was "The amount of miles driven by people has dramatically increased, auto fatalities has decreased."

For the questions of type "Describe the relationship between Y1 and Y2 in the highlighted region", participants did make a handful of errors (3 total), and all were for connected scatterplots. Two of these errors seem to reflect inappropriate reliance on data interpretation habits learned from more traditional graph formats. One participant noted that the segment of the *Driving Safety* data from the 1990's (a line sloping downward and to the right) was "the reverse" of the 1960's (a line sloping upward and to the right) across both measures, even though the reversal was only for the fatalities measure, not the miles driven measure. In a traditional graph with time on the x-axis, this inference would have been valid for such a mirror flip across the horizontal meridian. But in a connected scatterplot, reverses on both measures requires a 180-degree rotation of line orientation. Another error was found in a highlighted section around 1990 in the *Army* dataset. The participant noted that "The number of troops and the defense budget were both very minimal," even though the number of troops was relatively high. This error may also reflect a habit of drawing inferences primarily from y-values (which were minimal in this case), but not x-values (which were not in this case). A third error was found in a participant observation that "auto fatalities significantly decreased" in the 1960's, when they actually increased. This error may reflect an accidental flip of the polarity of the y-axis.

Descriptions. The way that data is depicted in a graph can drastically impact which patterns participants notice and the types of conclusions that they draw [43]. To evaluate such differences across the two graphical formats, we examined the types of descriptions produced by participants. Figure 7 depicts typical parts of each graph format picked out by participants, along with key phrases used to describe those parts.

Participants noted several salient visual features of each graph format, particularly for the questions that explicitly asked about visual patterns and the shape of the graph. Many participants commented that the graphs for the DALC examples both had a global X-shape (an intersection), and that the lines for each measure would at times "converge" and "diverge." For the CS format, participants frequently noted the L-shape in the *Army* example, with one participant noting, "Right angle? Weird graph". Participants noted two other features of the connected scatterplot format that they found to be particularly unusual. In the *Driving Safety* example, which contained a loop in the graph, participant responses (separated by semicolons) included "I don't know what's going on with the loop-the-loop there; I notice a loop… I've never seen that before… I don't know how that reads… it is confusing; quite erratic, it crosses itself at one point which is uncommon; erratic." In the *Army* example, where time runs from right to left across the first half of the graph, participant responses included, "The graph goes from right to left when plotting time; It is not really a line graph because the times are not chronological; It is going reverse chronologically."

A post-hoc review of the verbal descriptions suggests that these differences in visual format have the potential to lead participants to different conclusions. For example, there were trends toward differences among the types of metaphors used by participants for subsets of data [52]. DALC formats produced many descriptions related to mountainous terrain, with data patterns going *downhill*; *falling off*; *mountain ranges with a high peak*; *peak* (3), *plateau* (2), *steep rate or drops* (5), *uphill*, *valley with two cliffs*, while CS formats only had five examples of such terms. CS formats, in contrast, have relatively more examples of superlative descriptions of trends, due primarily to the strong horizontal and vertical lines in the *Army* example: *huge/sharp increase/leap* (4), *roller-coaster*, *skyrockets* (3), *stagnant* (2), *steadiness*, *takes a turn (at the L-junction)*.

Another post-hoc examination revealed a consistent difference in the use of terminology related to correlation. DALC formats produced terms such as *converging*, *diverging* (2), *correlated* (2), *direct relationship* (2), *directly proportional* (2), *exact opposite*, *inverse proportional*, *inverse relationship* (5), *inversely proportionate*, *inversely related*, *negative exponential*, *negatively correlated* (4), *opposing/opposite trends* (5) far more often than the mere handful of such terms mentioned for CS formats: *direct relationship*, *inverse relationship*, *close to a linear relationship*.

3.3 Discussion

In summary, both DALC and CS formats allowed high performance on objective measures of understanding in our college student population. The CS format produced two mistakes among the qualitative questions, where participants appeared to have relied inappropriately on conventions from the better-known DALC format: (1) that opposite trends tend to reflect across a horizontal meridian, unlike the 180 degree rotations in a CS, and (2) a low value on the y-axis does not mean that all values are low – a CS reader must also inspect the x-axis value.

Qualitative responses showed that participants found loops and possible right-to-left ordering to be a surprising and salient feature in the CS format, with the loop provoking substantial uncertainty across participants. There was a greater prevalence of mountainous metaphors for DALC formats, and a greater prevalence of strong descriptions ('skyrockets') of long linear sequences in the CS format (primarily from the *Army* example). The strongest trend was the extreme difference in the use of correlational language. Although this analysis was post-hoc, the strong trend – 28 examples for the DALC format, and only 3 for the CS format – warrants further evaluation. The long-term experience that participants have with DALC formats may cue them to recognize familiar patterns – parallel lines suggest a positive correlation, while an X-shape suggests a negative correlation [43]. CS formats are not likely to have these associations between correlation types, and particular diagonal orientations of a single line – and such associations are a core requirement for complex thinking [21, 33]. Without learning these associations, CS readers may have more difficulty in drawing more sophisticated inferences from the presented data.

