Data-Driven Mark Orientation for Trend Estimation in Scatterplots

Tingting Liu, Xiaotong Li, Chen Bao
Shandong University, Qingdao, China
sduhammer,xiaotonglig,baochen95@gmail.com

Michael Correll
Tableau Software, United States
mcorrell@tableau.com

Changhe Tu
Shandong University, Qingdao, China
chtu@sdu.edu.cn

Oliver Deussen
University of Konstanz, Germany
oliver.deussen@uni-konstanz.de

Yunhai Wang*
Shandong University, Qingdao, China
cloudseawang@gmail.com

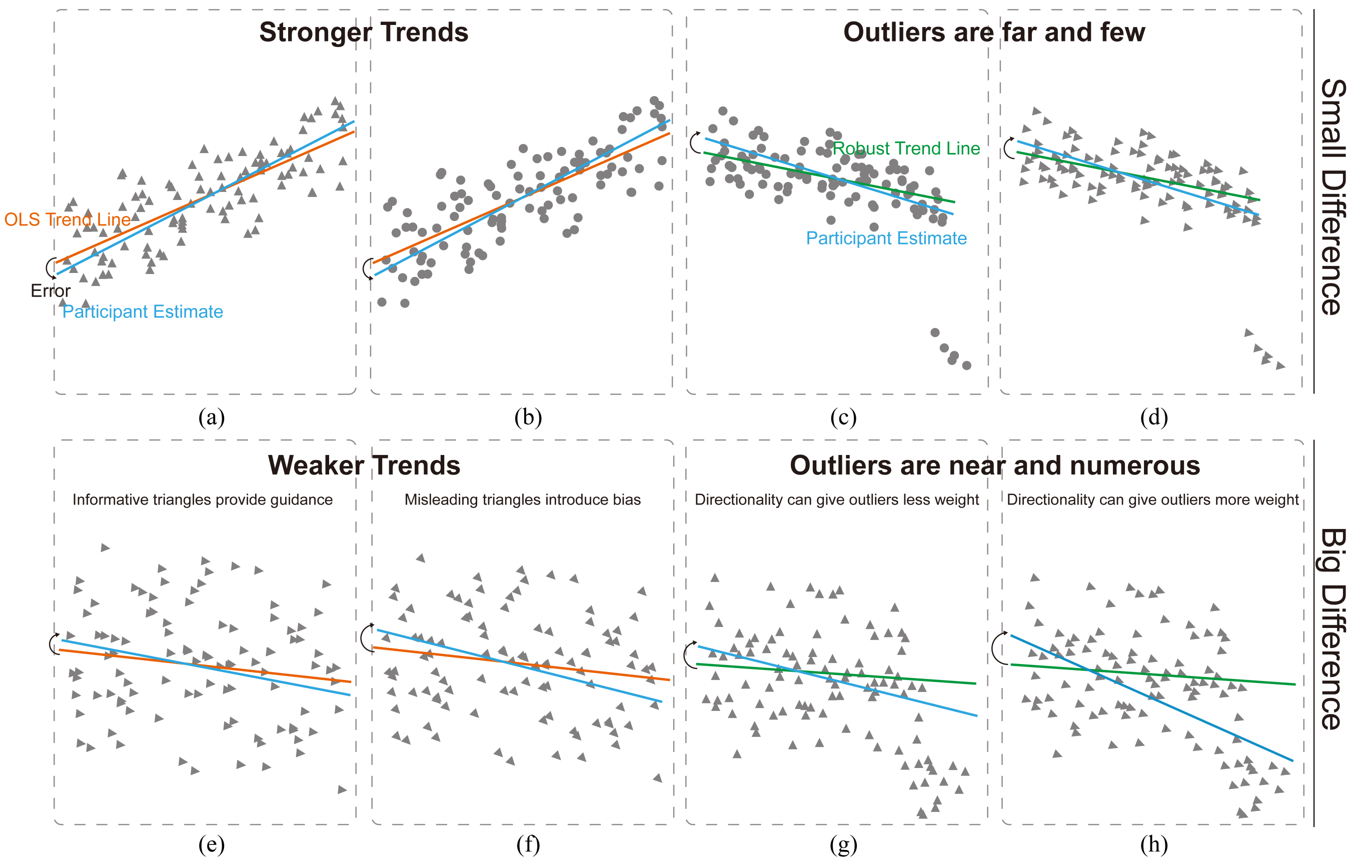


Figure 1: The impact of mark orientation on trend estimation. When a trend in the data is unambiguous, either due to a strong correlation (a,b) or the presence of an obviously outlying cluster irrelevant to the central trend (c,d) mark orientation does not greatly impact the perception of trends in the data. However, when trends are ambiguous, mark orientation can introduce a bias. By using a *data-driven orientation* of marks in such cases, we can guide viewers to more accurate estimates of weak trends (e,f), or trends impacted by nearby outliers (g,h). This results in estimates that are robust even for cases when assumptions about an underlying linear modeling are violated.

*corresponding author

CHI '21, May 08–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-XXXX-X/18/06. . . $15.00
https://doi.org/10.1145/1122445.1122456

ABSTRACT

A common task for scatterplots is communicating trends in bivariate data. However, the ability of people to visually estimate these trends is under-explored, especially when the data violate assumptions required for common statistical models, or visual trend estimates are in conflict with statistical ones. In such cases, designers may need to intervene and de-bias these estimations, or otherwise inform viewers about differences between statistical and visual trend estimations. We propose *data-driven mark orientation* as a solution in such cases, where the directionality of marks in the scatterplot guide participants when visual estimation is otherwise unclear or ambiguous. Through a set of laboratory studies, we investigate trend estimation across a variety of data distributions and mark directionalities, and find that data-driven mark orientation can help resolve ambiguities in visual trend estimates.

1 INTRODUCTION

We rely on scatterplots for a variety of bi-variate tasks including the perception of trends, correlations, and outliers [22]. Famous examples such as Anscombe's quartet [1] show that the *visual* representation of bi-variate data can reveal patterns that would ordinarily be hidden by common summary statistics. Likewise, common statistical trend measures as Ordinary Least Squares (OLS) can be simplistic and fail to accurately model data in the presence of outliers, multiple clusters, or heteroskedasticity (unequal variance) [19]. However, it is highly unlikely that the perceptual process by which people assess trends in scatterplots is anchored in a viewer's literal statistical calculations: recent evidence instead suggests that perceptual "proxies" [14, 18] based on one or more visual features of a scatterplot [36] drive our assessment of these statistical properties.

This results in uncertainty about how to make use of visual perception for trends estimation in scatterplots. On the one hand, we might wish to use the *power and flexibility* of human judgments to disambiguate different model spaces or select more reasonable fits when the assumptions of particular models are violated. On the other hand, we might want to provide *guidance and feedback* for viewers when they face possible human biases or misinterpretations of trends. Striking the right balance between trust in human visual judgments and assistance from statistical models requires a baseline comparison of human perception and statistical models: In which cases can we rely on humans to make sufficiently accurate judgments about trends, when might authors or designers need to intervene?

Initial work from Correll & Heer [7] suggests that people are able to perform a general sort of "regression by eye" with simple Gaussian stimuli and a single class of outliers. However, it is unknown whether this sort of visual estimation is reliable in more complex scenarios, or if there are strategies by which a visualization designer can intervene in order to promote better estimates. After studying corresponding cases, we propose the use of **data-driven mark orientations**, where marks in a scatterplot are oriented into the direction of the trend, as a potential design choice that relies on regression by eye when the trend in the data is unambiguous and strong, but provides useful guidance when the trend is ambiguous or weak. Unlike a mere superposition of a trend line on top of a scatterplot, such data-driven marks are a kind of weaker "suggestion" for a trend direction, and can be ignored more easily when unnecessary or inappropriate.

We performed three experiments to justify the use of mark orientation for encoding trend information. Our research questions for these experiments were: 1) to what degree does the orientation of marks in a scatterplot create a bias for regression by eye? and 2) in cases where traditional linear models like ordinary least squares (OLS) under-perform (such as distributions with heteroskedasticity or outliers), can a bias introduced by a carefully selected mark orientation result in estimates that are more robust than OLS?

Our experiments show that, while in most cases viewers can accurately estimate trends regardless of mark shapes, the use of data-driven marks with a directionality that visually reinforces a trend direction can assist viewers in trend estimation. This assistance is particularly prominent when the trend is unclear (as in scatterplots with large amounts of dispersion) or when the trend is ambiguous (for outlying or secondary clusters that would shift the OLS line of best fit). This suggests that the ability to visually estimate trends, even in the absence of an explicit line of best fit, is robust even in cases when methods like OLS do not perform well. In cases where the trend is uncertain or unclear, the use of data-driven orientation as a supplemental encoding can visually reinforce trend lines and reduce perceptual errors. Data tables, study materials and additional analyses are included in our supplemental material and can also be found at https://osf.io/hzdf3.

2 RELATED WORKS

There is an extensive body of works on graphical perception tasks in scatterplots. Rensink [20] states that, much as geneticists use model organisms with simple genomes to model more complex behavior, scatterplots can function as "fruit flies" for graphical perception research questions in visualization. Scatterplots become especially important when moving beyond atomic tasks such as extraction and comparison of individual values that are common in foundational graphical perception work such as Cleveland & McGill [5] to "ensemble" tasks [27] such as the perception of average value and variance. The assessment of a trend is such an ensemble task, potentially requiring viewers to consider every point in the scatterplot in a holistic way, although recent work [14, 18] suggests that "perceptual proxies" such as point envelopes and ranges likely guide visual statistical judgments in scatterplots rather than explicit calculations.

Much existing work focuses on the perception of correlations in scatterplots [12, 15, 21, 25, 32, 36], often involving pairwise comparisons on which of two scatterplots contain data that is more highly correlated. The perception of a particular trend, however, is comparatively under-explored. Correll & Heer [7] find that human estimations of trends in simple gaussian stimuli are largely similar to the results of methods like OLS, except for biases introduced by the design of a visualization or the presence of outlying points. We reanalyze their results and discuss modifications of their experimental methods in the following section

The choice of mark shapes in scatterplots is important and has great impact on the discriminability of clusters [16, 29] as well as the legibility of other encoding channels such as color [26]. Furthermore, there seems to be a qualitative difference between closed and open shapes [4] for the assessments of similarity and difference within classes [8]. Despite this potential impact, many (or even most) visualizations rely on default mark shapes [2], typically circles. Of particular interest to us is the work by Ziemkiewicz & Kosara on the "implied dynamics" [37] of visualizations. Their work on marks in scatterplots specifically suggests that "attractor" points [38] can create a bias for results of graphical perception tasks (even to the extent of irrational decision-making, as pointed out by Dimara et al. [9, 10]).

