Perception and Cognitive Implications of Logarithmic Scales for Exponentially Increasing Data: Perceptual Sensitivity Tested with Statistical Lineups

Emily A. Robinson 1
Department of Statistics, California Polytechnic State University - San Luis Obispo
and
Reka Howard 2
Department of Statistics, University of Nebraska - Lincoln
and
Susan VanderPlas 3
Department of Statistics, University of Nebraska - Lincoln

January 11, 2024

Abstract

Logarithmic transformations are a standard solution to displaying data that span several magnitudes within a single graph. This paper investigates the impact of log scales on perceptual sensitivity through a visual inference experiment using statistical lineups. Our study evaluated participant's ability to detect differences between exponentially increasing data, characterized by varying levels of curvature, using both linear and logarithmic scales. Participants were presented with a series of plots and asked to identify the panel that appeared most different from the others. Due to the choice of scale altering the contextual appearance of the data, the results revealed slight perceptual advantages for both scales depending on the curvatures of the compared data. This study serves as the initial part of a three-paper series dedicated to understanding the perceptual and cognitive implications of using logarithmic scales for visualizing exponentially increasing data. These studies serve as an example of multi-modal graphical testing, examining different levels of engagement and interaction with graphics to establish nuanced and specific guidelines for graphical design.

Keywords: log scales, visual inference, graphical testing

1 Introduction

Effective communication of data is critical in influencing people's opinions and actions. This consideration was particularly true during the COVID-19 pandemic, where data visualizations and dashboards were vital in informing the public and policymakers about the outbreak's status. Local governments relied on graphics to inform their decisions about shutdowns and mask mandates, while residents were presented with data visualizations to encourage compliance with these regulations. A major issue designers encountered when creating COVID-19 plots was how to display data from a wide range of values ~~[8, 5]~~ [8, **(author?)** [5]]. When faced with data that span several orders of magnitude, we must decide whether to show the data on its original scale (compressing the smaller magnitudes into a relatively small area) or to transform the scale and alter the contextual appearance of the data. Log axis transformations have emerged as a standard solution to this challenge, as they allow for the display of data over several orders of magnitude within a single graph.

Exponential data is one such example of a function that compresses smaller magnitudes into a smaller area; Fig. 1 presents hard drive capacity over the past forty years on both the linear and log scales to demonstrate the usefulness of log scales when dealing with data spanning multiple magnitudes. Logarithms facilitate the conversion of multiplicative relationships (displaying 1 & 10 with a distance of 10 units apart and displaying 10 & 100 with a distance of 90 units apart) to additive relationships (displaying 1 & 10 and 10 & 100 an equal distance apart), highlighting proportional relationships and linearizing power functions [17]. Logarithms also have practical applications, simplifying the computation of small numbers such as likelihoods and transforming data to conform to statistical assumptions. Although log scales have a long history of use in fields such as ecology, psychophysics, engineering, and physics [9, 24], there is still a need to understand the implications of their