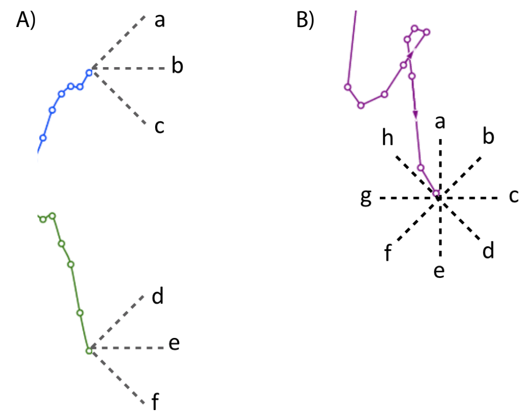


Fig. 8. Participants were asked to indicate the direction of the line using letters in Study 1b (Section 4). They had the choice between the possible principal directions in both the DALC (A) and the CS (B).

4 Study 1b: From Words to Lines

In the second part of Study 1, participants had to translate qualitative statements about the data into a prediction of the next step in either a DALC or a CS.

4.1 Materials and Procedure

The materials used in part b were similar to those in a. However, we slightly modified the datasets, e.g., by shifting some points in the *Driving Safety* dataset in order to ensure that the connected scatterplot contained line segments in each of the eight possible cardinal directions (Figure 8B).

Participants were presented with a set of 'quantitative' statements of the form "Y1 increases, and Y2 increases", and were asked to show which way the line(s) on the graph should move to be consistent with the statement. Figure 8 depicts the response selection, for dual-axis line graphs and connected scatterplots. There were nine such questions, consisting of the three possible states of each variable (increase, decrease, no change) times the three possibilities for the other variable. A second set of eight such questions (skipping the condition where both series do not change) repeated the process for questions of the more 'contextualized' style, e.g., "Oil Embargo: people drive less. They also drive more slowly, leading to a drop in fatalities", requiring participants to take a small inferential step before determining the appropriate changes in the graph.

4.2 Results

When participants were asked to predict what the next section of a graph should look like based on statements such as, "Y1 increases, and Y2 is constant", performance was quite high. Participants scored 13/14 on average for these questions, and performance only dropped slightly, to 11.8/14 for the questions that added context. Splitting the data according to the type of graph, performance on the dual-axis line chart version of questions (6.5/7) was numerically only slightly better than performance on the connected scatterplot version of questions (6.1/7).

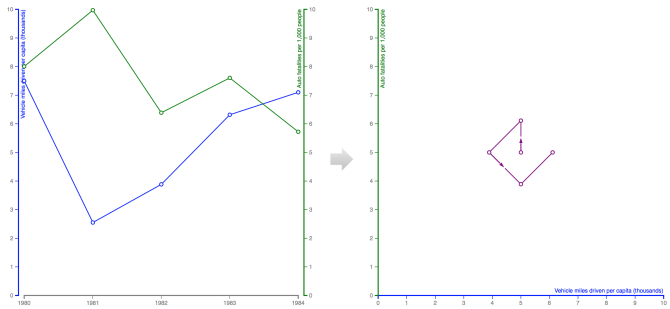


Fig. 9. Initial configuration of the translation study, in DALC-to-CS mode for the driving safety dataset. The data is shown on the left, the user moves each point on the right to match. The initial layout of right graph's points were a randomly rotated and mirrored variation of the example shown.

One trend that appeared was that participants had trouble with backward diagonals in connected scatterplots, relative to the equivalent pattern in dual-axis line charts. When the correct answer was *h* according to Figure 8B, collective participant errors (out of 28 trials) included a *b* response three times, and *f* once, which may reflect less experience dealing with the types of relationships depicted by diagonal lines within this format. Participants only made one collective error (out of 28 trials) for the equivalent dual-axis line chart trial, choosing *c* and *d* instead of *a* and *f*, respectively, reflecting a swap of the trends across the two measures.

For the CS format, when asked to plot the next step when neither variable changed, almost all participants were initially confused. The correct answer is that the line does not move, and that multiple points overlap. Most participants could tell that none of the directional options were appropriate, but typically were not sure of what the correct answer should be. We take this result only suggestively, because the answer choices did not include a center option reflecting this possibility, so it is possible that the design of the response mechanism contributed toward this confusion.

4.3 Discussion

Despite the lack of familiarity, there were few errors for the connected scatterplot, involving leftward ("backward") trends, and some possible confusion surrounding the lack of change on the graph when neither dataset changed. Our participants displayed an ability to understand the quantitative patterns of the connected scatterplot.

