

# An Empirical Model of Slope Ratio Comparisons

(Corrected February 1, 2013)\*

Justin Talbot, John Gerth, and Pat Hanrahan

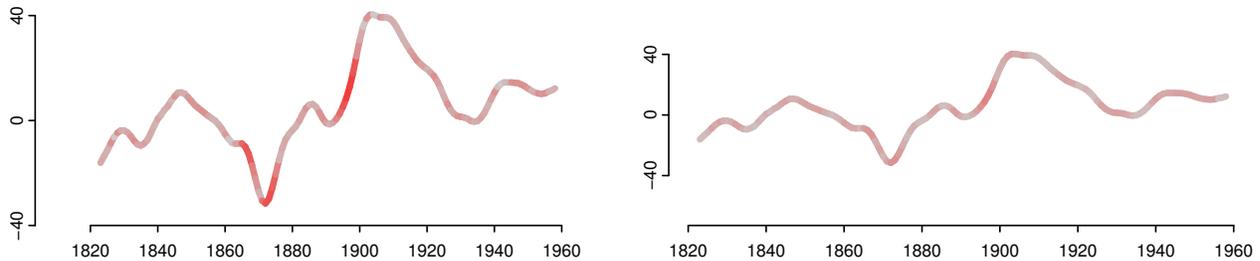


Fig. 1. Both plots show the same data set—the change in the length of a day (in microseconds) over 140 years—with different aspect ratios. The segment redness corresponds to the error that viewers will make when comparing the slope of that segment to all other slopes in the plot—as predicted by our empirical model. The aspect ratio on the right minimizes the total absolute predicted error.

**Abstract**—Comparing slopes is a fundamental graph reading task and the aspect ratio chosen for a plot influences how easy these comparisons are to make. According to *Banking to 45°*, a classic design guideline first proposed and studied by Cleveland et al., aspect ratios that center slopes around 45° minimize errors in visual judgments of slope ratios. This paper revisits this earlier work. Through exploratory pilot studies that expand Cleveland et al.’s experimental design, we develop an empirical model of slope ratio estimation that fits more extreme slope ratio judgments and two common slope ratio estimation strategies. We then run two experiments to validate our model. In the first, we show that our model fits more generally than the one proposed by Cleveland et al. and we find that, in general, slope ratio errors are *not* minimized around 45°. In the second experiment, we explore a novel hypothesis raised by our model: that visible baselines can substantially mitigate errors made in slope judgments. We conclude with an application of our model to aspect ratio selection.

**Index Terms**—Banking to 45 degrees, slope perception, orientation resolution, aspect ratio selection.

## 1 INTRODUCTION

Banking to 45° is a classic design guideline in information visualization and statistical graphics due to Cleveland, McGill, and McGill [5, 4, 3]. It recommends that the aspect ratio of a plot be chosen such that the slopes of the plot’s line segments are centered around 45°. The guideline has been widely and successfully used by visualization designers to manually select aspect ratios. And it has inspired a variety of algorithms [5, 3, 10, 15] that automate this task.

Despite the practical success of this guideline, its perceptual underpinnings remain unclear. Cleveland et al. justified the guideline with an experiment that showed that placing the mid-angle of two lines (the angle halfway between them) at 45° minimizes errors made in judging the ratio of their slopes. However, examination of their experimental design suggests that this conclusion might not be generally applicable. First, their experiment only tested moderate slope ratios and moderate mid-angles. It’s unclear if their observed trends will hold for more extreme slope comparisons. Second, they restricted how their subjects made the slope ratio comparisons, instructing them to compare only the heights (y-extents) of the line segments. Thus, their results may not apply if other comparison methods are also used in practice.

This paper seeks to improve our understanding of slope ratio estimation in line plots through empirical modeling and experimentation. The paper is organized as follows. Section 2 provides details on the original study and other background material. In Section 3 we

\*Our original analysis relied on slightly wrong mid-angles for our experimental stimuli. This revision updates our plots and estimated values to correct this error. Our arguments and conclusions are not affected.

- Justin Talbot is with Stanford, e-mail: jtalbot@stanford.edu.
- John Gerth is with Stanford, e-mail: gerth@graphics.stanford.edu.
- Pat Hanrahan is with Stanford, e-mail: hanrahan@cs.stanford.edu.

describe our methodological approach. Section 4 describes our pilot studies and Section 5 presents a new empirical model for slope ratio judgments. Sections 6 and 7 describe two experimental studies of the new model. The first study demonstrates that our new model fits observed data well. Additionally this model provides strong evidence that the minimum slope ratio estimation error does not occur at 45 degrees. The second study explores a novel hypothesis raised by the new model—that slope ratio estimates are dramatically affected by the presence or absence of a baseline. In Section 8 we briefly discuss the implication of the new model for aspect ratio selection. We conclude with a discussion of the limitations of our study and recommendations for future work.

## 2 PREVIOUS WORK

This section described related work in three categories—perception, aspect ratio selection algorithms, and line chart visualization design.

### 2.1 Perception

In the Cleveland et al. study [5], 16 subjects were each asked to estimate the ratio (as a percentage) between the slopes of 44 pairs of lines with equal x-extents by comparing the y-extents of the two lines. The line pairs were chosen so that the true ratio percentage between their slopes ( $p_{ij}$ ) varied between 50% and 100% and the slopes themselves varied from about 0.1 to slightly more than 1. Subjects were shown each pair of lines for only 2.5s encouraging quick estimates. The resulting estimates,  $\hat{p}_{ij}$ , were then used to fit an empirical model of the absolute estimation error [4]:

$$|\hat{p}_{ij} - p_{ij}| = 4.39 - 0.47(p_{ij} - 100) - 1.14r_{ij} + \epsilon_{ij} \quad (1)$$

where  $p_{ij}$  is the true percentage and  $r_{ij}$  is the angle between the lines in degrees. The model indicated that the absolute error was minimized

when the lines were centered around  $45^\circ$ .

As far as we can tell, there is no other work in perception on slope ratio estimation. However, there is substantial work on related perceptual estimation problems that might provide insight. Fisher [9] found that people under and over estimated the orientation of a single 2D line by as much as 3 degrees. The effect was largest for orientations near  $45^\circ$ . Kennedy et al. [13] showed that angle perception is biased by the lengths of the angle’s legs. The effect was as large as 15 degrees for extremely scalene angles. Recent work on 2D angle perception have explained the traditional effect that small angles are overestimated and big angles are underestimated as a result of a cognitive process that attempts to perceive the angles in 3D instead of 2D. The distribution of angles that occur in the real-world bias this cognitive process to under- or overestimate 2D angles [14, 12]. The effect is up to 2 degrees and is largest for angles near  $20^\circ$ . Fermüller and Malm proposed a unifying image processing-based explanation for many types of illusions [7].