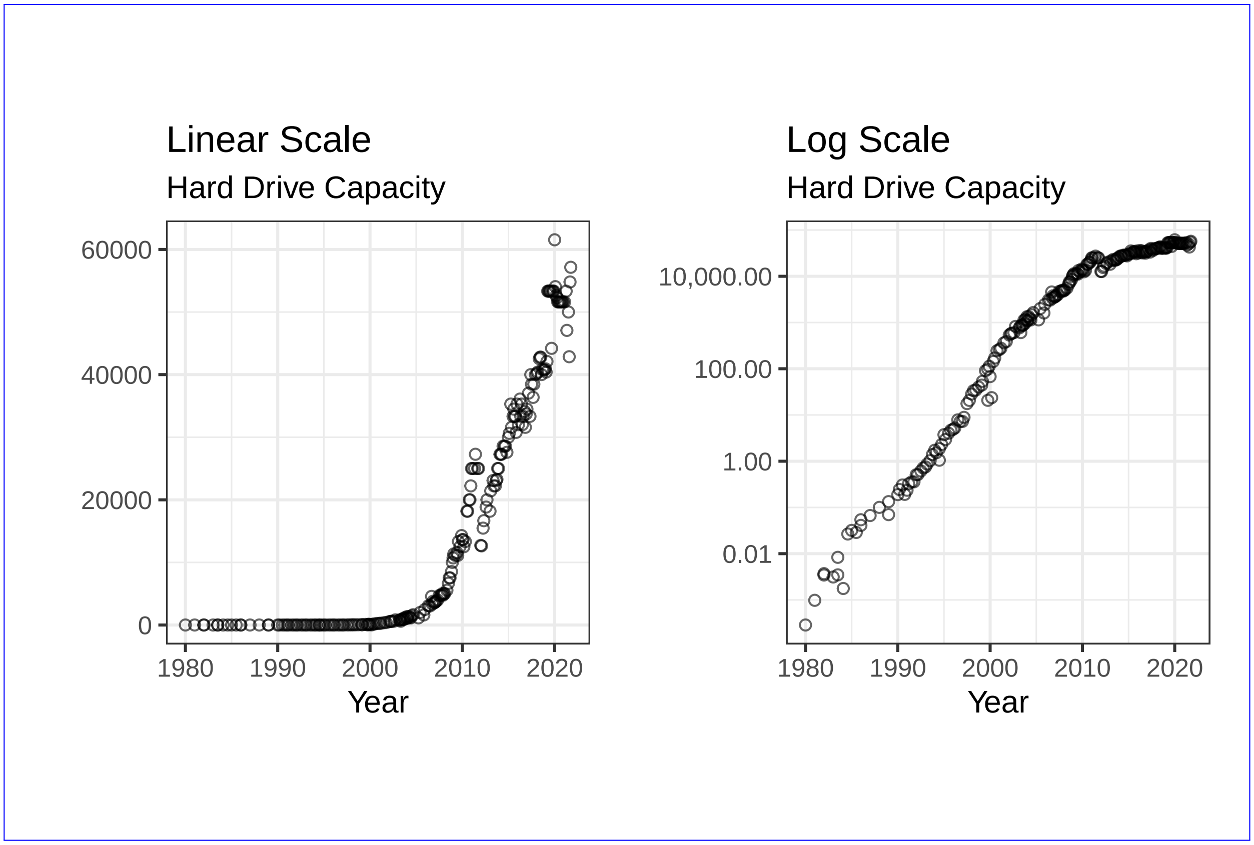


Figure 1: These plots present hard drive capacity over the past forty years on both the linear and log scale and illustrate the use of the log scale when displaying data which spans several magnitudes.

use and provide best practices for their implementation.

Apart from the biases resulting from using log scales, there is a general misinterpretation of exponential growth. Early stages of exponential growth often appear to have a small growth rate, while the middle stage seems to exhibit more quadratic growth. It is only in the later stages that the exponential growth becomes apparent. Fig. 2 highlights the three stages and associated appearances of exponential growth at each stage [23]. This misinterpretation can lead to decisions made under inaccurate understanding, resulting in potential consequences.

Previous studies have explored the estimation and prediction of exponential growth

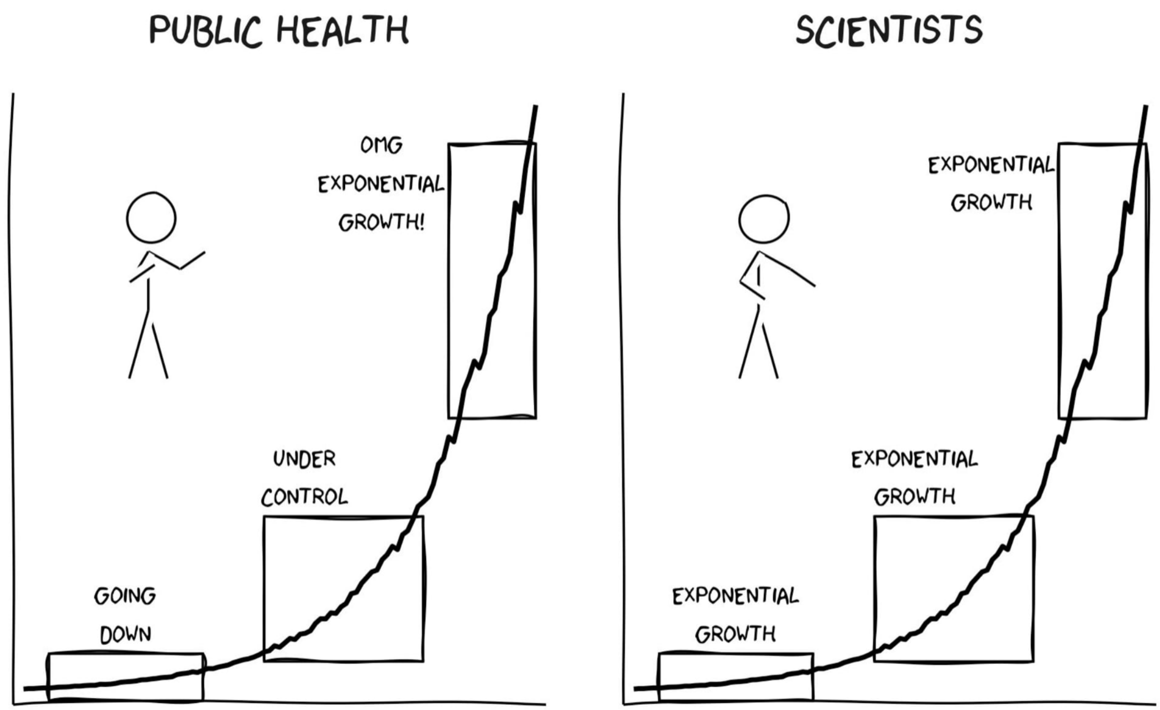


Figure 2: This figure highlights the three stages and associated appearances of exponential growth at each stage. Early stages of exponential growth often appear to have a small growth rate, while the middle stage seems to exhibit more quadratic growth. It is only in the later stages that the exponential growth becomes apparent.