Because we did not anticipate high performance for connected scatterplots among our participants, we omitted the no-change question for that condition. We felt that the visual response option (a letter at the center of the choices) would be too confusing. In hindsight, we suspect that this condition may have been the only one that would have shown high error rates, and we plan to test this in future work.

5 Study 2: Axis Direction and Sequence

The interviewed journalists generally agreed that the sequence in connected scatterplots should progress from left to right, though that was considered much more important in print and other static graphics, compared to the web, where interaction and animation might help guide viewers toward a non-conventional reading direction.

Most of our studied examples from the news media arrange variables across the axes such that the time lines flow generally in a left-to-right progression, but this is by no means necessary. The left-to-right progression is consistent with literature that suggests that visual exploration tend to start at the left before moving right [11], and is a dominant direction for imagining a sequence unfolding over time [8], including in graphical representations. One root of such biases may be an individual's learned reading direction [41], at least in the West.

This experiment aims to measure the potential directional confusion by asking subjects to re-imagine how a graph in one format would look when translated to another. We asked a group of participants to take data depicted in a DALC and replot it as a CS, and vice versa. As a control condition, we also asked the participants to copy DALCs to DALCs and CSs to CSs to obtain a baseline level of error for replotting a graph with known coordinates.

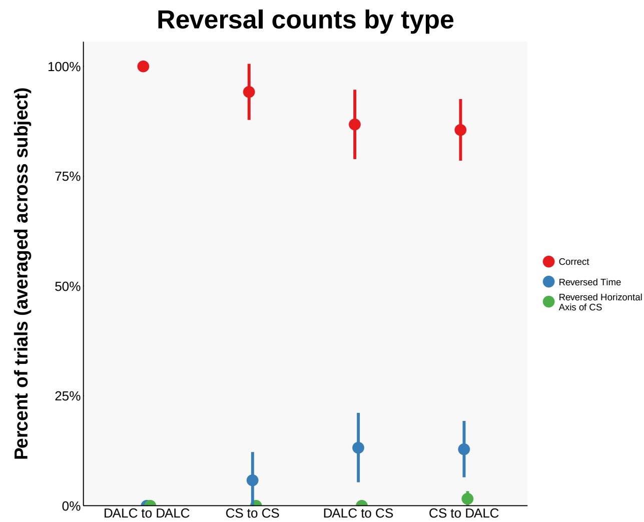


Fig. 10. The percent of each type of errors for each condition. Error bars are 95% CI.

5.1 Materials and Procedure

Study participants were shown two charts next to each other, each either a CS or a DALC. Their task was to transfer the points from the left to the right chart (Figure 9). We presented all four possible combinations, so in half of the cases participants had to translate from DALC to CS or CS to DALC, and in the other half they had to merely copy the data to the same kind of chart. All participants performed all four types of task, allowing within-participant comparison of results. The task type was blocked, with the order of blocks randomized between subjects and the order within each block also randomized. Participants each saw 4 tasks $\times$ 7 repetitions = 28 different datasets. Each consisted of five data points, with their shapes based loosely on examples abstracted from news graphics or constructed to mimic certain features such as loops.

35 participants were recruited on Amazon's Mechanical Turk platform [24] using the default worker requirements. Mechanical Turk ID numbers were recorded, so no one could participate more than once. They first clicked through a short tutorial that showed the correspondence between a DALC and a CS in five consecutive steps. Participants were paid $5 to perform the study, which took up to 45 minutes to complete. We note that while our other reported experiments paid more than minimum wage based on actual completion times, our underestimate regrettably caused this study to pay less in some cases. The locations of the response points and the response time (RT) were both recorded.

We used performance on the simple copying conditions as a filter to determine if participants were actually attending and performing the task. Consequently, we excluded 8 of the 35 participants from the analysis due to response patterns that were clearly at chance levels (possibly due to lax subject requirements). From the remaining participants, we also discarded 2 total trials that were completed in under 5 seconds (likely the results of an accidental click).

5.2 Results

While participants might make quantitative errors in the exact placement of points along axes, we focused our analyses on the categories of errors that reflect their understanding of the coordinate space of each chart, and how these spaces interact. Using automated analyses of error patterns, we classified each trial into a best match for four classes:

Correct. The translation was qualitatively correct.

Reversed Time. The temporal ordering of the points was reversed.

Reversed Horizontal Axis of CS. The polarity of the horizontal axis (defined by the CS) was reversed.

Reversed Vertical Axis of CS. The polarity of the vertical axis (defined by the CS) was reversed (this error never occurred).

The types of mistakes made by participants can be seen in Figure 10.

Copy conditions. In the more familiar DALC to DALC condition, there were almost no qualitative errors, as expected. In the CS to CS condition, there were no errors involving a reversed horizontal axis. But despite the task consisting of a simple copying operation, participants reversed the temporal ordering of the points on 5% of trials.