Xiong et al. [33] show that this sort of "perceptual pull" can influence ensemble tasks. Relying on this impact, Lu et al. [17] use data-driven "winglets" around marks in scatterplots to suggest implicit clusters. Our assumption therefore is that the directionality inherent in marks such as triangles could act as a form of implied dynamics that subtly influences the estimation of a trend, and therefore act as a way of mitigating potential biases (see Wall et al. for a discussion of the benefits and costs of bias mitigation techniques in visualization [31]). Similar to winglets, we use orientation to implicitly communicate structures in scatterplots. However, we believe our work aims at solving a more general and common problem (global trend estimation versus cluster boundary identification under uncertainty) in a more transferable way (a global, trivially computed orientation as compared to a more complex per-mark calculation involving uncertainty quantification).

We base our experimental methods on those employed by Correll & Heer [7]. The authors created scatterplots based on an initial set of Gaussian residuals of varying bandwidths (corresponding to an increasing level of dispersion) to which a particular trend was added, allowing orthogonal control over the slope of the trend line and the dispersion of the points (c.f. Rensink & Balridge [21] and Harrison et al. [12] in which correlation and slope are entangled). Participants then manipulated a trend line with a slider until they felt that it matched the trend in the data, with the error measured as the difference (both absolute and signed) in the user defined trend and the actual source trend. We extend and modify their methods to address the specific case of marks with inherent orientations and more complex data distributions.

3 METHODS

Expanding and extending the work of Correll & Heer [7], we performed a set of three in-person experiments focusing on two open points: the effect of *mark orientation* and the effect of the *internal distribution* of the data points such as outliers and non-uniform densities on the visual estimation of trends in scatterplots, with an explicit interest in the *interaction* between mark type and scatterplots with complex or ambiguous trends. With the belief that certain mark orientations could guide (or mislead) the viewer, our experiments investigated the accuracy or precision of trends that were estimated using regression by eye for various scatterplots that meet the OLS assumptions (Experiment 1) or violate them due to outliers (Experiment 2) or non-uniform data densities (Experiment 3).

We explore pre-defined as well as *data-driven* mark orientations, which are determined by analyzing trends in the data. When the internal distributions of the data violate the assumptions of the OLS model, we expected to find *systematic differences* between model and visual estimations. We focused on linear trends in this experiment for the ease of modeling but also because there are general similarities in human performance between linear and non-linear trends as observed by Correll & Heer [7].

We chose synthetic stimuli to be able to compare our work with prior approaches such as [7, 12, 18, 21], which in most cases use synthetic stimuli. A second reason was to have precise control over the features in the scatterplots. This allows us to directly assess our research questions (which primarily dealt with the robustness and feasibility of data-driven marks under specific conditions).

Mark Shape and Orientation. While there exists a vast palette of potential mark shape options for which we would anticipate differing patterns of perceptual judgments [8], we focus on triangles and circles, which are the most widely used marks. For example, over 97% of all single-class scatterplots in the VizNet [13] dataset use one of those two mark types. Besides a default upward oriented orientation, we test triangles rotated using data-driven factors, with circles as a default option without inherent orientation. We also found that most triangle marks in VizNet [13] dataset are equilateral triangles; thus, while isosceles triangles have less ambiguity in their encoding of orientation, we use equilateral triangles. Additionally, unlike isosceles triangles, the exact center of equilateral triangles is easier to estimate, which might have benefits for distributional estimation tasks.

Although other shapes can also be used such as diamonds, squares or even arrows, we omitted them for the sake of experimental parsimony. In terms of statistical power, we anticipated significant inter-participant variances and relatively small effect sizes, therefore we aimed at keeping all factors within-subjects while maintaining a reasonable task time to avoid participant fatigue.

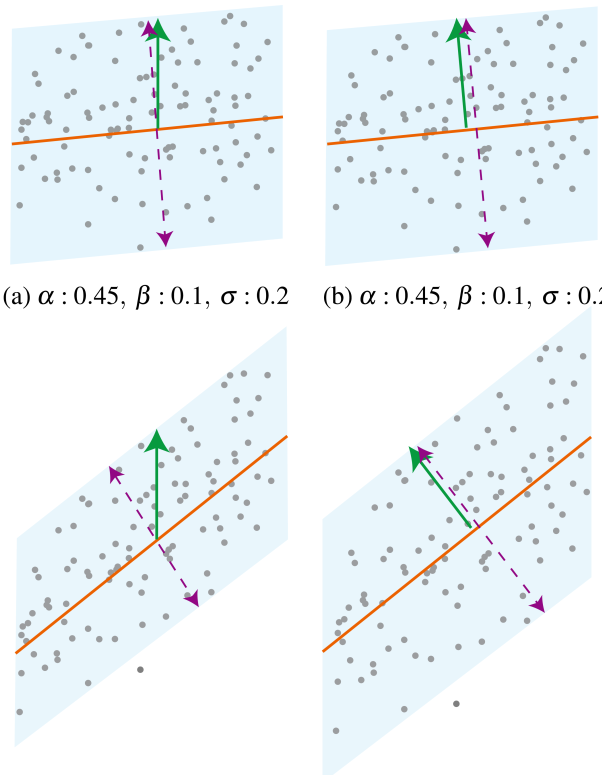
Baseline Stimuli. We adapted the data generation of Correll & Heer [7] to create the stimuli. To have independent control over the relevant statistics, we generated the distributions based on the standard linear regression model:

$$y = \alpha + \beta x + \varepsilon, \tag{1}$$

where ε follows the Gaussian distribution $N(0, \sigma)$. Here, α is the intercept, β is the slope and the bandwidth σ controls the dispersion of the points. Larger values of σ result in larger residuals and weaker trend lines, and this way act as a control over the task difficulty. We synthesize each baseline distribution by uniformly sampling 100 points along a trend in the x direction and then adjusting these points with the Gaussian residual. Note that we alleviate the effect of overplotting by requiring the points to have a minimal amount of occlusion. We explored more complex data models and distributions in experiments described later.

However, perturbing the points sampled from a trend line y (see Eq.(1)) along the vertical direction might result in an incorrect trend estimation. The reason for this is the "sine illusion" [30]; humans appear to have a preference for assessing a line width based on its orthogonal rather than its vertical width. Fig. 2 shows that a scatterplot with a small slope (a) has a larger apparent residual bandwidth (purple line) than for a larger slope (c). In other words, the residual bandwidth attached to a large trend slope seems to be smaller, which could impact our estimation task.

We re-analyzed the data in Correll & Heer [7] and identified an interaction between trend slope and residual bandwidth consistent with this illusion, as shown in Fig. 3. To account for this effect, we inject the residuals in the orthogonal direction to the trend line instead of the vertical direction. Fig. 2 shows a comparison of the baseline



(a) $\alpha: 0.45,\ \beta: 0.1,\ \sigma: 0.2$ (b) $\alpha: 0.45,\ \beta: 0.1,\ \sigma: 0.2$

(c) $\alpha: 0.1,\ \beta: 0.8,\ \sigma: 0.2$ (d) $\alpha: 0.1,\ \beta: 0.8,\ \sigma: 0.2$

Figure 2: Comparison of point sets generated by adding Gaussian residuals along the vertical direction in blue (a,c) and orthogonal direction in purple (b,d) to the trend line (shown in orange).

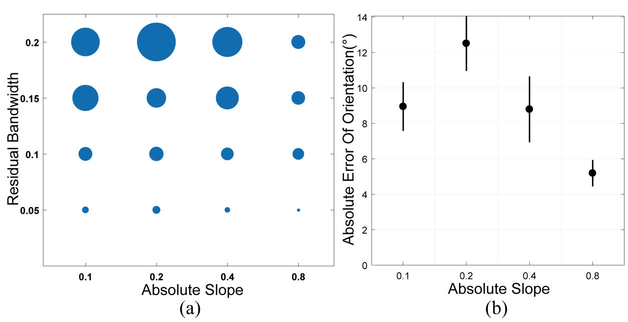


Figure 3: Joint analysis of estimation error resulted by residual bandwidth and trend slope. (a) resulting error (point size) in relation to absolute slope and residual bandwidth; (b) mean values and errors (shown as 95% confidence intervals (CI)) of the estimated slope for different given slopes when the residual bandwidth is 0.2.

distributions generated by these two methods using the same residual bandwidth σ. The bandwidths of the scatterplots in Figs. 2(b,d) seem to have a larger variance than the ones in Figs. 2(a,c).

Participants and Apparatus. Since we were showing participants stimuli which could violate the assumptions of the OLS model, we wanted a pool of study participants with at least some data science expertise and thus did not perform an online study. For the three experiments we recruited 88 participants (41 female and 47 male) from undergraduate and master students of the schools of computer science and information engineering from our university (average age 23.6, median age 25), where the numbers of participants for experiment 1, 2 and 3 were 67 (27 female and 40 male), 73 (37 female and 36 male), and 69 (28 female and 41 male), respectively. Since it is not easy to find many participants with experience in reading scatterplots, we allowed participants from experiment 1 to participate in experiment 2 or 3. Although this might introduce a possible learning effect, we alleviated the effect by having a large time interval (more than one week) between the two experiments. Each experiment lasted 30 minutes on average and the students were compensated with $8.00 USD. The study was conducted on a desktop machine with a 3.4GHz Intel i7-6700 CPU, 8 GB of RAM, Windows 10 operating system using a 24-inch LCD display with a resolution of 1920 x 1080 pixels, and the Logitech M510 Wireless Computer Mouse.

Interface and Measures. We used an interface similar to the one used in Correll & Heer [7] that presented participants with a scatterplot and asked them to interactively adjust a trend line to fit the data best. Our interface required users to adjust the orientation of the trend line and the y-intercept (the intersection point of the line and the *y*-axis) with two independent sliders. Note that instead of manipulating the slope directly (in direct coefficients from say [-1,1]), the participants manipulated the *orientation* of the line, in degrees. Slope and orientation are related by the non-linear arctan function, and Talbot et al. [28] suggest that human judgments are more closely connected to the orientation of the line rather than its slope. As such, in our experimental apparatus, users alter the orientation of their estimated line rather than the slope, and we measure error in terms of degrees of difference between the participant line and the source trend.

After adjusting both sliders, participants had to confirm their estimated trend line before proceeding to the next trial. Before starting each experiment, we explained the concept of trend estimation to each participant using three examples and then conducted a training session with 8 trials. To avoid the influence of mark orientation, all training examples used circle marks. In all experiments, we did not tell participants about the meaning of oriented triangles and also did not tell them that there were outliers and that they should exclude outliers from the trend estimation in Experiment 2.

For each trial, we measured the errors in orientation (in degrees) with respect to the slope, value with respect to the y-intercept, as well as the per-trial response time (in milliseconds). We rejected all trials from participants with high average orientation errors ($\geq$20 degrees). This amounted to 0/67, 2/73, 1/69 of participants for experiments 1, 2, and 3, respectively. Our figures report all measures with 95% t-confidence intervals. Raw data tables, stimuli, additional analyses and figures are available in our supplemental material at https://osf.io/hzdf3.

Interview. After completion of each experiment, we conducted a short interview with the participants in which we asked them three questions:

- Q1:"How might visual marks affect your trend estimation, if at all?";
- Q2:"How do you estimate trends?"; and
- Q3: "Which marks do you prefer?".