## 2.2 Aspect Ratio Algorithms

Banking to  $45^\circ$  has led to the design of automated algorithms for picking good aspect ratios. Since it is not clear how to extend the Cleveland et al. experimental results from plots with two line segments to plots with many line segments, multiple algorithms have been recommended. In the original paper, it was suggested that the aspect ratio of plots should be chosen such that the median segment slope was 1, an approach they called *median absolute slope* (MS). Cleveland later suggested an alternative method, *length-weighted average orientation* (AWO), that sets the length-weighted average of the absolute segment orientations (the angle made with the horizontal) to  $45^\circ$  [3]. Later, Heer and Agrawala [10] proposed the *global orientation resolution* (GOR) method that selects the aspect ratio by maximizing the sum of squares of the angles between all pairs of segments in the plot. A computationally cheaper approach is to only consider the angles between adjacent pairs of segments. Heer and Agrawala called this approach the *local orientation resolution* (LOR). Recently, Talbot et al. [15] recommended an approach based on minimizing the arc length of the data curve and demonstrated that this approach is more robust than previous work.

## 2.3 Line Chart Visualization

Beattie and Jones [1] examine business graphics in light of the slope ratio hypothesis and conclude that viewer perception of business performance is more accurate when graphs are banked to  $45^\circ$ . Best [2] compare how extrapolation of visual trends varied across a number of different chart types including line charts and find that the error depended upon the trend’s functional form. Correll et al. [6] explore how aggregate comparisons are made in line charts and suggest an alternate visualization based on a color encoding to make these comparisons easier. Horizon plots [8, 11] permit compressing the vertical axis of a line chart without changing the aspect ratio by overlaying horizontal slices of the plot.

## 3 RESEARCH GOALS AND METHODS

Our objective is to better understand the theory behind aspect ratio selection. To do this, we revisit the question raised in the original Cleveland et al. study: how does slope ratio estimation accuracy depend on the true slope ratio and on the aspect ratio of the line segments (expressed as the mid-angle between the segments)? Our approach to answering this question is twofold:

1. Based on insights gained from a series of pilot studies (Section 4), we develop an empirical model of human slope ratio judgments (Section 5).
2. Then, through two formal experiments, we demonstrate that the model fits observed data well and we provide a perceptual interpretation of its components (Sections 6 and 7).

In our model building process and in our experiments we follow the high-level design of the original study. However, we modify the study design to address two major limitations of their work.

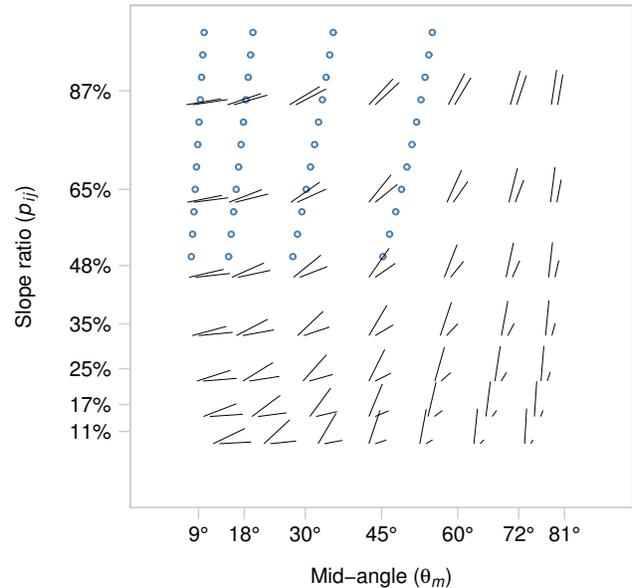


Fig. 2. Space of line comparisons parameterized by mid-angle and a non-linear transformation of the true slope ratio. Line pairs used in this study are plotted here. The spacing on the x-axis is not linear due to a change in our preferred parameterization during post-experiment data analysis. Line pairs used in Cleveland et al. are indicated with blue points.

First, the original design only covers a small portion of the space of slope pairs. In our work, we sample a much larger portion of the space. The range of line pairs studied in our experiment is shown in Figure 2 parameterized by the *mid-angle*,  $\theta_m$ , the angle halfway between the line segments, and by the *slope ratio* (expressed as a percentage),  $p_{ij}$ , between two segments  $i$  and  $j$ . (Due to a change in parameterization part way through our analysis, our mid-angle samples are spaced slightly unevenly.) In comparison, the set of line pairs studied in Cleveland et al., shown with blue points in Figure 2, spans a much smaller range. To keep the sample count manageable, we sample less frequently along the slope ratio-axis than Cleveland et al; however, our pilot studies and Figure 6 in Cleveland and McGill [5] indicate that the slope estimation error function changes slowly in this direction so a lower sampling rate makes better use of limited subject time.

Second, in the Cleveland et al. design, subjects were explicitly instructed to make their slope estimates by comparing the heights (y-extent) of both lines. This approach is exactly correct if the x-extent of both lines is the same. However, we wonder if this is how most visualization users would naturally approach the slope estimation task. To explore how users make slope comparisons in practice, we omit this instruction in our studies, instead allowing subjects to use their own approach to the problem.

## 4 PILOT STUDIES

To help us better understand the slope ratio estimation task, we ran a series of 11 pilot studies on Amazon Turk. We used these informal, rapid studies to quickly iterate on our experimental design and to explore a variety of model possibilities.

For example, in early iterations we used a study design with 100 distinct slope comparison tasks (the product of 10 mid-angles and 10 slope ratios), but found that this resolution was unnecessary since the shape of the response function was relatively smooth. In later iterations, we settled on the design shown in Figure 2 with 49 ( $7 \times 7$ ) slope comparisons, allowing us to run double the number of replications at the same cost. We also found that linear sampling of the slope ratio resulted in many line pairs that were visually indistinguishable. So we modified our design to vary the true slope ratio nonlinearly making the change in the visual angle between line segments equal.

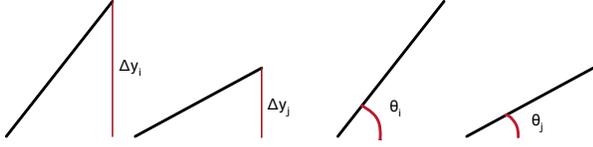


Fig. 3. Our model assumes that people do not make slope ratio judgments directly. Rather they approximate the slope ratio using either the ratio of the line heights or the ratio of the angles each line makes with the horizontal. Both approximations can introduce error into the ratio estimation. This error is captured by the first term in our model.