Translation. In both DALC to CS and CS to DALC conditions, time was again reversed, but even more frequently (13% in each case). Furthermore, in the CS to DALC condition, participants flipped the horizontal axis 1.6% of the time. It appears that while larger values are expected to be on the top of the vertical axes of a DALC, some participants occasionally read larger values from the left of the horizontal axis of the CS. There was no evidence of this axis reversal in the DALC to CS condition.

Response time. Although the CS to CS condition was about 20% faster than the other conditions, this effect is likely due to the lower number of points needed to be observed and specified. We found no other statistically reliable difference between the response times for the other conditions, nor did we find a significant response time effect from reversals.

5.3 Discussion

Performance on this difficult translation task was high overall (92%), but there were two noteworthy patterns of error: participants often reversed the flow of time when dealing with a CS, particularly when translating back and forth with a DALC, but even when simply copying a CS. And there were a small but significant number of trials where participants assumed that high values in DALC should be placed on the left, instead of right, side of the horizontal axis of the CS.

The occurrence of this type of error raises the question of what design changes – such as using animation or varying size over time – might be able to mitigate this confusion. One study [39] compared animation in scatterplots similar to gapminder [1], with a line that traced the history of each dot. Graphs with these traces were effectively a collection of connected scatterplots. These traces were not found to yield significantly better accuracy than animation. However, there were many traces visible at the same time, leading to clutter.

6 Study 3: Engagement

The journalists that we interviewed claimed that a substantial benefit of connected scatterplots over more traditional formats was added engagement. They especially spoke highly of loops and other unusual shapes that can arise in connected scatterplots, and that these shapes draw in potential readers to more closely examine the chart. Loops were considered a critical component, with one calling them "the delightful part." They generally believed that any initial difficulty would be compensated for by the increased introspection and engagement.

But are these intuitions accurate? We predicted that study participants would be preferentially drawn to more closely inspect CSs because they have two properties known to attract attention and engage a viewer: novelty and challenge [28, 48, 26]. The graphical format is unfamiliar and presents the viewer with a puzzle to be solved.

There are several ways to measure task engagement. For a single task, participants can self-report engagement levels at randomly sampled intervals, or when they realize that they have become more or less engaged in a task [44]. There are also physiological signals that correlate with task engagement and effort, including the activation levels of particular brain regions, electrophysiological reflections of brain responses to mistakes made during a task, cortisol levels, or heart-rate variability [45].

But when seeking to determine which of multiple possible visual images a viewer prefers to engage with, it is often most direct to measure preferential viewing with an eyetracking or user-choice technique, as in the marketing literature [12, 37]. We pit the CS and DALC formats against one another for the attention of study participants by simulating the experience of a reader glancing across the pages of a newspaper or website. We previously used this technique for the similar task of comparing engagement and viewing preference for bar graph styles [22].

6.1 Materials and Procedure

We created six datasets, each including a 2-3 line description and three short annotations highlighting points of interest in the data (included in the supplemental material). The experiment application presented participants with a row of six thumbnails of visualizations of the six unique datasets (Figure 11A), with three represented in DALC format and three in CS format. No one saw both the CS and DALC version of a given dataset, and the order of the datasets was randomized between participants. The images were small enough to relay the overall structure of the visualization, but not the text or other details.

Participants were told that we were studying the types of information that most interested them, and they would not be tested on any of the visualizations. No explanation of the techniques was provided in order to emulate the scenario of naive readers. They were allowed five minutes to explore the set of six visualizations, during which time they could select a thumbnail to view clearly in full screen for as long as they liked (Figure 11B). They could click again to return to the thumbnail view, and select a new image to view.

25 university students participated in this experiment (12 men and 13 women). They were paid $5 each.

6.2 Results

Figure 12 shows what proportion of subjects were viewing CSs or DALCs during each one-second time interval. In the first half of the experiment, subjects spent more time viewing CSs (57.4%, 95% CI [49.3%, 65.6%]). They then shifted their attention to the remaining charts, which were DALCs. This shift is visible in Figure 12 around two minutes.

We did not find a viewing bias in the second half of the experiment (49.2%, 95% CI [40.6%, 57.7%]). However, a post-hoc inspection of the data revealed that most subjects completed viewing all of the charts before the experiment was finished. In the remaining time, many subjects returned to the CSs to view them for a second time, which explains the increase in CS viewing during the last minute.

A linear regression of chart index and time showed a trend of subjects progressing from left to right when selecting which charts to view. A positive slope of .452 (95% CI [.428, .476]) indicates that subjects consistently progressed from charts with a low index (on the left) to charts with a high index (on the right). To confirm that the CS prioritization was not simply an ordering effect, we simulated what would happen if each subject simply progressed from the left to the right and viewed each chart for an equal amount of time. The random ordering actually placed more DALCs on the left than the right 62% of the time, which would have yielded the opposite trend – a prioritization of DALCs – if viewing order were the only factor. Instead, subjects preferentially viewed CSs.