Since Experiment 2 focuses on the influence of outliers, we change Q2 to the following one:

- Q2:"How might outliers affect your trend estimation, if at all?".

We report the interview results for each experiment below.

4 EXPERIMENT 1: BASELINE DATA

Our first experiment studies investigated three independent variables and their interaction with trend estimation: trend slope, residual bandwidth, and mark orientation.

4.1 Data Generation

We synthesized 32 distributions with 8 slopes $\beta = \pm\{0.1, 0.2, 0.4, 0.8\}$ and 4 bandwidths of Gaussian residuals $\sigma = \{0.05, 0.1, 0.15, 0.2\}$ based on Eq. 1. For each distribution, we selected five types of visual marks: two kinds of *statically oriented marks*: circles, upward triangles (denoted as *Tri-Up*), and three triangle forms with *data-driven, dynamic* orientations. Data-driven marks are either *informative* (when pointing in the same direction as the trend) or *biasing* (when pointing in a direction away from the intended trend line). We refer to these conditions as *Tri-0*, *Tri-20* and *Tri-40*, where the number indicates the orientation of the mark (in angles) away from the trend line. Although clockwise or counterclockwise rotation was taken randomly for each stimuli, overall there was about the same number of participants assigned to clockwise and counterclockwise orientations. The first two mark shapes, circle and Tri-Up, are commonly used marks [8]. While a circle does not have any specific orientation, the Tri-Up has a fixed orientation and therefore might amplify a steep trend line or suppress a more horizontal one, see an example stimuli in Fig. 4.

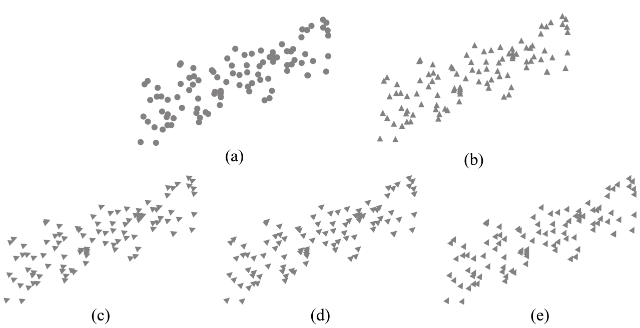


Figure 4: Scatterplots with five different marks explored in Experiment 1: (a) circle; (b) Tri-Up; (c) Tri-0, (d) Tri-20, and (e) Tri-40.

Participants saw one example of each combination of mark shape and data distribution and 4 additional validation stimuli for attention check, in total $32 \times 5 + 4 = 164$ trials. We did not impose any time constraints for the trials, but almost all participants completed their trials in less than 30 minutes. We recorded the response time, the estimated trend orientation, and the estimated intercept for each trial.

4.2 Hypotheses

For this experiment, our major goal was to investigate how mark orientation would affect trend estimation. Thus we had three hypotheses:

(1) **Mark orientation would have a strong effect on trend estimation**: an orientation consistent with the trend direction results in a higher accuracy. In other words, we expect mark shapes to guide or bias the estimates of the participants. Prior work indicates that carefully designed mark orientation can direct the perception of cluster structures in scatterplots [3, 17].

(2) **The effect of mark orientation on trend accuracy would be most pronounced when residual bandwidth was large**: that is, a small bandwidth would result in similar accuracy for different marks, while a large bandwidth would result in large differences in accuracy across different marks. Prior work [7, 21] shows that the dispersion of points influences accuracy in trend and correlation judgments, with very high accuracy when dispersion from the trend is small. Insofar as we anticipated that mark orientation would guide or bias participants in their estimates, we suspected that the impact of this guidance would be less pronounced for stimuli where the dispersion is low and so accuracy is high in any event.

(3) **Trend slope would not have a significant effect on trend estimation.** In Correll & Heer [7], there was not significantly different accuracy in trend estimation across different trend slopes. We believed that this finding would also hold in our setting.

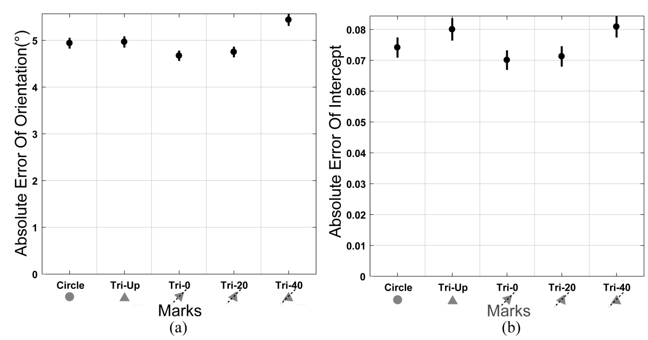


Figure 5: Results of Hypothesis 1 in Experiment 1: effect of mark orientation on the accuracy of estimating trend orientation (a) and trend intercept (b). Triangle marks with an orientation consistent with the actual trend result in the smallest error, while Tri-Up results in the largest errors.

4.3 Results

We performed a three-way repeated measures analysis of variance (rANOVA) of the effects produced by trend slope, residual bandwidth and mark type for trend estimation. Estimation accuracy was measured by orientation error and intercept error, which were defined as the absolute difference between the measured estimates and the OLS trend orientation/intercept. Note that the OLS trend is almost the same as the ground truth, since the data agrees with the assumptions of the OLS model.

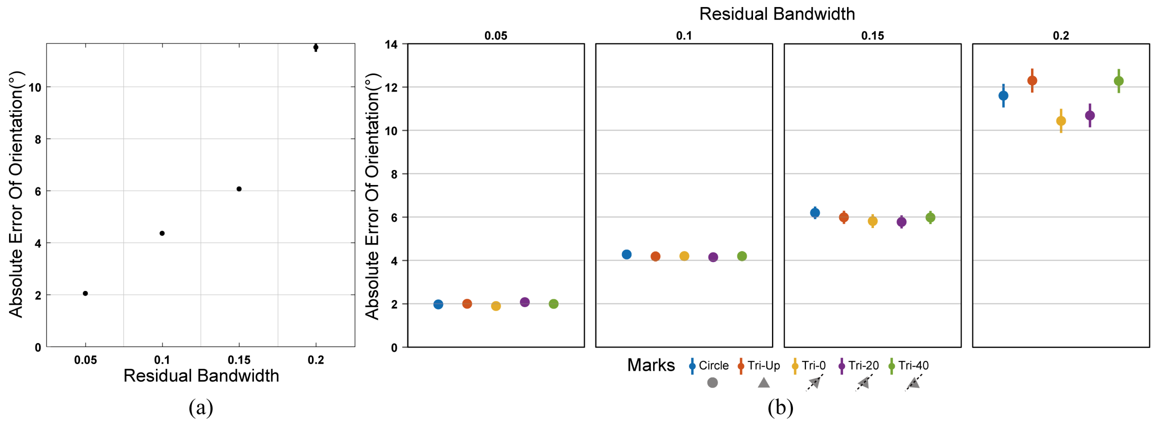


Figure 6: Results of Hypothesis 2 in Experiment 1: (a) Effect of residual bandwidth on accuracy of estimating trend orientation. The absolute orientation error increases monotonically as the bandwidth increases. (b) Effect of residual bandwidth and mark orientation on estimating trend orientation.

Mark Orientation. Mark orientation had a significant effect on the orientation estimation error ($F(4,264) = 10.437; p < .0001$) and trend intercept ($F(4,264) = 11.725; p < .0001$), but there was no significant effect on response time ($F(4,264) = 1.08; p > .05$). As shown in Fig. 5, Tri-0 results in the smallest error for the estimated trend orientation and Tri-40 in the largest error, with the interquartile means of the absolute errors being 4.66 and 5.56. We performed a post-hoc Tukey's Honest Significant Difference (HSD) test to examine pairwise difference and we were able to confirm that there are significant differences between Tri-Up/Tri-40 and other mark shapes ($p < 0.05$), but not between the other mark shapes.

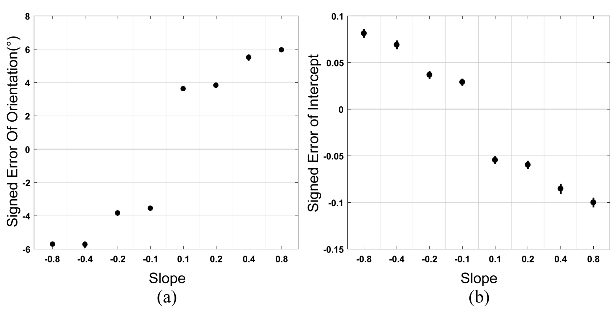


Figure 7: Results of Hypothesis 3 in Experiment 1: (a) Effect of trend slope on the accuracy of trend orientation estimation. A negative slope results in underestimation of the orientation, while a positive slope results in overestimation. (b) Effect of trend slope on the accuracy of trend intercept estimation. A negative slope results in an overestimation of the intercept, while a positive slope results in underestimation.

Residual Bandwidth. The residual bandwidth also had a significant effect on the estimation accuracy of trend orientation ($F(3,198) = 399.64; p < .0001$) and trend intercept ($F(3,198) = 377.92; p < .0001$). Fig. 6(a) shows that the interquartile mean of the absolute trend orientation error increases monotonically with the bandwidth, from 2 degrees when the bandwidth is 0.05 to 12 degrees for a width of 0.2. Similar results for the intercept error can be found in the supplemental material. Note that the residual bandwidth also does not have a significant effect on the response time for trend estimation ($F(3,198) = 0.54, p = .66$).

There was also a significant interaction effect between mark orientation and residual bandwidth on trend estimation accuracy ($F(12,792) = 4.83, p < .0001$). To further investigate this interaction, we show the interquartile mean of orientation errors for different marks at different bandwidths in Fig. 6(b). When the bandwidth is 0.15 or less, all marks result in similar orientation errors. For example, when the bandwidth is 0.05, mark orientation does not have any effect on trend estimation ($F(4,264) = 1.92, p = .11$). By contrast, mark orientation has a significant effect($F(4,264) = 9.64, p < .0001$) when the residual bandwidth was 0.2. Different marks have different interquartile means: Tri-0 has the smallest error while Tri-40 has the largest, similar to Tri-Up. This observation is consistent with the one shown in Fig. 5.

Trend Slope. Contrary to our expectations, trend slope also had a significant effect on trend estimation accuracy ($F(7,462) = 16.71; p < .0001$) as well as on estimating the trend intercept ($F(7,462) = 48.75; p < .0001$). Fig. 7 shows that the interquartile mean of trend orientation errors is heavily influenced. For positive trends, the trend orientation is overestimated, while the trend intercept is underestimated.For negative trends, the trend orientation is underestimated and the trend intercept is overestimated. For both, positive and negative trends, the absolute orientation error monotonically increases with the absolute trend slope from 0.1 to 0.8.