We paid 3¢ for each slope comparison or about 7 dollars an hour if a participant took 15 seconds per estimate—somewhat on the slow side. A total of 148 workers participated across all 11 studies, some of whom participated in multiple studies, making up to 300 slope comparisons in total. Given the informal nature of these studies we chose to make no effort to screen participants and accepted every non-empty response. This meant that the quality of responses varied greatly, but that we were able to finish studies often in a matter of minutes, making iteration easy and effective. To compensate for the variance in quality, we ran multiple replications (3–10) of each comparison and used the median in our analyses. Each replication was completed by a different subject or set of subjects (since Amazon Turk allows multiple participants to collaborate on completing a replication).

## 5 A MODEL OF SLOPE RATIO COMPARISONS

Our iterative, exploratory use of pilot studies led us to make three assumptions about slope ratio estimation that we want reflected in our model.

First, slope ratios are not estimated directly, rather they are approximated using simpler, more direct visual quantities, such as lengths or angles. The approximation method used results in characteristic *approximation errors* that our model should capture. While Cleveland et al. instructed their subjects to make an approximation based on comparing the y-extents of the two lines (the left half of Figure 3), our pilot study results showed that many participants were making errors consistent with an approximation method based on comparing the angles that the two segments made with the horizontal (the right half of Figure 3). We denote these two approximation strategies HEIGHT and ANGLE.

If the observer uses the HEIGHT approach, then they approximate the slope of a line,  $\tan(\theta_i)$ , with its y-extent,  $\sin(\theta_i) l_i$ , leading to the slope ratio approximation:

$$p_{ij} = \frac{\tan(\theta_i)}{\tan(\theta_j)} \times 100 \approx \frac{\sin(\theta_i) l_i}{\sin(\theta_j) l_j} \times 100$$

This is a good approximation when the x-extents of the two lines are similar.

If the observer uses the ANGLE approach they estimate the slope of the line with the angle the line makes with the horizontal,  $\theta_i$ , leading to the approximation:

$$p_{ij} = \frac{\tan(\theta_i)}{\tan(\theta_j)} \times 100 \approx \frac{\theta_i}{\theta_j} \times 100$$

This is a good approximation when  $\tan(\theta) \approx \theta$  which is true for small angles.

Second, we found that after accounting for the expected approximation error itself, there was an additional consistent source of error that depended on the approximation method used. We hypothesize that this error arises from perceptual sources, such as biases in visual estimates of length or angle, or from higher level cognitive processes, for example, rounding a response to the nearest multiple 5 or 10. We

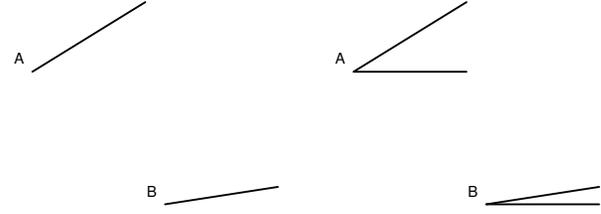


Fig. 4. Stimuli shown to subjects in Experiment 1 (left only) and 2 (both). Not to scale. Experiment 2 evaluates the impact that adding a baseline, making the comparison angle explicit, has on slope ratio estimation accuracy. The vertical and horizontal displacement keeps the observer from taking advantage of spatial alignment, which is known to aid many types of perceptual judgment.

call this the *judgment error*. In the case of the HEIGHT approximation method, the judgment error was roughly constant. For the ANGLE approximation, the judgment error appeared to be a linear function of the mid-angle.

Third, our pilot studies suggested that the overall slope ratio estimation process could be well modeled by a *linear combination* of the approximation method and the judgment error.

Combining these three assumptions, we suggest a model for  $\hat{p}_{ij}$ , an observer’s visual estimate of the slope ratio (expressed as a percentage) between two line segments  $i$  and  $j$ :

$$\hat{p}_{ij} = \begin{cases} \frac{\sin(\theta_i) l_i}{\sin(\theta_j) l_j} \times 100 & +\gamma & +\epsilon_{ij}^h & \text{if HEIGHT} \\ \frac{\theta_i}{\theta_j} \times 100 & +(\mu + \beta \theta_m) & +\epsilon_{ij}^a & \text{if ANGLE} \end{cases} \quad (2)$$

where  $\theta_i$  and  $\theta_j$  are the angles the two lines make with the horizontal,  $\theta_m$  is their mid-angle, the angle halfway between them, and  $\mu$ ,  $\beta$ , and  $\gamma$  are unknown parameters.

The model contains a two sub-models, one for each of approximating methods we observed in our pilot studies. The first term of each sub-model corresponds to the approximation method. The second term captures the *judgment error* and is constant for the HEIGHT sub-model and linear in the mid-angle for the ANGLE model as suggested by our pilot studies. The final term is a normally-distributed random variable that captures all other sources of variation. Note that our model predicts the slope ratio,  $\hat{p}_{ij}$ , unlike the Cleveland model which predicts the absolute error,  $|\hat{p}_{ij} - p_{ij}|$ .

## 6 EXPERIMENT 1: MODEL ESTIMATION AND VALIDATION

The purpose of our first experiment was to gather an independent data set to allow us to validate the form of our model and to estimate its unknown parameters  $\gamma$ ,  $\mu$ , and  $\beta$ .

### 6.1 Method

We asked 8 naïve subjects (4 female, 4 male, all Ph.D. students in visualization or computer graphics) to estimate the slope ratio (as a percentage) of the 49 line pairs shown in Figure 2. Subjects were informed that the true percentage lay between 0% and 100%. Mimicking the original study, our display placed the line segments in the upper-left and lower-right quadrants (Figure 4); this placement ensures that the subject is unable to make direct x- or y-extent comparisons between the lines. As in the original experiment, all pairs of lines had the same x-extent; the upper line, A, was always 200 pixels long and the length of B was varied to maintain the same x-extent as A. Each subject performed the experiment on their own computer, so the absolute size of the lines varied.

Subjects read a short set of instructions before beginning that showed 3 example stimuli (large slope percentage, medium percentage, and small percentage) and gave the correct percentage for each.