We did not find a reliable difference in total viewing time between the two techniques (CS: 148s, 95% CI [132, 163]; DALC: 130s, 95% CI [114, 146]). This result indicates that although subjects prioritized viewing CSs, they eventually returned to view the DALCs as well.

6.3 Discussion

In an informal interview after the experiment, we asked subjects which of the six charts was most difficult to understand. Although the spe-

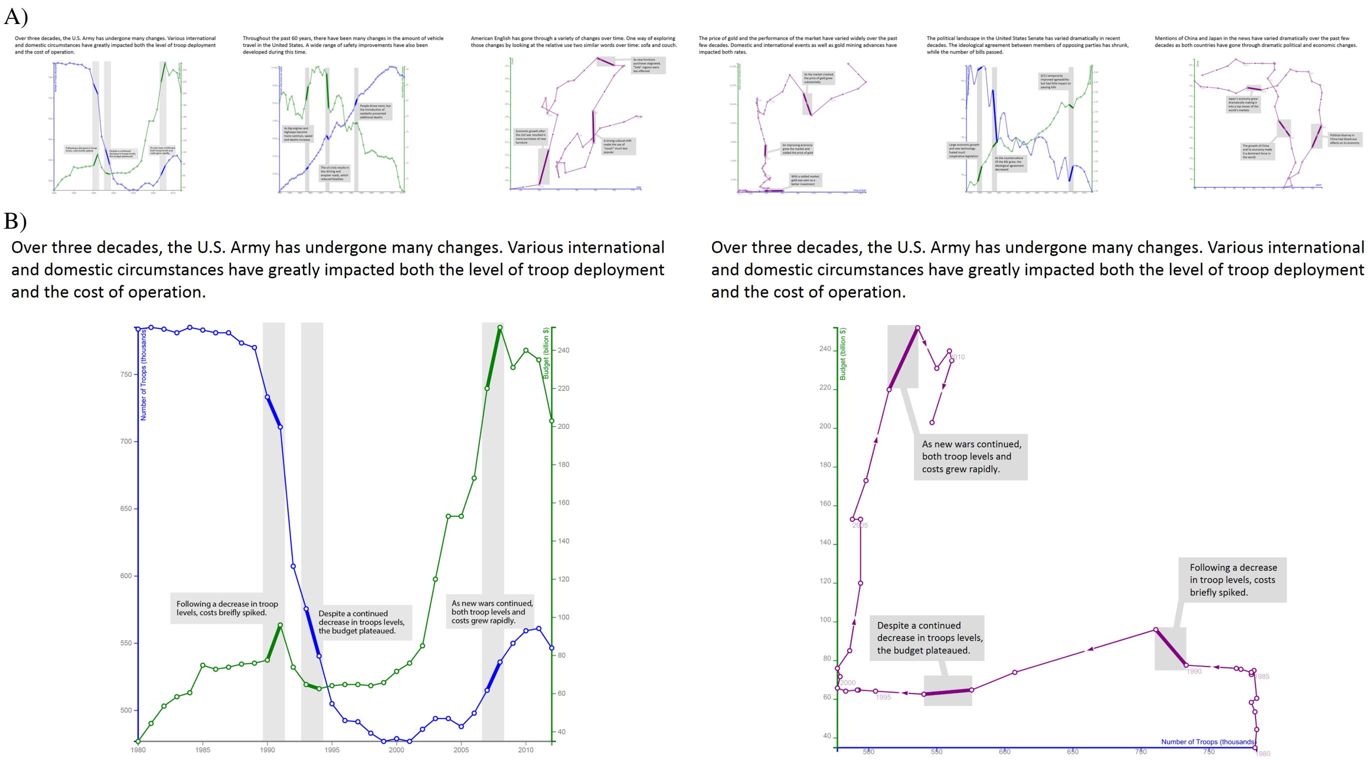


Fig. 11. Engagement study setup. Participants first saw the filmstrip-like display of six randomly ordered images with text descriptions (A). They were free to pick which images to examine in more detail, and we measured the time spent with each. Written descriptions and annotations were equivalent for both versions of the dataset (B).

cific chart chosen varied by subject, the most difficult chart was always a connected scatterplot. One made a general statement that they had a hard time with "all of the loopy ones." Perhaps due to this difficulty, the novelty, the sparseness of having one line instead of two, or the specific combination of angled line segments, CSs were more effective at grabbing a viewer's attention, at least in the initial stages of their opportunity to choose among formats. Viewers did not immediately disregard these new visualizations due to their unfamiliar format. Instead, they appeared to prioritize their initial attention towards them. While it will be important for future work to extend these results in a variety of contexts, our initial study suggests that when a reader sees a small thumbnail of a CS in a news aggregator or website, they may be more likely to engage with that visualization.

7 Guidelines and Questions

Our initial exploration of the understanding of the connected scatterplot reveals some distinct advantages over other techniques, but its use also carries some strong caveats. As this work is only an initial foray into studying user performance with connected scatterplots, we present preliminary guidelines for their use.