4.4 Post-Task Interviews

Q1. 74% of participants claimed that their estimates were influenced by the mark type; for instance: i)"*The visual mark has less effect for scatterplots with less dispersions, but I'd like to rotate the trend line along the mark orientation for the scatterplots with more dispersions;*" and ii)"*the mark orientation assists me to do fine-tuning when it is close to my estimation; but otherwise it might be misleading.*" By contrast, the remaining participants stated that visual marks did not affect their trend estimation. The participants explained: "*I completely did not notice the difference of visual marks;*" and

"*I mainly examined data distribution and visual marks have less influence especially for scatterplots with clear trends.*"

Q2. By analyzing the answers of the participants, we found three typical procedures for performing regression by eye. The majority (63/67) first examined the bounding box of all points and then fine-tuned the line to bisect the bounding box. Besides, three participants said "*find a line so that half of points are above it and half of them are below it*" and one participant mentioned that "*find a line that minimizes the distances from all points to it.*" Since points were symmetrically distributed along the trend line, taking the bounding box as a visual proxy was reasonable for trend estimation.

Q3. The participants (74%) whose trend estimation were influenced by mark orientation preferred Tri-0, as it provides direct information about the trend, especially when dispersion was large. The remaining participants preferred circles, because they were more familiar with circles and assumed that visual marks would not negatively impact their estimates.

5 EXPERIMENT 2: OUTLIERS

Correll & Heer [7] explore a single class of outliers, defined as points in the top 10% or bottom 10% of the graph (whichever is furthest away from the intended trend). We wanted to include a more general class of secondary point clusters not generated by the same Gaussian process as the main cluster, with variable distances to the main points. Therefore, we expanded their outlier definition. Our objective with this experiment was to see if these outlying clusters would bias the trend estimation in proportion to their distance from the main cluster, and if the influence of mark shapes observed in the prior experiment would furthermore attenuate this bias.

5.1 Data Generation

Similar to our other experiments, we generated an initial set of points using the following parameters: 8 slopes $\beta = \pm\{0.1, 0.2, 0.4, 0.8\}$ with 4 bandwidths of Gaussian residuals $\sigma = \{0.05, 0.1, 0.15, 0.2\}$. We then generated a secondary cluster via a Gaussian distribution and adjusted its density and distance in correspondence to the main cluster, see Fig. 8 for examples.

Outlier Position. We randomly placed the *x* position of the outlier cluster centered around the beginning, first third or the end of the major point set. The cluster was randomly placed on the *Top* or *Bottom* relative to the main cluster, with the central *y* position determined by:

$$y = \begin{cases} (h - (\alpha + \beta x))\gamma + \alpha + \beta x, & \text{on the Top,} \\ (\alpha + \beta x)(1 - \gamma), & \text{on the Bottom} \end{cases} \tag{2}$$

where α and β are the parameters for generating the main cluster (see Eq. 1) and *h* is the canvas height. The weight γ determines the distance and was set to $\{0.25, 0.50, 0.75\}$. Figs. 8(a,b,c) show three examples of outlying clusters with different γ values.

While in practice the distance of an outlying cluster to the majority of points defining the trend line is influenced by the slope, bandwidth, and γ, we define the distance as the vertical distance from the center of the outlier cluster to the robust trend line, as illustrated by the dashed line in Fig. 8.

Outlier Density. Once the residual bandwidth σ of the major cluster is selected, we generate the outlier cluster by using a Gaussian distribution with the same bandwidth σ. To create stimuli with varying outlier densities, we introduce a random factor δ from the set $\{0.5, 1, 2\}$ which we use to re-scale the positions of the points in the outlying cluster:

$$x = (x - x_c)\delta + x_c, \tag{3}$$

$$y = (y - y_c)\delta + y_c, \tag{4}$$

where (x_c, y_c) is outlying cluster center. Figs. 8(d,e,f) show three examples where the outliers have different densities, resulting from different values of δ.

Mark Orientation. In accordance to the first experiment, we use two mark types with static orientation (circle and Tri-Up) as well as data-driven triangle marks. In this experiment, however, we calculate two orientations, based on either the slope of the *robust* trend line (the fit which ignores the outlying cluster) or the standard OLS trend line (based on all points). Since the main cluster meets the assumptions of OLS, the robust tend line is directly defined by α and β. We refer to the two resulting triangle marks with the orientations of the robust and OLS trend lines as *Tri-R* (robust) and *Tri-O* (OLS), respectively. Fig. 9 shows example stimuli for all four different mark types.

Stimuli Generation. Since the effect of the slope in trend estimation is mainly determined by its sign (Experiment 1), we did not show every trend slope to our participants. To maintain a manageable number of stimuli, we kept cluster density δ, the number of outlier points, and the *x* position of the outlying clusters as random factors, while guaranteeing an equal distribution over all participants. Participants saw for all of the following factors one of each level: four marks, slopes of two different signs, four residual bandwidths, and three distance parameters, which summed up to a total of $4 \times 2 \times 4 \times 3 = 96$ stimuli plus 4 additional validation stimuli synthesized by Eq. 1 as attention checks.

5.2 Hypotheses

We had three main hypotheses for this experiment:

(1) **Mark orientation would affect trend estimation**: we believed that a mark orientation consistent with the robust trend line would encourage participants to ignore or otherwise downweight outliers. As with the prior experiment an prior work [3, 17], we anticipated that marks orientation might bias or guide judgments.

(2) **The distance between outliers and the major cluster would affect trend estimation**: as the distance between outlier cluster and major point cluster centroids increased, we anticipated that participants would downweight the outlying clusters when making their estimates. Prior work shows that participants downweight (but did not entirely discount) outliers when performing trend estimation [7], although only with very distance outliers. We wanted to explore the placement of outliers in greater detail.

(3) **The density difference between outliers and major trend has a significant effect**: with increasing density difference between main trend and outliers, participants would more easily ignore outliers. Previous research indicates that density difference affects human judgments in visual clustering [23].

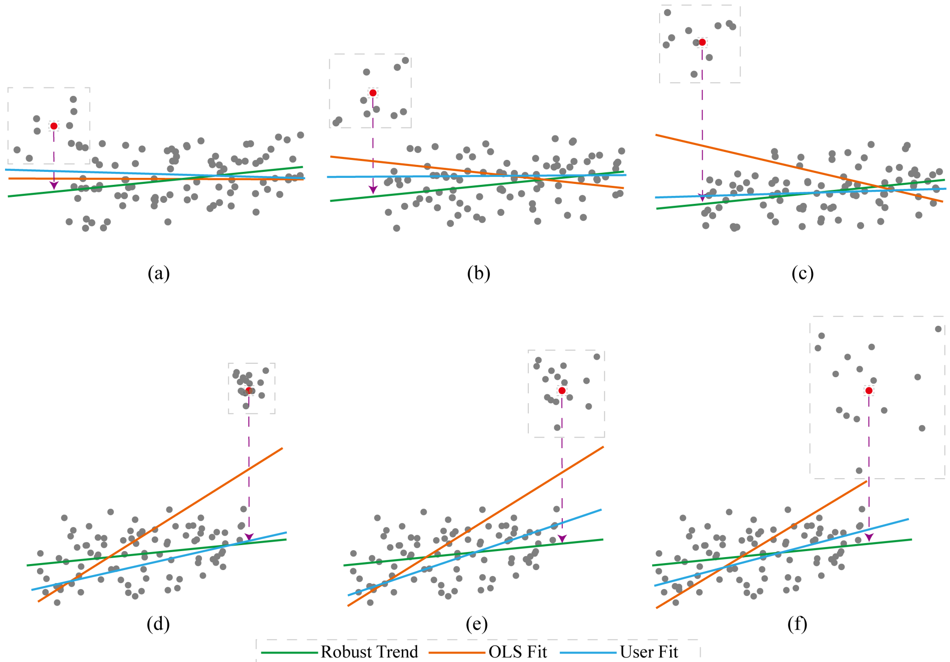


Figure 8: Three examples for outlier clusters with different distances (a,b,c) to the major cluster in y direction (distance indicated by the red dashed line) and three examples of outliers with different densities (d,e,f). The green line shows the robust trend line fitted only to the major cluster, while the orange line indicates the OLS trend line fitted to all points in both clusters.

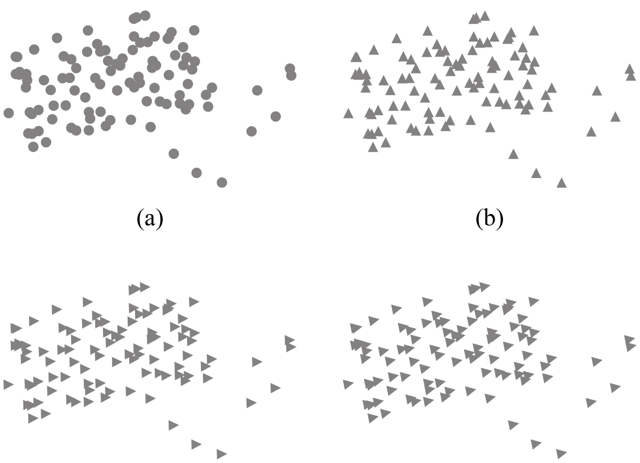


Figure 9: Scatterplots with four different marks explored in Experiment 2: (a) Circle; (b) Tri-Up; (c) Tri-R, and (d) Tri-O.

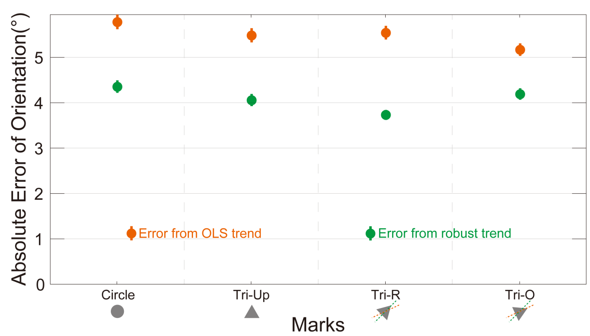


Figure 10: Results of Hypothesis 1 in Experiment 2: effect of mark orientation on the accuracy of trend estimation. The estimated trends are closer to the robust trend than the OLS trend. Tri-R and Tri-O yield less errors in terms of the robust trend and OLS trend, respectively, while Circle and Tri-Up perform similarly.

5.3 Results

We performed a repeated measures rANOVA analysis for understanding five factors: number of outliers, mark shape, outlier distance to the major cluster, outlier location, as well as outlier density. For analyzing the effect of the outlier distance, we computed all stimuli and then quantized them into five uniform groups based on the ratio of cluster centroid distance and the correlation of main cluster. In the supplementary materials we discuss other factors for which we did not have strong or relevant hypotheses.