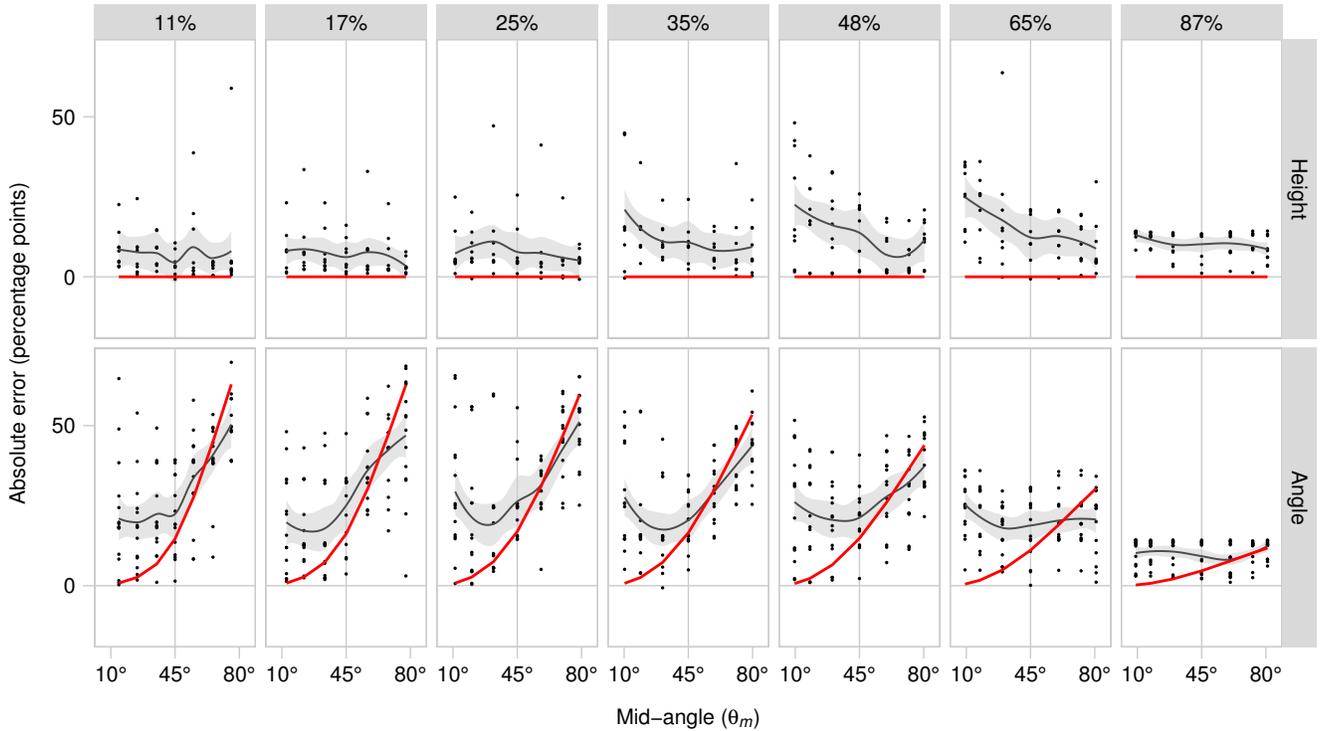


Fig. 5. The observed error ( $\hat{p}_{ij} - p_{ij}$ ) for Experiment 1 (black points and loess curve). The self-reported approximation strategies HEIGHT and ANGLE result in very different errors. The error predicted by the approximation method is shown in red.

Then each subject was asked to evaluate each line pair 3 times for a total of  $7 \times 7 \times 3 = 147$  comparisons. In each replication, the order of presentation was randomized and each replication had a different random order. In contrast to Cleveland et al., we did not attempt to control the timing of the stimuli; we instructed the subjects to make “quick visual estimates”, but each subject advanced at their own pace.

As in the pilot studies, we did not suggest any particular method for making the slope ratio comparison. In particular, we ensured that the instructions did not use the words “height” or “angle”, and we did not inform the participants that the x-extents were matched within line pairs. After each subject completed the study, we performed a follow-up interview in which we asked them if they had used “any specific strategy when making the slope comparisons” and then used this self-reporting to categorize subjects by their dominant approximation method—ANGLE, HEIGHT, or other.

## 6.2 Results

We collected a total of 1,176 responses from the 8 subjects. Most subjects finished the study in 10–20 minutes, indicating a per-task time of 4–8 seconds, including both estimation and response entry time. In comparison, Cleveland et al. showed each stimuli for 2.5 seconds after which it disappeared and the user had an unspecified amount of time to enter their response. We found no significant learning effect.

In the follow-up questioning, 5 subjects (4 female, 1 male) stated that they had primarily made angle comparisons while 3 said that they had primarily used height comparisons. Two of these subjects further stated that they had also tried estimating slopes by zooming into the plot and counting individual pixels. However, both reported quickly abandoning this approach due to its time-consuming nature. Those who reported making HEIGHT comparisons said it had been a conscious choice made after realizing that the x-extents were nearly equal (only 1 out of the 3 HEIGHT subjects reported realizing that the x-extents were exactly equal). In contrast, those who reported making angle comparisons indicated that they did not make a conscious ap-

proximation choice. Rather, all 5 subjects did not realize until the post-study interview that angle and slope estimates would produce different results.

We separated the responses from the self-reported HEIGHT and ANGLE groups for further analysis. The errors for the two groups are plotted in Figure 5 along with a loess curve in black. As predicted by our pilot studies, there are clearly observable differences resulting from the choice of approximation method. The HEIGHT error function is substantially lower and flatter than the ANGLE error which is higher, particularly for larger mid-angles, and distinctly asymmetric. The expected errors due to the HEIGHT and ANGLE approximation methods are shown in red. (Since we constrained the segments to have the same x-extents in all stimuli, the HEIGHT approximation was exact and its expected error is zero.) The expected error is a reasonable first-order approximation to the observed error for both cases. However, there is a substantial second-order error not explained by the approximation approach used. This effect is captured by the *judgment error* terms in our model.

As predicted by our pilot studies, for the HEIGHT subjects, there is an additional roughly constant error. Fitting the HEIGHT sub-model to this data gives an estimate for the parameter  $\gamma$ :

Parameter	Estimate	SE	t-value
$\gamma$	7.45	0.59	12.73

This error is potentially consistent with participants rounding their responses up to the nearest multiple of ten. However, inconsistent with our pilot study-derived model, in slope ratios 35%, 48%, and 65% there is also an obvious downward trend as a function of mid-angle that is not explained by our simple constant model.