Give salient cues for the flow of time. Viewers occasionally made mistakes in understanding the direction of the flow of time indicated by the connecting lines, even in simple copying tasks. Using a left-to-right global flow of time should minimize this error [31, 22], along with explicit annotation of temporal direction. Although we have primarily seen arrows used to annotate temporal direction, this approach does not appear to be sufficient, as time reversals were the dominant error made by readers. Future work should explore the many alternative annotation styles that are possible, including varying line thickness [39], color, or contrast over time.

Give explicit reminders of two oriented axes. Viewers will intuit that large values belong on the top of the vertical axis of a CS, but they may need to be reminded that large values belong on the right of the horizontal axis. They should also be aware that mirror-reversals along the horizontal axis do not indicate full reversals across both sets of values – only reversals along that single axis. Annotation or added embellishments to the horizontal axis may help avoid these misinterpretations.

Use for engagement and communication. The prioritized viewing of CSs – at least as compared to DALCs – makes them good candidates when the goal is to draw a viewer's attention. As the journalism examples show, the distinctive features also lend themselves to annotation and highlighting, further adding to its usefulness as a tool for communicating data. But it is not yet clear whether the preferential viewing arises from the technique per se, or its lack of familiarity. The general public is likely to continue to be less familiar with CSs than with DALCs.

A caveat on correlation. If a major purpose of a graph is to leave the viewer with an understanding of negative or positive correlation in a dataset, a more traditional DALC may make this conclusion far more salient. Our college student participants rarely used correlational language to describe patterns of data in the CS, in striking contrast to the same datasets depicted as DALCs. This is may be a result of low exposure to the visual features that indicate correlation within this chart type. Annotation that highlights these relationships in a CS may help associate its visual features with correlation.

We also identified a number of unanswered questions that can guide further work to increase our understanding of the connected scatterplot technique.

Questions about complexity. The news graphics examples we have studied all have highly unique shapes, and generally only show a small number of identifiable features. Having a small number of unique shapes and features will likely make for a more memorable and understandable graph [16, 23]. Future evaluations should seek the number of salient features that strikes a balance between tolerable complexity and desirable difficulties [26].

Questions for analysis. We focused on the presentation aspect of connected scatterplots because the inherent sequence in the technique lends itself well to narrative visualization and journalism [42]. The CS technique, however, likely has utility for data analytics and explo-

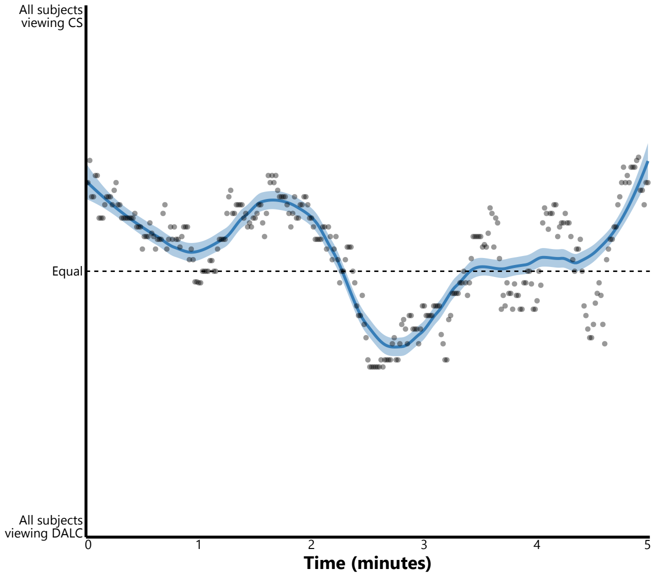


Fig. 12. The duration of the experiment was split into one-second intervals. The vertical axis shows the proportion of subjects viewing a CS or DALC at each point. The beginning of the experiment is time 0, and moving to right progresses through the duration of the experiment. The blue line is a locally weighted regression (LOESS) fit to the points, and the ribbon is the 95% confidence interval across subjects.

ration as well. Despite a variety of proposed techniques for rendering higher dimensional versions of these plots [20], using them to trace the animation history of a plot [38, 39], being a means of interacting with a plot's timeline [29], as well as navigating similarity spaces of graphs [47] or images [4] over time, the conditions in which they help or hurt a user in exploring data remain unclear.

8 Conclusions

The connected scatterplot is not a new technique, but one that is unfamiliar to most viewers and under-explored in the visualization community. With this paper, we hope to introduce it to a wider audience. It seems to be effective for communicating data and engaging viewers, and it shows some structures in data (like time-shifted patterns) in a different way compared with other techniques.

The studies in this paper only scratch the surface, but they also provide some first insights. Both college students (in the in-lab qualitative study) and Mechanical Turk participants (in the translation study) are generally able understand this unusual chart, with the notable exception of occasionally mirrored direction. We also found that study participants/readers were intrigued by the chart's unusual shape and chose to look at it first more often than at standard line charts. All these findings suggest that the technique, despite its lack of familiarity, has merit for presenting and communicating data.