Mark Orientation. In accordance to our hypothesis, mark orientation had a significant effect on the accuracy of the estimations for trend orientation ($F(3,210) = 5.06, p < .05$) and trend intercept ($F(3,216) = 5.86, p < .001$). As shown in Fig. 10, marks pointing towards the direction of the robust trend (Tri-R) created less error with respect to the robust trend. Similarly, marks pointing towards the full OLS line (Tri-O) produced less error with respect to this line, while the other two marks perform similarly. Our results also support the

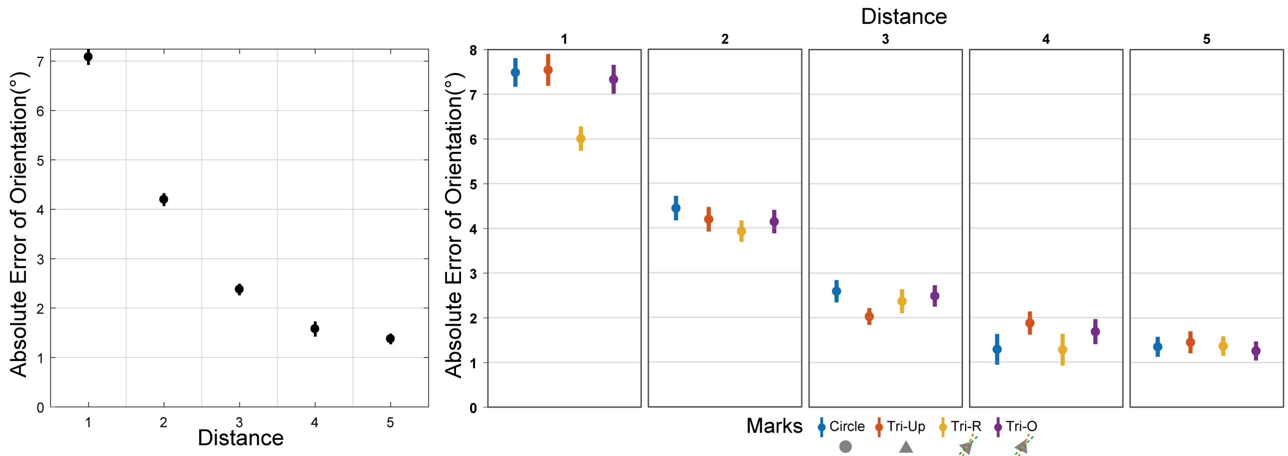


Figure 11: Results of Hypothesis 2 in Experiment 2: (a) Effect of outlier distance on the accuracy of trend orientation estimation; (b) interaction effect between outlier distance and mark orientation on the accuracy of trend orientation estimation.

finding of Correll & Heer [7] that participants down-weighted outlying poin For the following analyses we therefore used the robust trend line as our basis of comparison. Results for estimating intercepts can be found in the supplemental materials.

Outlier Distance. Corresponding to our hypothesis, outlying clusters further away from the remaining points were considered less important for trend estimation. With increasing distance of the outlying cluster our participants produced estimations closer to the robust trend line. We observed that the distance had an extremely significant effect on the estimation accuracy with respect to the robust trend orientation ($F(4, 284) = 27.94, p < .0001$) and the intercept ($F(4, 293) = 29.03, p < .0001$). Fig. 11(a) shows that the absolute orientation error drops rapidly from 7.1 to 1.3 when the outlier distance changes from the smallest to the largest values.

To learn how mark orientation interacts with outlier distance, Fig. 11(b) shows the CI error bars for different marks. Tri-R marks have the smallest error when the distance is small, while Tri-Up marks have the largest error. Since the outlier distance is associated with how ambiguous a trend is perceived, all marks result in smaller errors with increasing outlier distance. Similar to the prior experiment, mark orientation seems to have the strong impact when trends are otherwise ambiguous, see examples in Fig. 1.

Outlier Density. Our results do not support our hypothesis of differences in estimation error across cluster densities. Fig. 12(a) shows the CI error bars for the different outlier density groups in trend estimation, where three groups have similar errors.

Our ANOVAs show that the density difference between the majority of points and the outlier cluster does not have a significant effect on the estimation accuracy of trend slope ($F(2, 140) = 1.88, p > 0.05$) and trend intercept ($F(2, 144) = 2.90, p > 0.05$). However, Tukey's HSD Post Hoc Test only identifies the dense group have a significant interaction with the other groups in trend estimation (see Fig. 12(b)), while the sparse group does not have a significant difference compared to the other groups.

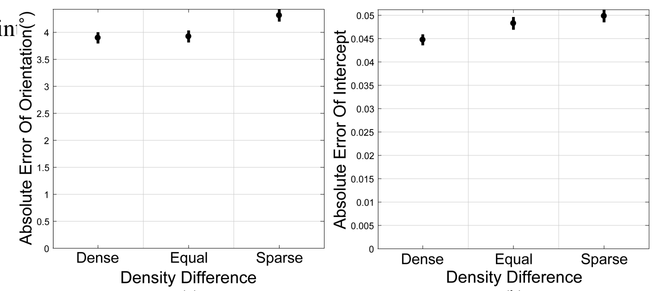


Figure 12: Results of Hypothesis 3 in Experiment 2: effect of outlier cluster density on estimating trend orientation (a) and intercept (b).

5.4 Post-Task Interviews

Q1. 57% of our participants mentioned that mark orientation had an influence on their estimates. For instance, participants stated that "*if the mark has an orientation, I first rotate the trend line to align the mark orientation and then fine-tune the line based on data distribution;*" "*I rotate the line along the mark orientation when the scatterplots have a higher amount of dispersion,*" and "*when the mark orientation is close to my estimation, it provides a reference for me to quickly fine-tune the trend line.*"

Q2. Regarding visual proxies, around 89% of the participants believed that their estimations were heavily influenced by distance of the outliers to the majority of points; 72% of the participants believed that density differences would also have an effect. Representative statements include that "*if the outlier cluster is compact or far from the major cluster, it can be ignored, otherwise it might have an influence;*" and "*if the number of points in the outlier cluster is large, I will move the trend line towards it.*"

Q3. Around 65% of participants preferred Tri-R and Tri-O marks. Some participants cited the ability of these data-driven marks to speed up decision-making and estimation. For example, the participants mentioned that "*when I am uncertain, I'd like to adjust trend* *line along the mark orientation.*" The rest preferred standard circle marks; some claimed that the inherent directionality in triangular marks had the potential to be misleading. For example, one participant stated that "*I am not sure whether the triangle orientation is correct and thus I am worried if the orientation misguide my estimation.*"

6 EXPERIMENT 3: NON-UNIFORM DENSITY

Local cluster density heavily influences the quality of the regression model [24]. On the other hand, non-uniform densities will result in heteroskedasticity, violating the assumptions of OLS and generating poor fits. We speculated that local density might also impact visual estimate of trends. Furthermore, as in our previous experiments, mark shapes with data-driven orientations could guide or bias the estimates for ambiguous stimuli. In this experiment we investigate how the regression by eye works for non-uniform data and identify the potential effects from local densities and mark orientations.

6.1 Data Generation

To generate scatterplots with varying densities, we fixed a point Set A and then merged it with another point Set B whose number of points and slope could be adjusted. Specifically, we first created a Set A with 40 points, slope of $\beta = 1$ and residual bandwidth of $\sigma = 0.2$ by using Eq. 1. Then, we merged Set A with a second Set B generated by combining 6 possible point numbers $\{40, 60, 80, 100, 120, 140\}$, with 4 slopes $\{0.1, 0.2, 0.4, 0.8\}$, and a residual bandwidth of 0.1. The y-intercept of the trend line for point Set A was zero, while the one for point Set B was chosen from the set {0.1, 0.3, 0.4, 0.45}. Here, we only analyzed 4 slopes instead of 8 (positive and negative) for two reasons: i) we had already studied effect of trend slope sign on trend estimation in our previous experiments and had no strong hypotheses around an interaction with cluster density (see Fig. 7) and ii) mixing a positive trend in Set A with a negative trend in Set B can only produce small regions with non-uniform densities. The number of total stimuli was therefore $4 \times 6 \times 4 = 96$ stimuli plus 4 additional validation stimuli synthesized by Eq. 1 as attention checks for this experiment.

Fig. 13 shows six examples generated by merging the green and orange point sets with overlaid green and orange trend lines. Figs. 13(a,b,c) demonstrate that changing the number of points in one set with a fixed slope influences the perceived density. On the other hand, changing the trend slope while keeping the number of points unchanged influences the bounding boxes, as shown in Figs. 13(d,e,f). Since bounding boxes are an efficient visual proxy for the perception of the correlations in scatterplots [36], we speculated that the size of the bounding box might also impact the visual estimates of the trends. Note that the colors used in Fig. 13 are for illustrative purposes only, Fig. 14 shows an example stimuli used in the experiment.

Mark Orientation. As with the prior experiments, we used static marks (circles and Tri-Up) as well as a set of marks with data-driven orientations. In this case, the triangle marks were rotated towards the orientation of the trend of Set A (*Tri-A*) or Set B (*Tri-B*), which serves as the upper or lower bound of the orientation of the user fit. In particular, Tri-A was fixed with 45°, while the orientation of Tri-B was a value in the set of $\{5.7°, 11.3°, 21.8°, 38.7°\}$.

6.2 Hypotheses

We had three hypotheses for this experiment:

(1) **Mark orientation would affect trend estimation for non-uniform data**: as per our previous experiments, we expected mark orientation to guide or bias the estimates. For instance, trend estimates for the Tri-A would be closer to the trend of Set A, and for Tri-B to be closer to Set B. Since the points of Set A by construction have a larger slope than the points of Set B, this would result in an increasing estimated slope for Tri-A, and a decreasing estimated slope for Tri-B.

(2) **Local density would affect trend estimation**: as the number of points in Set B increases, the estimated trend would hew closer to the trend orientation of Set B (and so would decrease). Previous research indicates that local density heavily affects human judgments in visual clustering [23].

(3) **The slope of the points in Set B would affect estimation precision**: we anticipated that as the slope increases, the resulting stimuli would be less "ambiguous" which, as in our previous experiments, would result in a more consistent performance. Similarly we anticipated that the impact of mark orientation would be less pronounced for less ambiguous stimuli.

6.3 Results

We performed a three-way repeated measures ANOVA analysis for measuring the impact of mark orientation, number of points and trend slope of the second set for trend estimation. Unlike in the prior experiment, where we could use the robust trend line as an anchor based on its dominance in participant responses, here there is no single "correct" fit: OLS is not an appropriate tool for intermixed heteroskedastic data. Due to the absence of a ground truth in this experiment, we are left only with measuring precision rather than accuracy: our analysis is directly based on the estimated trend orientation and intercept related to the individual clusters A and B. E.g., significant differences in the values estimated for different marks indicates different patterns of decision-making, with higher estimated slopes indicating a preference for the trend of Set A over the (by construction) lower slope of Set B.

Mark Orientation. Our results partially support the hypothesis that mark shapes would affect trend estimation. We observed a significant effect of the mark shape on the estimation of the trend slope ($F(3, 201) = 8.60, p < .0001$) and trend intercept ($F(3, 204) = 4.76, p = .003$). A post hoc Turkey HSD test showed that only Tri-B and Tri-Up performed significantly differently from each other when estimating slope ($p < .005$) and intercept ($p = 0.005$), whereas marks Tri-B and Tri-A did not have a statistically significant difference in estimates of the trend intercept ($p > .05$).