For the ANGLE subjects, the approximation method explains some of the error shape, but under-predicts the error for low mid-angles, and somewhat overestimates the error for high mid-angles. To understand the shape of this additional error, we plot the residual (the difference between the observed error and the expected approximation error) as

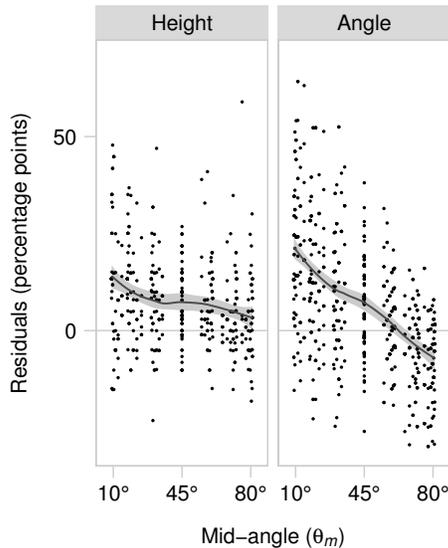


Fig. 6. Residuals of the observed responses in Experiment 1 after accounting for the approximation term in our model, but not the judgment error term. The overlaid loess fit provides evidence that, for the ANGLE approximation approach, the judgment error can be approximated by a linear function of the mid-angle.

a function of the mid-angle (Figure 6) along with a loess fit. There is plausible evidence of a linear trend across the entire range with only mild heteroscedasticity. As in our proposed model, it appears that a linear function of  $\theta_m$  is justified for the ANGLE judgment error term. Thus, we fit a simple linear model with parameters  $\mu$  and  $\beta$ :

Parameter	Estimate	SE	t-value
$\mu$	23.8	1.09	21.84
$\beta$	-0.40	0.02	-18.62

Plots of the residuals from this linear model show a good match to a normal distribution. We also tried fitting a linear model that included the true slope ratio,  $p_{ij}$ , as an additional predictor, but its coefficient estimate was not significantly different from zero and an F-test indicated no significant difference between the two linear models, confirming our decision to not include it the ANGLE model.

The overall ANGLE model, including both the approximation and the perceptual error terms, has  $R^2 = 0.690$ . Models fit with only one of the two error terms had worse  $R^2$  and F-tests indicated that the difference was significant. We also tried fitting a model that included a coefficient on the approximation term (the fit value was 0.94).  $R^2$  was slightly improved (0.692), but not enough, in our opinion, to require complicating the model.

### 6.3 Qualitative Evaluation of Fit

In Figure 7 we compare the absolute error derived from our model’s slope ratio prediction (red), the absolute error directly predicted by the Cleveland model (blue), and a loess curve fit to the observed absolute error (black). Since the Cleveland model was not fit on our data we expect it to fit worse; so we are more interested in whether it matches the overall trend of the observed error.

**Quality of the HEIGHT Submodel** We first focus on the HEIGHT approximation since this was explicitly tested in the Cleveland et al. experiment. Their limited sampling space only included the left halves of the 48%, 65%, and 87% plots (indicated with the thicker blue lines). Note that in these halves, their model (blue) fits the shape of our data pretty well; our subjects appear to have consistently higher error than in the original study. However, when extrapolating the model beyond their sample space the fit is quite poor. Where our data shows a continued downward trend as the mid-angle increases, their model goes back up, creating a false minimum at 45°.

As we further extrapolate their model to the 11%, 17%, and 25% panels, the fit is simply wrong. In these panels our data show a relatively flat error, not the predicted U-shape. Part of the problem is that the downward trend visible in the rightmost four plots, does not occur in the left three. Thus, the Cleveland et al. model is fitting a local, rather than global, trend.

In contrast, our simple constant sub-model does not fit any of the plots particularly well, but the overall fit is better.

**Quality of the ANGLE Submodel** There are two high-level data trends that are important in the ANGLE case. First, as the true slope ratio decreases, the error function becomes increasingly asymmetric, with very large errors for tall-narrow aspect ratios, and small errors for short-wide aspect ratios. Second, this effect leads to a shift in the minimum location from near 45° to below 30°. Our descriptive model captures these trends quite well for most slope percentages. But our model fails to fit the data at the very upper end of the slope percentage (86.7%). The Cleveland et al. model was not designed to handle slope estimation via angle approximation and does not capture the asymmetry.

### 6.4 Discussion

This study has demonstrated that (1) people use both angle and height approximations when making slope judgments and (2) that while the Cleveland et al. model fit our data well in the regions considered in the original study, it fails to extrapolate to either larger mid-angles or smaller true slope percentages. Further, we have seen that slope ratio estimation accuracy is not, in general, minimized at 45°.

Our model does not explain the downward trend in the right four HEIGHT panels in Figure 7 that was observed both in our experiment and in the original study. One speculative hypothesis that might explain this trend is that some HEIGHT subjects may switch to making ANGLE approximations in this region of the slope comparison space, perhaps without realizing it. The approximation error in this region is low, so such a switch makes sense as a cognitive shortcut. But switching also adds the ANGLE judgment error, which is comparatively high in this region.

Another notable failure of our model is in the ANGLE 87% panel. The observed error is roughly flat across all mid-angles while our model predicts increasing error. This may also indicate a place where subjects switched approximation strategies. In this region of the slope comparison space, the lines are nearly parallel. Perhaps subjects switched from angle comparisons to judgments of parallelness with an associated change in the approximation error.

More work is needed to understand these subtle effects.

## 7 EXPERIMENT 2: SOURCE OF ANGLE’S JUDGMENT ERROR TERM

The judgment error term in our model represents the perceptual or cognitive error introduced in the process of making a slope ratio estimate. Our exploratory studies and the first experiment allowed us to empirically fit a linear form to the this term in the ANGLE sub-model. However, the source of this error remains unclear. Experiment 1 indicated that the linear trend being fit by the second term appears only in the ANGLE data, not the HEIGHT data (c.f. Figure 6). Thus, whatever is causing the trend is likely due to the angle estimation process itself.

There has been a large amount of previous work on angle estimation in the perception literature. We attempted to find previously described effects that could explain the error we are seeing. However, the effects we found were either too small in magnitude (we have to explain a 20% deviation from the predicted value—implying an angular misperception of 2-4°) or went in the wrong direction.

But in comparing the stimulus of Experiment 1 to experiments in the angle discrimination literature one property stands out. Our stimulus in Experiment 1 shows only one of the angle’s legs. The other leg (the horizontal base) must be imagined by the subject. We hypothesized that the source of the linear trend in the second term of ANGLE is a misperception of the angle ratio caused by the lack of a visible baseline.