9 Demos and Experiment Materials

An interactive tool for generating connected scatterplots as well as materials for the experiments are available at: `http://steveharoz.com/research/connected_scatterplot`

10 Acknowledgements

The authors wish to thank Evan Applegate (Bloomberg News), Jorge Camoes (ExcelCharts), Amanda Cox (NY Times), Hannah Fairfield (NY Times), Christopher Ingraham (Brookings Institution), and Katie Peek (Popular Science) for supplying the data used in their connected scatterplots, and for answering our questions.

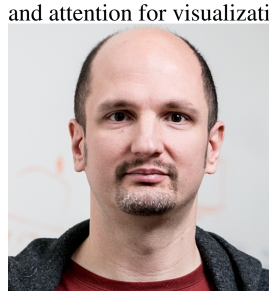
We are grateful to Matt Hong and Kevin Harstein for their assistance in experimental material design and data collection. Analysis and plots were made using the R packages ggplot2 [49], tidyr, and dplyr [50]. This work was supported in part by NSF awards IIS-1162037 and BCS-1162067.

References

[1] Gapminder. `http://gapminder.org/`.
[2] Anonymous. Daily chart: The tracks of arrears. `http://econ.st/1Fguao4`, 2015.
[3] J. Ashkenas, A. Parlapiano, and H. Fairfield. How the u.s. and opec drive oil prices. `http://nyti.ms/1KNjS0t`, 2015.
[4] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568, Jan 2016.
[5] P. Bump. Obama's approval versus the economy. `http://wpo.st/d1_b0`, 2015.
[6] A. Cairo. *The Functional Art*. New Riders Press, 2012.
[7] J. Camoes. Chart redraw: Troops Vs. Cost (Time Magazine). `http://www.excelcharts.com/blog/redraw-troops-vs-cost-time-magazine/`, 2013.
[8] D. Casasanto and L. Boroditsky. Time in the mind: using space to think about time. *Cognition*, 106(2):579–93, Feb. 2008.
[9] A. Cox. Oil's Roller Coaster Ride. `http://nyti.ms/yfZB8X`, 2008.
[10] P. Coy, E. Applegate, and J. Daniel. The Rise of Long-Term Joblessness. `http://www.businessweek.com/articles/2013-02-07/the-rise-of-long-term-joblessness`, 2013.
[11] C. A. Dickinson and H. Intraub. Spatial Asymmetries in Viewing and Remembering Scenes: Consequences of an Attentional Bias? *Attention, Perception, & Psychophysics*, 71(6):1251–1262, 2009.
[12] S. Djamasbi, M. Siegel, and T. Tullis. Generation Y, web design, and eye tracking. *International Journal of Human Computer Studies*, 68(5):307–323, 2010.
[13] J. C. R. Dow and L. A. Dicks-Mireaux. The Excess Demand for Labour. A Study of Conditions in Great Britain, 1946-56. *Oxford Economic Papers*, 10(1):1–33, 1958.
[14] H. Fairfield. Driving shifts into reverse. `http://www.nytimes.com/imagepages/2010/05/02/business/02metrics.html`, 2010.
[15] H. Fairfield. Driving Safety, in Fits and Starts. `http://nyti.ms/PB07e2`, 2012.
[16] S. L. Franconeri. The nature and status of visual resources. *Oxford handbook of cognitive psychology*, 8481:147–162, 2013.
[17] S. Furth. In short-term unemployment data, good and bad news. `http://on.wsj.com/1pPPwDl`, 2014.
[18] D. Gately. Do Oil Markets Work? Is OPEC Dead? *Annual Review of Energy*, 14:95–116, 1989.
[19] T. Giratikanon and A. Parlapiano. Janet L. Yellen, on the Economy's Twists and Turns. `http://nyti.ms/19jPM2o`, 2013.
[20] S. Grottel, J. Heinrich, D. Weiskopf, and S. Gumhold. Visual Analysis of Trajectories in Multi-Dimensional State Spaces. *Computer Graphics Forum*, 33(6):310–321, 2014.
[21] G. S. Halford, N. Cowan, and G. Andrews. Separating Cognitive Capacity from Knowledge: A New Hypothesis. *TRENDS in Cognitive Sciences*, 11(6):236–242, 2007.
[22] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1191–1200. ACM, 2015.
[23] S. Haroz and D. Whitney. How Capacity Limits of Attention Influence Information Visualization Effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2402–2410, Dec. 2012.
[24] J. Heer and M. Bostock. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings CHI*, pages 203–212, 2010.
[25] W. Hickey. The m. night shyamalan twist. `http://53eig.ht/1i1jDon`, 2015.
[26] J. Hullman, E. Adar, and P. Shah. Benefitting InfoVis with Visual Difficulties. *Transactions on Visualization and Computer Graphics*, 17(12):2213–2222, 2011.
[27] C. Ingraham. Graduation, marijuana use rates climb in tandem. `http://wapo.st/XNfDOe`, 2014.
[28] W. A. Johnston, K. J. Hawley, and J. M. Farnham. Novel popout: Empirical boundaries and tentative theory. *Journal of Experimental Psychology:*