Fig. 15 illustrates estimation accuracy for different mark shapes. Tri-B resulted in the smallest mean orientation (30.1°) and the largest mean y-intercept (0.09), while Tri-Up resulted in the largest mean orientation (31.9°) and the smallest mean y-intercept (0.05). Tri-A, Circles and Tri-Up all performed similarly for the estimation of the trend intercept.

Local Density. Our results support the hypothesis that estimates are influenced by the density of the secondary cluster. We observed that

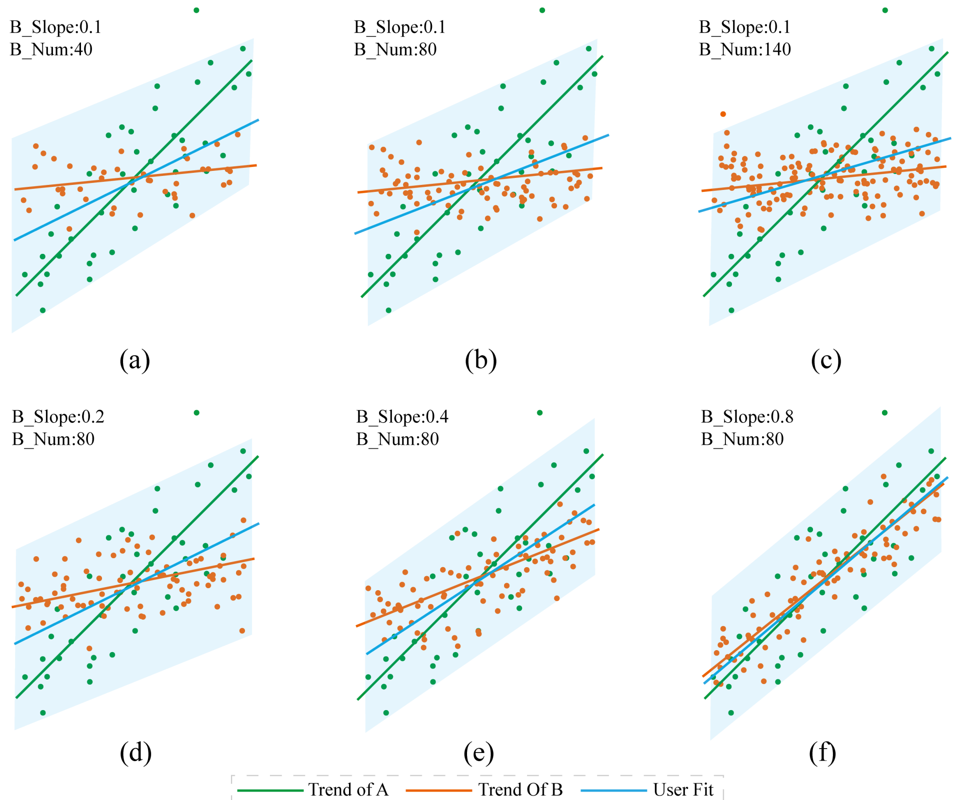


Figure 13: Scatterplots with varying point numbers and slope: (a,b,c) show varying point numbers in Set B; (d,e,f) show variations of the slope of Set B. The overlaid green line is the fitted trend line for point Set A, the orange line for point Set B; and the blue line is the user fit. The bounding box shown for each scatterplot is computed by discarding outliers.

the number of points in Set B has a significant effect on the estimation of the trend slope ($F(5,335) = 218.12, p < .0001$) and trend intercept ($F(5,340) = 76.92, p < .0001$). Fig. 16(a,b) illustrates the results.

As the number of points in Set B increases, the estimated trend orientation decreases and the intercept increases (see Fig. 16(a,b)). This observation indicates that estimates are influenced by the point density, which was enlarged by the increased number of points in Set B. However, the interaction effect between density of the secondary cluster and mark orientation is not significant ($F(15,1005) = 0.95, p > 0.05$). Fig. 16(c) shows that Tri-B and circles result in similar and smaller estimation results when the point number in Set B is bigger than 40. Tri-Up results in the largest estimated trend orientation, which is consistent with the observation in Fig. 15.

Trend Slopes. Our results support the hypothesis that estimates are influenced by the slope of Set B. We observe that the slope of Set B has a significant effect on the estimation of the trend orientation ($F(3,201) = 467.02, p < .0001$) and intercept ($F(3,204) = 301.98, p < .0001$). As shown in Fig. 17(a,b), the estimated mean trend orientation increases and the variance decreases as the slope increases for Set B. This is consistent with our observation that Set B becomes more distinct with increasing slope and the ambiguity of the stimuli is reduced.

Fig. 17(c) further shows the interaction between trend slopes and visual marks, which has a significant effect on estimating mark orientation ($F(9,603) = 4.07, p < .0001$). We can see that the difference of the estimated mean orientations between Tri-A and Tri-B rapidly decreases from 5° to almost zero as the slope increases. Meanwhile, the local density grows larger and the bounding boxes become smaller, see example in Figs. 13(d,e,f). Previous work [36] shows that the bounding box and local density are efficient visual proxies for the perception of correlations in scatterplots; however, it is unclear which plays the more important role.

6.4 Post-Task Interviews

Q1. About 22% of the participants believed that mark orientation does not have an influence on their estimates and utilized the bounding box of the whole point set and the one of high density areas to estimate trends. The rest of the participants (78%) believed that the mark orientation influenced their estimation, with quotes such as "*I will unconsciously rotate the trend line along the mark orientation,*" or "*I will first rotate the trend line along the mark orientation and then fine-tune it based on the data,* and "*when the mark orientation is close to my estimation, I will be sure that it is the orientation of trend line and thus rotate the trend line to align it.* "

Q2. We found almost all participants estimate trends by using bounding boxes or local densities with three different procedures. The two

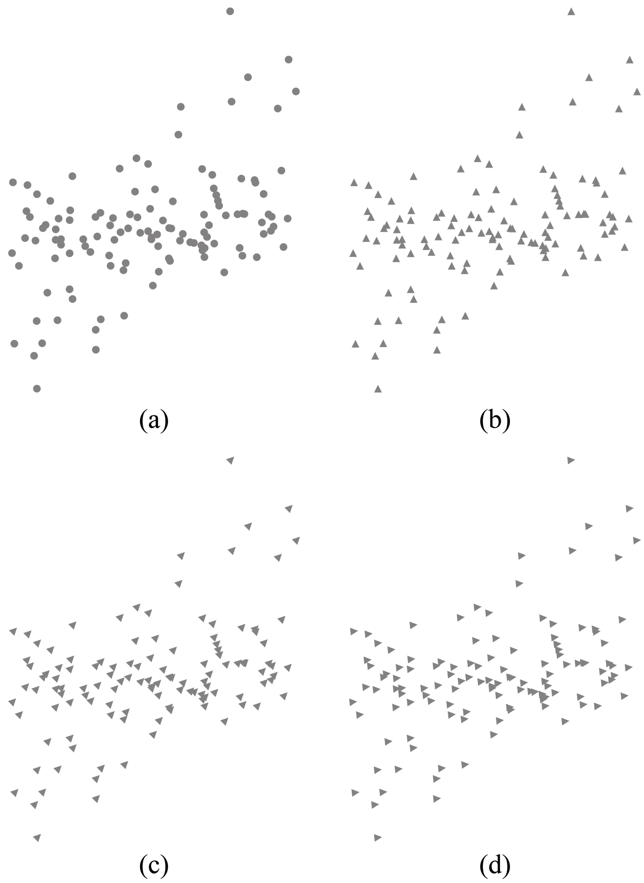


Figure 14: Scatterplots with four different marks explored in Experiment 3: (a) Circle; (b) Tri-Up; (c) Tri-A, and (d) Tri-B.

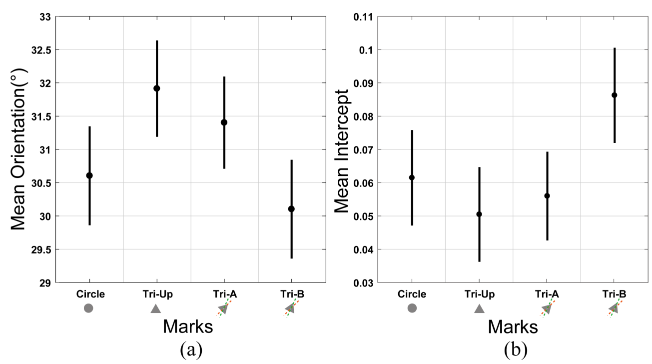


Figure 15: Results of Hypothesis 1 in Experiment 3: for our third experiment with intermixed point sets A and B, the effect of mark orientation on the estimation of trend orientation (a) and intercept (b).

major procedures were: i) "*I first adjust the trend line based on dense areas and then fine-tune it along sparse points*"; and ii)"*I first look at the bounding box of the whole data, then find a line to bisect the bounding box, and finally fine-tune the line in terms of the density difference in scatterplots,*" whereas only two participants mentioned "*I only consider dense areas of the scatterplot for adjusting trend lines.*" We conclude that local densities are one major visual proxy for estimating trends in scatterplots with non-uniform densities.

Q3. The majority (66%) of participants preferred circular marks and some participants expressed concern that marks with orientation might mislead them. That is, a preference for circular marks over data-driven marks appears to be a worry over their undue influence over their estimates, and a clear assumption of impact. For example, two participants mentioned that "*Given the rotated triangle marks, I often unconsciously rotated the trend line along their orientation. However, I am not sure if such orientation is correct and thus I prefer circle marks to avoid incorrect estimations.*"

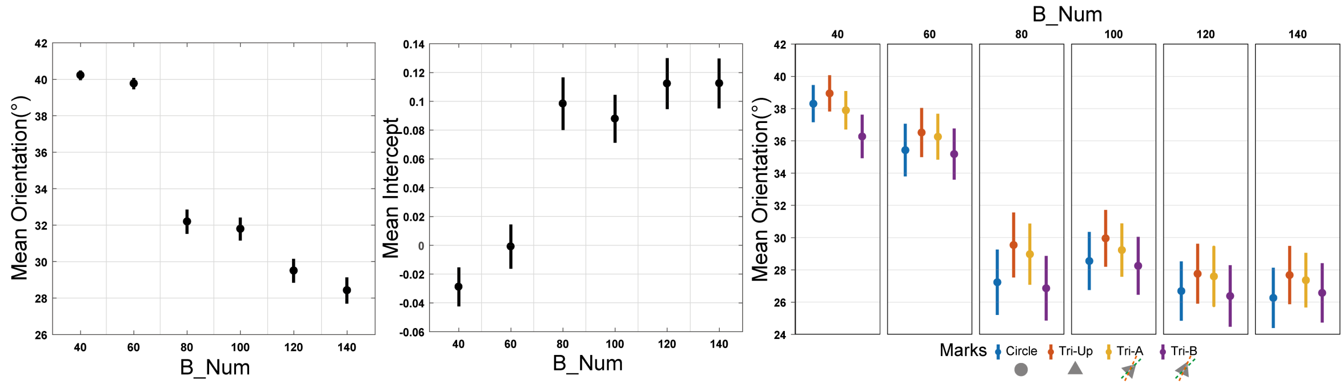
7 DISCUSSION

Directly showing trendlines in scatterplots is a clear way to indicate trends, however, the model enforced by such trendlines might be inappropriate for some data. Even with an appropriate model, viewers may not have enough statistical expertise to interpret the included trendlines. We propose data-driven orientation in marks as a way to *guide* viewers' perceptions of trends without dogmatically *enforcing* a particular model on the data. However, advocating for this technique required us to more fully investigate the strengths and limitations of regression by eye. Building on previous work [7], we further investigated how mark orientation influences trend estimates.