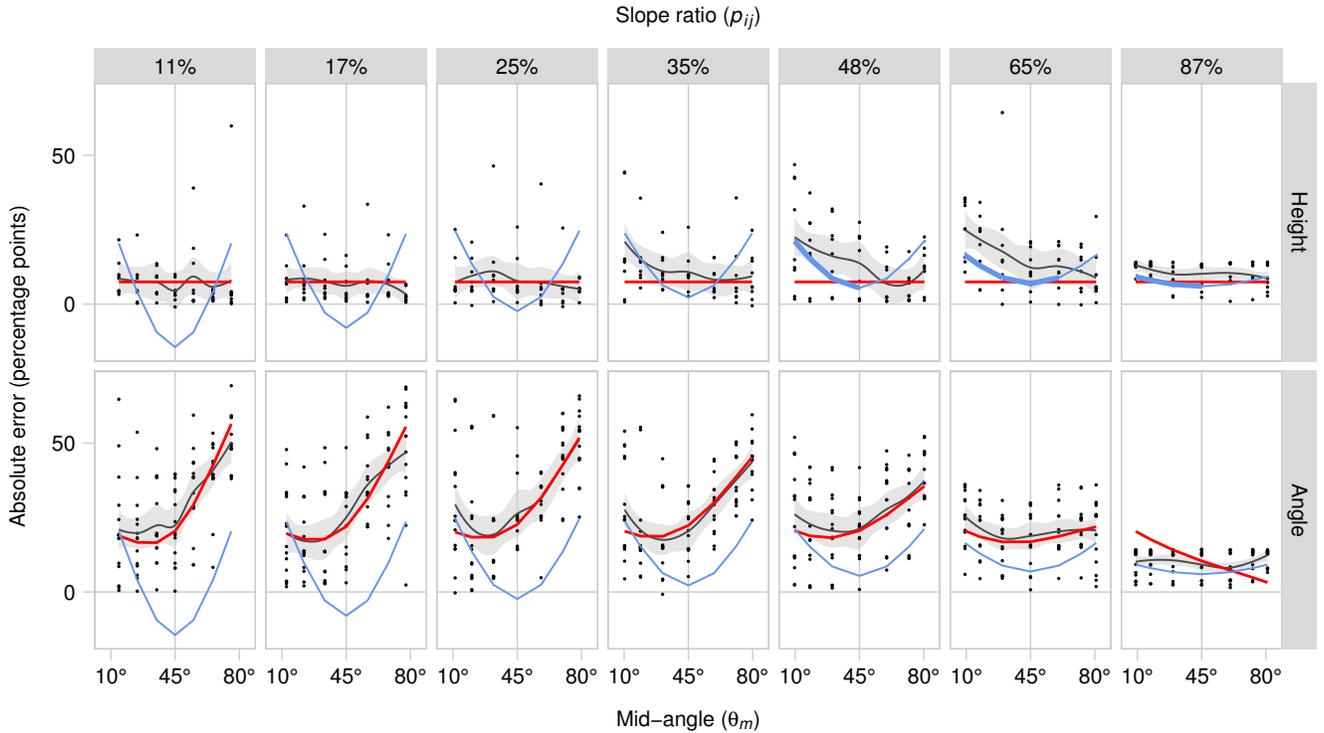


Fig. 7. Comparison of our fit model (red), the Cleveland model [4] (blue), and a loess curve fit to the observed absolute error (black). Our model fits the prominent trends in the `ANGLE` case well, but the simple constant function in the `HEIGHT` case does not fit the downward trends in the right four plots. In the region of the slope comparison space studied in Cleveland et al. (shown in thick blue line), their model matches our data relatively well. However, its shape fails to adequately match the trends seen in the rest of the data. In particular, the assumption that the error will be minimized at a mid-angle of  $45^\circ$  is not generally true.

This second experiment tests this hypothesis by comparing the judgment error made by subjects shown a baseline with the error of those who are not shown a baseline. If our hypothesis is correct, the linear trend in the judgment error should be largely eliminated by the presence of a baseline.

### 7.1 Method

We asked 20 naïve subjects (18 males, 2 females, 19 are Ph.D. students, post-docs, or staff members of a university computer science department, 1 is a Physics Ph.D.) to perform the same task as in Experiment 1, except half were randomly assigned to see line segments without baselines and half to see line segments with baselines (see Figure 4). Additionally, rather than ask participants to make 3 judgments of the same stimulus, as in Experiment 1, we further subdivided the mid-angle axis of the sample space by a factor of 3 in order to get finer-grained estimates of the shape of the error function. Thus each participant still made a total of  $21 \times 7 = 147$  comparisons.

Other than these two changes, the experimental set up followed Experiment 1. Again we did not mention the slope estimation strategies to subjects, but in a follow up interview identified whether they were principally making height or angle comparisons.

### 7.2 Results

We collected a total of 2,940 responses from the 20 subjects. Of the 10 subjects in the baseline condition, 2 subjects reported using height approximation and examination of their response patterns confirmed this. Of the 10 subjects in the without-baseline condition, 2 subjects also reported using height approximation, but examination of their response patterns suggested that only one of them had actually done so. We attributed this to miscommunication between the subject and the interviewer and included their response with the angle data. Addition-

ally, one subject reported using the angle method, but was aware that they were going to overestimate large angles, so mentally applied a correction to their responses. Examination of their responses revealed a pattern completely distinct from either the `ANGLE` or `HEIGHT` predictions. The subject had extremely low error for high mid-angle comparisons, but extremely high error for low mid-angle comparisons. We held this subject’s data out from the current analysis. We found no learning effect.

The response patterns of those who used the height approximation were similar to those seen in Experiment 1. But given the small number of responses, we didn’t further analyze this case.

Figure 8 shows the observed errors for the `ANGLE` subjects in the without and with-baseline cases. The addition of a baseline dramatically improves the error behavior in the low mid-angle regions of the leftmost 5 plots. After subtracting out the approximation term of our model, we again plotted the remaining error as a function of mid-angle (Figure 9). The without-baseline plot is a plausible match to the linear trend found and modeled in Experiment 1 (c.f. left half of Figure 6). However, the baseline case is substantially different. As predicted, the addition of a baseline nearly eliminates the judgment error for mid-angles less than  $45^\circ$ . The error here is now nearly as small as in the height approximation results from Experiment 1. But, unpredicted, the linear trend was not eliminated for mid-angles larger than  $45^\circ$ . Here the trend continues downward as it did before.

### 7.3 Discussion

Our hypothesis is not fully confirmed. The `ANGLE` judgment error was mitigated in part of the domain but was not completely eliminated by the addition of a baseline. However, this result does provide evidence that the judgment error arises at least in part due to errors made in the angle ratio estimation portion of slope estimation.