Human Perception and Performance, 19(1):140–153, 1993.
[29] B. Kondo and C. M. Collins. Dimpvis: Exploring time-varying information visualizations by direct manipulation. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2003–2012, 2014.
[30] D. Mancino. Il giocattolo si è rotto. http://www.wired.it/attualita/2014/06/18/nord-sud-150-anni-differenze-italiane/, 2014.
[31] A. L. Michal and S. L. Franconeri. The order of attentional shifts determines what visual relations we extract. *Journal of Vision*, 14(10):1033–1033, 2014.
[32] D. P. Miller. The mysterious case of james watt's ' "1785" steam indicator': Forgery or folklore in the history of an instrument? *International Journal for the History of Engineering & Technology*, 81(1):129–150, 2011.
[33] G. A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 101(2):343–352, 1994.
[34] M. Murray and C. Szymanski. The Fed's Balancing Act. http://www.reuters.com/investigates/graphics/fed/, 2014.
[35] R. Olson. Wealth and height in the netherlands, 1820-2013. http://www.randalolson.com/2014/06/23/why-the-dutch-are-so-tall/, 2015.
[36] K. Peek. Helium Supply. *Popular Science*, page 36, August 2013.
[37] R. Pieters, M. Wedel, and R. Batra. The Stopping Power of Advertising: Measures and Effects of Visual Complexity, 2010.
[38] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, F. Drexler, B. Neubauer, and N. Suchy. Visually Exploring Multivariate Trends in Patient Cohorts Using Animated Scatter Plots. In *Ergonomics and Health Aspects of Work with Computers*, volume 6779 LNCS, pages 139–148. 2011.
[39] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of Animation in Trend Visualization. *Transactions on Visualization and Computer Graphics*, 14(6):1325–1332, 2008.
[40] P. Rodenburg. The remarkable transformation of the UV curve in economic theory. *The European Journal of the History of Economic Thought*, 18(1):125–153, Feb. 2011.
[41] A. Román, A. El Fathi, and J. Santiago. Spatial biases in understanding descriptions of static scenes: the role of reading and writing direction. *Memory & cognition*, 41(4):588–99, May 2013.
[42] E. Segel and J. Heer. Narrative Visualization: Telling Stories with Data. *Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.
[43] P. Shah and J. Hoeffner. Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, 14(1):47–69, 2002.
[44] J. Smallwood, J. B. Davies, D. Heim, F. Finnigan, M. Sudberry, R. O'Connor, and M. Obonsawin. Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, 13(4):657–690, 2004.
[45] M. Tops, M. a. S. Boksem, A. E. Wester, M. M. Lorist, and T. F. Meijman. Task engagement and the relationships between the error-related negativity, agreeableness, behavioral shame proneness and cortisol. *Psychoneuroendocrinology*, 31(7):847–858, 2006.
[46] A. Tribou, D. Ingold, and J. Diamond. Holdouts Find Cheapest Super Bowl Tickets Late in the Game. http://www.bloomberg.com/infographics/2014-01-16/tracking-super-bowl-ticket-prices.html, 2014.
[47] S. van den Elzen, D. Holten, J. Blaas, and J. van Wijk. Reducing snapshots to points: A visual analytics approach to dynamic network exploration. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):1–10, Jan 2016.
[48] Q. Wang, P. Cavanagh, and M. Green. Familiarity and pop-out in visual search. *Perception & psychophysics*, 56(5):495–500, 1994.
[49] H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media, 2009.
[50] H. Wickham. Tidy data. *Journal of Statistical Software*, 59(10), 2014.
[51] J. Wolfers. Wage Growth Is No Longer as Sensitive to Labor Market Conditions. http://nyti.ms/1rThNrM, 2014.
[52] C. Ziemkiewicz and R. Kosara. The Shaping of Information by Visual Metaphors. *Transactions on Visualization and Computer Graphics*, 14(6):1269–1276, 2008.


Steve Haroz researches how our brain perceives and comprehends visual information. Steve is a postdocoral fellow in the Psychology Department at Northwestern University. He received his PhD from University of California, Davis on perception and attention for visualization.


Robert Kosara is Research Scientist at Tableau Software, where he focuses on visualization for data presentation and storytelling. He was formerly Associate Professor at UNC Charlotte. Robert also runs the visualization blog https://eagereyes.org/.


Steve Franconeri is a Professor of Psychology at Northwestern University, and Director of the Northwestern Cognitive Science Program. He studies visuospatial thinking and visual communication, across psychology, education, and information visualization.