Accuracy of Regression by Eye. Our first collection of findings deals with our capacity to perform regression by eye in scatterplots that differ from the single cluster gaussian collections of points common in prior experiments [7, 12, 21]. Without a strong belief in the ability of viewers to make reasonably accurate estimates of trends from scatterplots, it would be difficult to argue for any intervention other than the explicit annotation of trend lines. This would also be a troubling result for other visualization techniques that rely on visual estimations of aggregates, such as assessing distributional fits via Q-Q plots, or labeling points as outliers.

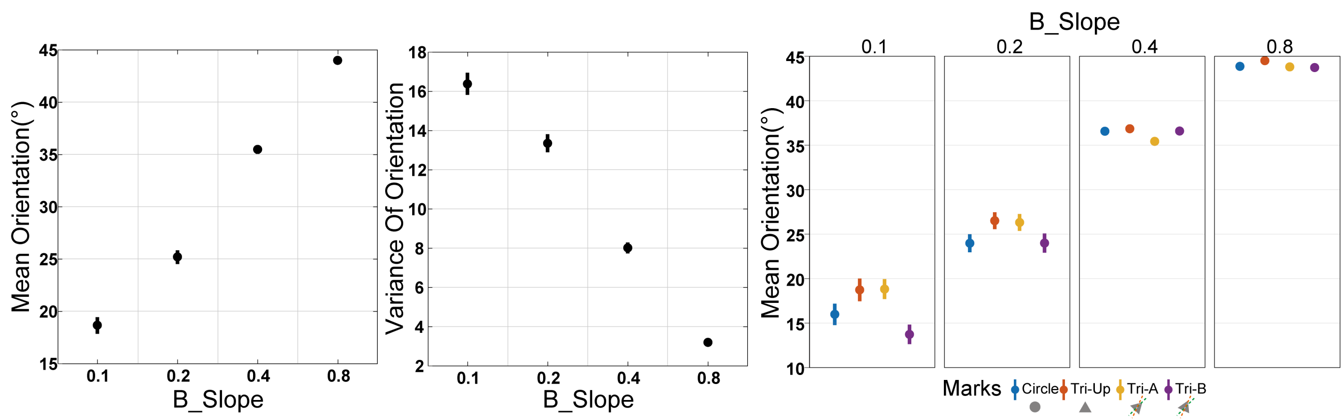
Overall, we find that, in our constrained estimation settings (where, for instance, the user is adjusting the parameters of a linear model, rather than choosing a particular model type directly), our participants made estimates that were, on the whole, reasonable. We reflect on our experimental apparatus, as well as provide additional analyses, in our supplemental materials. Participants consistently ignored far-away outliers and non-dominant interior clusters. Although not strictly an apples-to-apples comparison due to, among other reasons, the different ways in which residuals were added to data (which we would anticipate having the largest impact when the bandwidth of residuals are large), Fig. 18(a) compares participant estimates in our first two experiments using circle marks (the two experiments where we have selected a ground truth trend) to results in Correll & Heer [7], showing similar patterns of performance, and in particular low error rates (less than 5° for our two most tightly correlated data sets). We note that in such tightly correlated stimuli, performance among different mark types was similar to circle marks.

Influence of Mark Orientation. Prior work has found that visualizations with ambiguous features can support multiple interpretations [35]. Marks with inherent orientation appear to function as a tool to prime or bias participants towards a particular interpretation. We found consistent impacts of shapes with inherent orientations



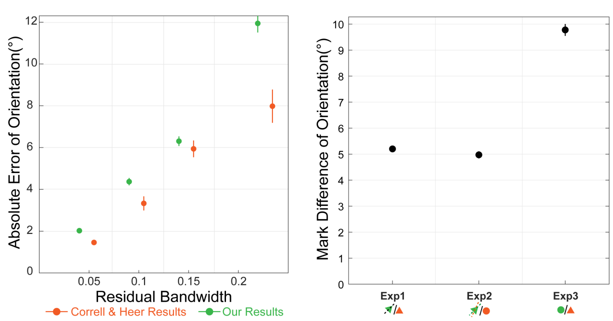
(a) (b) (c)

Figure 16: Results of Hypothesis 2 in Experiment 3: for our third experiment with intermixed point sets A and B, the effect of the number of points in Set B on the estimation of trend orientation (a) and intercept (b). (c) Interaction between the number of points in Set B and mark orientation on the estimation of the orientation.



(a) (b) (c)

Figure 17: Results of Hypothesis 3 in Experiment 3: for our third experiment with intermixed point sets A and B, the effect of the slope of Set B on the estimation of trend orientation, in terms of both mean (a) and variance (b); (c) Effect of the interaction between the slope of Set B and mark orientation on trend estimation.



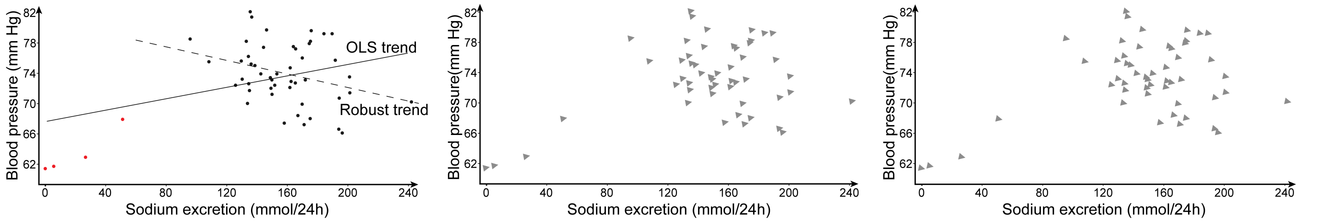
(a) (b)

Figure 18: (a) The accuracies of estimated trend orientation in the study of Correll & Heer [7] and ours under the conditions of different residual bandwidths; (b) the maximal differences of the estimated mark orientations in three experiments.

on visual estimates of trend. For each experiment, we found the pairs of shapes with the highest average differences in estimates: Fig. 18(b) shows such maximal differences with 95% CIs, where the corresponding two most different marks are Tri-0 and Tri-Up in Experiment 1, Tri-R and Circle in Experiment 2, and Tri-Up and Circle in Experiment 3, respectively. In the first two experiments, the largest differences were between marks (Tri-0 and Tri-R, respectively) that directly and accurately encoded trend values, compared to shapes without any such orientation information. This points to the potential power of orientation as an encoding channel for regression by eye.

We summarize our findings with respect to mark orientation as follows:

- Mark orientation that is consistent with a trend line can reduce errors when the residual bandwidth is large.
- In the presence of a large outlying clusters close to the major cluster, mark orientations with an orientation consistent with the major cluster can reduce visual estimation errors, down-weighting the importance of the outlying cluster.



(a) (b) (c)

Figure 19: The scatterplots with different visual marks for studying the relation between sodium excretion and blood pressure. (a) The points with outliers shown in red and the OLS and robust trend lines; (b,c) two scatterplots with two different visual marks, Tri-O (b) and Tri-R(c).

- When two clusters are intermingled with similar densities and slopes, mark orientations consistent with one cluster can bias estimates towards this cluster.

However, the bias introduced by data-driven marks is small when the trend line is more certain. This is the case if the residual bandwidth is small, outliers are far away from the major cluster or have different densities, or if the intermingled clusters have radically different slopes. In such cases there is little ambiguity in the trend, and so estimates are accurate even without guidance from mark orientation.

Implications for Scatterplot Design. The above findings suggest that mark orientation can be a visual cue to implicitly encode trend information. A trend line directly encodes the model in a dogmatic way, whereas data-driven marks act as less dogmatic guidance that can be ignored or downweighted when they seem inappropriate or are clearly at odds with the depicted data, functioning as a form of "implicit uncertainty visualization" [6]. We have shown viewers will, in their own visual estimates, ignore things like outliers that would be a problem for explicit trend lines based on models of insufficient complexity. Fig. 19 shows how data-driven marks could be used to disambiguate trends in a health care dataset for investigating how the blood pressure is related to sodium excretion [11]. We can see that the slopes of the OLS trend and the robust trend have opposing signs (see Fig. 19(a)). Using data-driven marks can suggest a resolution to this ambiguity for the reviewer by downweight (b) or reinforcing (c) the contribution of the outlying points to the overall trend. Although it is hard to make a fair comparison between such implicit and explicit trend visualizations (for instance, Correll & Heer [7] believed that participants would adhere to explicit trend lines so closely if present that they used such stimuli as action checks), we believe that data-driven orientation can assist in regression by eye, especially for cases where standard models like OLS are inappropriate, or there are multiple valid interpretations of the data (as in Xiong et al. [34]).

In summary, if we do not want to guide users towards a trend or the trend is clear, we suggest to use the most common visual marks: circles. But, if we think that the trend is an important component of the data that should be conveyed, and this trend would be difficult to parse without guidance, then chart designers should be encouraged to encode trend via mark orientations.

8 CONCLUSION & FUTURE WORK

In this paper, we study the impact of different data features on visual estimations of trends in scatterplots, and how mark orientation, especially data-driven mark orientation, can impact these estimates. Our results show that participants are sensitive to features like dispersion, outlying clusters, and local density in ways that can systematically differ from common linear models of data. We also find that, in cases where the trend is ambiguous or uncertain, data-driven marks can be used to influence trend estimates.

Our interview results show that participants report that they estimate trends by combining bounding boxes and local densities. Many of our participants reported mentally drawing a bounding box for all points, and then removing outliers. Since the trend is not strictly determined by these perceptual proxies [14, 18], we cannot directly compute how these global or local visual structures were used. In the future, we will quantitatively model how the estimated trend is related to these proxies and to model a larger set of mark shapes and sizes. Since participants might have used different strategies for dealing with outliers (e.g., considering or discarding) in Experiment 2, the effects of outliers are likely to differ based on the individual strategy . In the future, we will model individual differences and preform a secondary analysis of the experimental data as did by Kay and Heer [15]. In addition, we are particularly interested in how the mark size (and its influence on visual artifacts like overplotting and occlusion) could skew assessments of local density and the resulting estimated trend. We are also interested in expanding our study of data-driven orientation to wider use cases, especially under additional "adversarial" scatterplot conditions [18] (such as marks with transparency, high degrees of overplotting, non-linear or non-uniform trends, or where likely perceptual proxies like bounding boxes are uninformative).