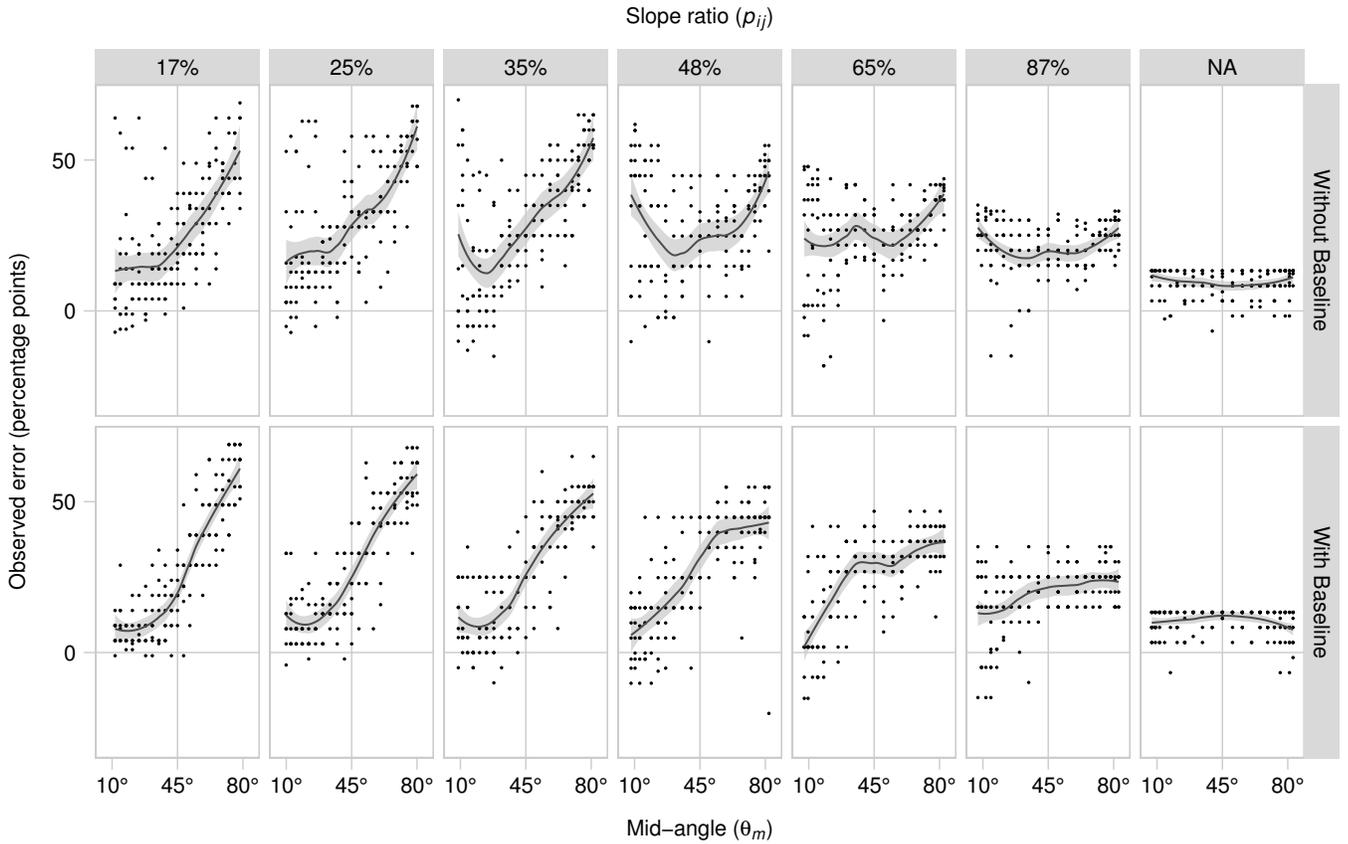


Fig. 8. Comparison of observed error in the `ANGLE` group without and with a visible baseline. The addition of a baseline dramatically decreases the error in the low mid-angle region, indicating that a substantial portion of the `ANGLE` judgment error is due to misestimation of angle size in the absence of a baseline.

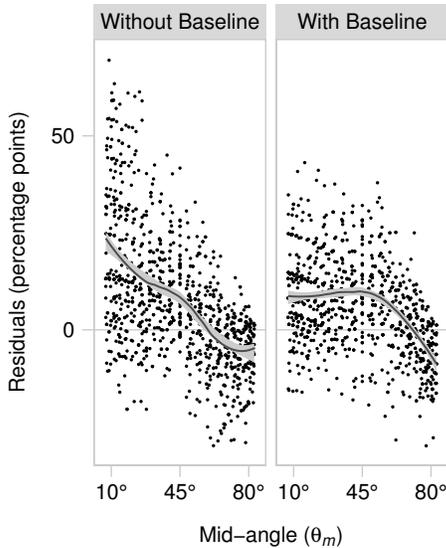


Fig. 9. Plot of the `ANGLE` judgment error in Experiment 2 as a function of mid-angle. The presence of a baseline partially mitigates the error for mid-angles less than  $45^\circ$ , but has no impact on mid-angles larger than  $45^\circ$ .

It's unclear what causes the error for larger mid-angles. One possible hypothesis is that for large slopes observers actually estimate the angle it makes with the vertical, rather than with the horizontal. If this is true, the addition of a vertical baseline could make the error completely disappear. Alternatively, this error appears to arise primarily for steeper angles where our choice to keep x-widths the same led to comparisons between very long and very short lines. Perhaps this length difference, or simply small absolute lengths, leads to this observed error.

## 8 APPLICATION TO ASPECT RATIO SELECTION

The frequent use of the angle approximation for slope comparisons in our study suggests that most people will make large errors when comparing two lines with large slope, suggesting that flatter aspect ratios should be preferred. This, in retrospect, seems obvious, but has not been included in automatic aspect ratio selection algorithms.

To verify this, we use the model proposed in Section 5 with the coefficients fit in Experiment 1 to estimate optimal aspect ratios for a set of time series analyzed previously in Talbot et al. [15]. We selected the aspect ratios by minimizing the sum of the predicted error under the `ANGLE` submodel over all pairs of line segments in each plot. The results are shown in Figure 10. As expected, compared to previously proposed aspect ratio selection algorithms, our empirical model consistently selects flatter, wider aspect ratios.

Our work predicts that the aspect ratios chosen by previous methods will result in higher slope ratio estimation errors; but slope ratio estimation is just one graph reading task among many. For example, we also want visualization users to have low error when making estimates of the function itself (locations of maximums and minimums,

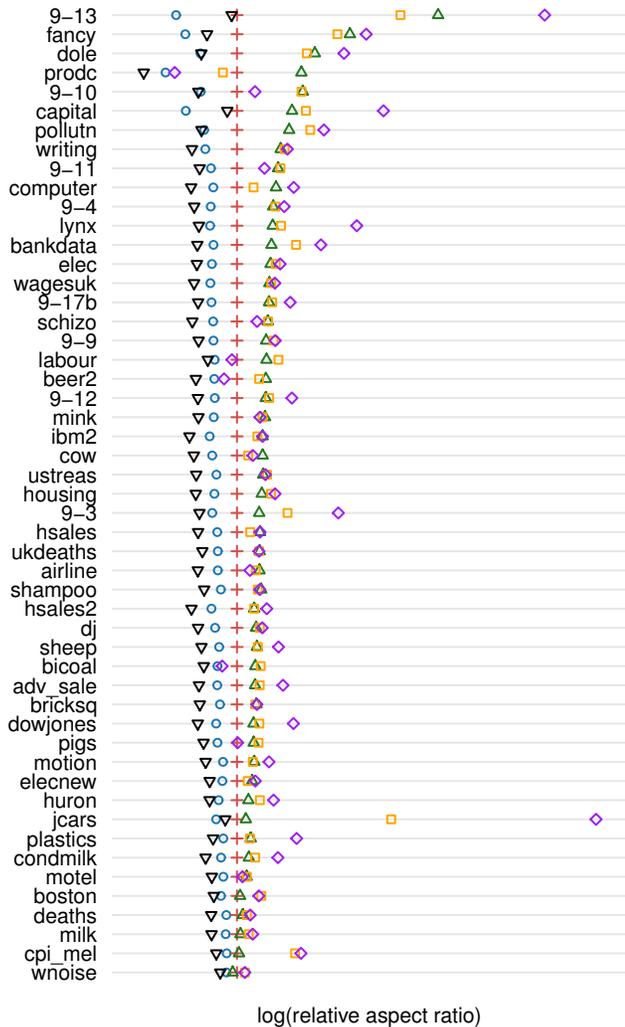


Fig. 10. Comparison of the aspect ratios produced by minimizing the error predicted by our empirical model (black upside down triangles) with previously proposed methods: AWO (circle), arc length (cross), GOR (square), MS (triangle), and LOR (diamond). Our empirical model consistently chooses flatter aspect ratios than other methods.

sign changes, etc.) and of the second derivative (changing rates of growth or inflection points). Any specific aspect ratio necessarily has to make a tradeoff between reducing errors for some tasks and increasing it for others. It’s likely that these previous automated techniques are trading off slope ratio estimation accuracy for better accuracy on something else. Our descriptive model of slope ratio estimation is a first step towards understanding the full set of error functions that should be considered when automatically selecting an aspect ratio.

We have not yet incorporated our baseline results into an automatic aspect ratio selection mechanism. For a plot with gridlines or for a data curve that is close to, or crosses, the x-axis, the baseline results suggest that an aspect ratio selection algorithm could have more flexibility in selecting flatter aspect ratios than otherwise might be possible, because the visible reference line will reduce slope ratio estimation errors.

## 9 DISCUSSION AND FUTURE WORK

The high-level conclusion we draw from this work is that the theory of aspect ratio selection is not as simple as it once seemed. Minimizing the error in slope ratio estimation does not directly lead to a simple design guideline. Substantial future work remains to flesh out a full theory of aspect ratio selection. In previous sections, we called out patterns observed in our data that were not captured by our model.

Teasing apart the low-level details here will require careful experimental work. Further, we did not gather data from the extremes of the comparison space. Having more data from the extremes might provide further insight into the source of the remaining unexplained error. Additionally, the fact that we had few subjects choose to use height comparisons means that we did not gather enough data to further analyze the interesting trends that we saw there.

A major limitation shared by Cleveland et al.’s original study and our study is that both only look at slope comparisons between pairs of lines in isolation. But this is not how graph reading is done; graphs typically consist of many line segments, and frequently graphs have enough line segments that the data curve looks continuous. It is still unclear if the results derived in our studies for pairwise discrete comparisons will transfer to real plots. Additionally, our finding that angle ratio comparisons are more common than height comparisons also needs to be verified using real plots and real plot-reading tasks.

We have presented and validated a new model for slope ratio comparisons. We’ve also shown that the model can generate interesting hypotheses for further exploration and can inform existing aspect ratio selection practice. But there remains substantial work to be done to build a solid understanding of aspect ratio selection.

## ACKNOWLEDGMENTS

The first author was supported by FODAVA grant 0937123 and NSF grant CCF-1111943.

## REFERENCES

- [1] V. Beattie and M. J. Jones. The impact of graph slope on rate of change judgments in corporate reports. *Abacus*, 38(2):177–199, 2002.
- [2] L. A. Best. Visual extrapolation of linear and nonlinear trends: Does the knowledge of underlying trend type affect accuracy and response bias? In T. Sobh, editor, *Advances in Computer and Information Sciences and Engineering*, pages 273–278. Springer Netherlands, 2008.
- [3] W. S. Cleveland. A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics*, 2(4):323–343, 1993.
- [4] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988.
- [5] W. S. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society. Series A (General)*, 150(3):192–229, 1987.
- [6] M. Correll, D. Albers, S. Franconeri, and M. Gleicher. Comparing averages in time series data. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI ’12, pages 1095–1104, New York, NY, USA, 2012. ACM.
- [7] C. Fermler and H. Malm. Uncertainty in visual processes predicts geometrical optical illusions. *Vision Research*, 44(7):727–749, 2004.
- [8] S. Few. Time on the horizon. [http://www.perceptualedge.com/articles/visual\\_business\\_intelligence/time\\_on\\_the\\_horizon.pdf](http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf), June/July 2008.
- [9] G. H. Fisher. An experimental study of linear inclination. *Quarterly Journal of Experimental Psychology*, 26(1):52–62, 1974.
- [10] J. Heer and M. Agrawala. Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12:701–708, 2006.
- [11] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI ’09, pages 1303–1312, New York, NY, USA, 2009. ACM.
- [12] C. Q. Howe and D. Purves. Natural-scene geometry predicts the perception of angles and line orientation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4):1228–1233, 2005.
- [13] G. J. Kennedy, H. S. Orbach, and G. Loffler. Effects of global shape on angle discrimination. *Vision Research*, 46(8-9):1530–1539, 2006.
- [14] S. Nundy, B. Lotto, D. Coppola, A. Shimpi, and D. Purves. Why are angles misperceived? *Proceedings of the National Academy of Sciences*, 97(10):5592–5597, 2000.
- [15] J. Talbot, J. Gerth, and P. Hanrahan. Arc length-based aspect ratio selection. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2276–2282, Dec. 2011.