ACKNOWLEDGMENTS

This work is supported by the grants of the NSFC (61772315, 61861136012), the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (No.VRLAB2020C08), and the CAS grant (GJHZ1862).

REFERENCES

[1] Francis J Anscombe. 1973. Graphs in Statistical Analysis. *The American Statistician* 27, 1 (1973), 17–21. https://doi.org/10.1080/00031305.1973.10478966

[2] Leilani Battle, Peitong Duan, Zachery Miranda, Dana Mukusheva, Remco Chang, and Michael Stonebraker. 2018. Beagle: Automated extraction and interpretation of visualizations from the web. In *ACM CHI Conf. Human Fact. Comp. Sys.* ACM, 594. https://doi.org/10.1145/3173574.3174168

[3] Rongzheng Bian, Yumeng Xue, Liang Zhou, Jian Zhang, Baoquan Chen, Daniel Weiskopf, and Yunhai Wang. 2021. Implicit Multidimensional Projection of Local Subspaces. *IEEE Trans. Vis. & Comp. Graphics* (2021). https://doi.org/10.1109/TVCG.2020.3030368

[4] David Burlinson, Kalpathi Subramanian, and Paula Goolkasian. 2017. Open vs. Closed Shapes: New Perceptual Categories? *IEEE Trans. Vis. & Comp. Graphics* 24, 1 (2017), 574–583. https://doi.org/10.1109/TVCG.2017.2745086

[5] William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554. https://doi.org/10.1080/01621459.1984.10478080

[6] Michael Correll and Michael Gleicher. 2015. Implicit Uncertainty Visualization: Aligning Perception and Statistics. In *Proceedings of the 2015 Workshop on Visualization for Decision Making Under Uncertainty*.

[7] Michael Correll and Jeffrey Heer. 2017. Regression by eye: Estimating trends in bivariate visualizations. In *ACM CHI Conf. Human Fact. Comp. Sys.* ACM, 1387–1396. https://doi.org/10.1145/3025453.3025922

[8] Çağatay Demiralp, Michael S Bernstein, and Jeffrey Heer. 2014. Learning perceptual kernels for visualization design. *IEEE Trans. Vis. & Comp. Graphics* 20, 12 (2014), 1933–1942. https://doi.org/10.1109/TVCG.2014.2346978

[9] Evanthia Dimara, Gilles Bailly, Anastasia Bezerianos, and Steven Franconeri. 2018. Mitigating the attraction effect with visualizations. *IEEE Trans. Vis. & Comp. Graphics* 25, 1 (2018), 850–860. https://doi.org/10.1109/TVCG.2018.2865233

[10] Evanthia Dimara, Anastasia Bezerianos, and Pierre Dragicevic. 2016. The attraction effect in information visualization. *IEEE Trans. Vis. & Comp. Graphics* 23, 1 (2016), 471–480. https://doi.org/10.1109/TVCG.2016.2598594

[11] Intersalt Cooperative Research Group. 1988. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. *BMJ: British Medical Journal* (1988), 319–328. https://doi.org/10.1136/bmj.297.6644.319

[12] Lane Harrison, Fumeng Yang, Steven Franconeri, and Remco Chang. 2014. Ranking Visualizations of Correlation Using Weber's Law. *IEEE Trans. Vis. & Comp. Graphics* 20, 12 (2014), 1943–1952. https://doi.org/10.1109/TVCG.2014.2346979

[13] Kevin Hu, Snehalkumar'Neil'S Gaikwad, Madelon Hulsebos, Michiel A Bakker, Emanuel Zgraggen, César Hidalgo, Tim Kraska, Guoliang Li, Arvind Satyanarayan, and Çağatay Demiralp. 2019. Viznet: Towards a large-scale visualization learning and benchmarking repository. In *ACM CHI Conf. Human Fact. Comp. Sys.* 1–12. https://doi.org/10.1145/3290605.3300892

[14] Nicole Jardine, Brian D Ondov, Niklas Elmqvist, and Steven Franconeri. 2019. The Perceptual Proxies of Visual Comparison. *IEEE Trans. Vis. & Comp. Graphics* 26, 1 (2019), 1012–1021. https://doi.org/10.1109/TVCG.2019.2934786

[15] Matthew Kay and Jeffrey Heer. 2015. Beyond Weber's Law: A Second Look at Ranking Visualizations of Correlation. *IEEE Trans. Vis. & Comp. Graphics* 22, 1 (2015), 469–478. https://doi.org/10.1109/TVCG.2015.2467671

[16] Jing Li, Jarke J van Wijk, and Jean-Bernard Martens. 2009. Evaluation of Symbol Contrast in Scatterplots. In *2009 IEEE Pacific Visualization Symposium*. IEEE, 97–104. https://doi.org/10.1109/PACIFICVIS.2009.4906843

[17] Min Lu, Shuaiqi Wang, Joel Lanir, Noa Fish, Yang Yue, Daniel Cohen-Or, and Hui Huang. 2019. Winglets: Visualizing Association with Uncertainty in Multi-class Scatterplots. *IEEE Trans. Vis. & Comp. Graphics* 26, 1 (2019), 770–779. https://doi.org/10.1109/TVCG.2019.2934811

[18] Brian D Ondov, Fumeng Yang, Matthew Kay, Niklas Elmqvist, and Steven Franconeri. [n.d.]. Revealing Perceptual Proxies with Adversarial Examples. *IEEE Trans. Vis. & Comp. Graphics* ([n. d.]). https://doi.org/10.1109/TVCG.2020.3030429 Accessed Oct. 23, 2020.

[19] Jason Osborne and Elaine Waters. 2002. Four assumptions of multiple regression that researchers should always test. *Practical assessment, research & evaluation* 8, 2 (2002), 1–9. https://doi.org/10.7275/r222-hv23

[20] Ronald A Rensink. 2014. On the Prospects for a Science of Visualization. In *Handbook of human centric visualization*. Springer, 147–175. https://doi.org/10.1007/978-1-4614-7485-2

[21] Ronald A Rensink and Gideon Baldridge. 2010. The Perception of Correlation in Scatterplots. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 1203–1210. https://doi.org/10.1111/j.1467-8659.2009.01694.x

[22] Alper Sarikaya and Michael Gleicher. 2018. Scatterplots: Tasks, Data, and Designs. *IEEE Trans. Vis. & Comp. Graphics* 24, 1 (2018), 402–412. https://doi.org/10.1109/tvcg.2017.2744184

[23] Michael Sedlmair, Andrada Tatu, Tamara Munzner, and Melanie Tory. 2012. A taxonomy of visual cluster separation factors. *Computer Graphics Forum* 31, 3pt4 (2012), 1335–1344. https://doi.org/10.1111/j.1467-8659.2012.03125.x

[24] L. Shao, A. Mahajan, T. Schreck, and D. J. Lehmann. 2017. Interactive Regression Lens for Exploring Scatter Plots. *Computer Graphics Forum* 36, 3 (2017), 157–166. https://doi.org/10.1111/cgf.13176

[25] Varshita Sher, Karen G Bemis, Ilaria Liccardi, and Min Chen. 2017. An Empirical Study on the Reliability of Perceiving Correlation Indices Using Scatterplots. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 61–72. https://doi.org/10.1111/cgf.13168

[26] Stephen Smart and Danielle Albers Szafir. 2019. Measuring the Separability of Shape, Size, and Color in Scatterplots. In *ACM CHI Conf. Human Fact. Comp. Sys.* 1–14. https://doi.org/10.1145/3290605.3300899

[27] Danielle Albers Szafir, Steve Haroz, Michael Gleicher, and Steven Franconeri. 2016. Four Types of Ensemble Coding in Data Visualizations. *Journal of Vision* 16, 5 (2016), 11–11. https://doi.org/10.1167/16.5.11

[28] Justin Talbot, John Gerth, and Pat Hanrahan. 2012. An empirical model of slope ratio comparisons. *IEEE Trans. Vis. & Comp. Graphics* 18, 12 (2012), 2613–2620. https://doi.org/10.1109/TVCG.2012.196

[29] Lothar Tremmel. 1995. The Visual Separability of Plotting Symbols in Scatterplots. *Journal of Computational and Graphical Statistics* 4, 2 (1995), 101–112. https://doi.org/10.1080/10618600.1995.10474669

[30] Susan VanderPlas and Heike Hofmann. 2015. Signs of the Sine Illusion - Why We Need to Care. *Journal of Computational and Graphical Statistics* 24, 4 (2015), 1170–1190. https://doi.org/10.1080/10618600.2014.951547

[31] Emily Wall, John Stasko, and Alex Endert. 2019. Toward a Design Space for Mitigating Cognitive Bias in Vis. In *Proc. IEEE Conf. on Visualization*. IEEE, 111–115. https://doi.org/10.1109/VISUAL.2019.8933611

[32] Yunhai Wang, Zeyu Wang, Tingting Liu, Michael Correll, Zhanglin Cheng, Oliver Deussen, and Michael Sedlmair. 2019. Improving the Robustness of Scagnostics. *IEEE Trans. Vis. & Comp. Graphics* 26, 1 (2019), 759–769. https://doi.org/10.1109/TVCG.2019.2934796

[33] Cindy Xiong, Cristina R Ceja, Casimir JH Ludwig, and Steven Franconeri. 2019. Biased Average Position Estimates in Line and Bar Graphs: Underestimation, Overestimation, and Perceptual Pull. *IEEE Trans. Vis. & Comp. Graphics* 26, 1 (2019), 301–310. https://doi.org/10.1109/TVCG.2019.2934400

[34] Cindy Xiong, Lisanne van Weelden, and Steven Franconeri. 2018. Perceiving Graphs as Ambiguous Figures. *Journal of Vision* 18, 10 (2018), 1327–1327.

[35] Cindy Xiong, Lisanne van Weelden, and Steven Franconeri. 2019. The curse of knowledge in visual data communication. *IEEE Trans. Vis. & Comp. Graphics* (2019). https://doi.org/10.1109/TVCG.2019.2917689

[36] Fumeng Yang, Lane T Harrison, Ronald A Rensink, Steven L Franconeri, and Remco Chang. 2018. Correlation judgment and visualization features: A comparative study. *IEEE Trans. Vis. & Comp. Graphics* 25, 3 (2018), 1474–1488. https://doi.org/10.1109/TVCG.2018.2810918

[37] Caroline Ziemkiewicz and Robert Kosara. 2010. Implied dynamics in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces*. 215–222. https://doi.org/10.1145/1842993.1843031

[38] Caroline Ziemkiewicz and Robert Kosara. 2010. Laws of attraction: From perceptual forces to conceptual similarity. *IEEE Trans. Vis. & Comp. Graphics* 16, 6 (2010), 1009–1016. https://doi.org/10.1109/TVCG.2010.174