and found that individuals often underestimate exponential growth when presented values numerically and graphically [25]. The hierarchy of plot objects, such as lengths and angles, as described by **(author?)** [7], offers a possible explanation for the underestimation observed in exponentially increasing trends. Experiments conducted by **(author?)** [25], **(author?)** [11], and **(author?)** [15] aimed to improve estimation accuracy for exponential growth. While contextual knowledge or experience did not enhance estimation, instruction on exponential growth reduced underestimation by prompting participants to adjust their initial starting value [25, 11]. Furthermore, providing immediate feedback to participants about the accuracy of their predictions improved estimation [15].

Log transforming the data may address our inability to predict exponential growth accurately. However, this transformation introduces new complexities, as most readers may need to be mathematically sophisticated enough to intuitively understand logarithmic math and translate it back into real-world effects. ~~**(author?)** [17] surveyed ecologists to determine their encounter frequency and~~ Despite the transformative power of logarithmic scales in facilitating accurate data representation, **(author?)** [17]'s survey of ecologists highlights the challenges associated with the widespread comprehension of log-scaled data~~in the literature. The authors identified misconceptions encountered when presented with data on log-log scales which arise~~. Notably, the study identifies prevalent misconceptions arising from linear extrapolation assumptions in log-log space, ~~neglecting~~ a factor that often leads to neglect of the underlying exponential ~~relationship~~ relationships in linear-linear space.

~~In this paper, we evaluated the benefits and drawbacks of using log scales and examine their impact on perceptual sensitivity by conducting a visual inference experiment using statistical lineups [4]. The experiment focused on a participant's ability to identify differences between exponentially increasing curves with varying levels of curvature using both linear~~

~~and log scales. The study did not require participants receive any mathematical training or have prior understanding of exponential growth or logarithmic scales, focusing instead on the fundamental nature of the ability to identify differences in charts.~~

Building upon the need for a nuanced understanding of data representation, **(author?)** [4] introduced statistical lineups as a framework for statistical inference and graphical tests. Statistical lineups treat a data plot as a visual statistic, summarizing the data as a numerical function or mapping. Evaluation of a panel in a statistical lineup requires visual inspection by a person, and if visual evaluations lead to different results, two visualization methods are deemed significantly different. Recent studies have utilized statistical lineups to quantify the perception of graphical design choices [10, 14, 13, 22]. Statistical lineups provide an elegant way of combining perception and statistical hypothesis testing through graphical experiments [16, 20, 26].

The term 'lineup' is an analogy to police lineups in criminal investigations, where witnesses identify the criminal from a group of individuals. Similarly, researchers present a statistical lineup plot consisting of smaller panels and ask the viewer to identify the panel that contains the actual data from a set of decoy null plots. Researchers generate null plots containing data generated according to a prespecified hypothesis using permutation or simulation. Typically, a statistical lineup consists of 20 panels, with one target panel and 19 null panels. If the viewer can identify the target panel from the null panels, it suggests that the actual data is visually distinct from the data generated under the null model. ~~Fig. 5 presents examples of statistical lineups. The statistical lineup on the left presents increasing exponential data displayed on a linear scale with panel (5 × 2) + 3 as the target. The lineup on the right shows increasing exponential data plotted on a log scale with panel 2 × 2 as the target.~~

While explicit graphical tests direct the participant to a specific feature of a plot to answer a particular question, implicit graphical tests require the user to identify both the purpose and function of the plot in order to evaluate the plots shown. Furthermore, implicit graphical tests, such as lineups, simultaneously test for multiple visual features, including outliers, clusters, and linear and nonlinear relationships [21]. Researchers can collect responses from multiple viewers using crowd-sourcing websites such as Prolific and Amazon Mechanical Turk.

In this paper, our primary focus is to evaluate the benefits and drawbacks of using log scales, specifically delving into their impact on perceptual sensitivity towards the degree of curvature. To address this, we conducted a visual inference experiment employing statistical lineups [4]. Although our findings could have broad applications to various functions resulting in curvature, our experiment deliberately centered on participants' ability to identify differences in the curvature of exponentially increasing curves when presented with both linear and log scales. We discuss the nuances and challenges of testing the perception of exponential growth in the appendix. Importantly, this investigation did not necessitate participants to undergo mathematical training or possess a prior understanding of exponential growth or logarithmic scales. Instead, it aimed to unravel the inherent ability to identify differences in curvature within charts, focusing on the fundamental nature of visual perception.

~~The lineup plot on the left displays increasing exponential data on a linear scale with panel (2 x 5) + 3 as the target. The lineup plot on the right displays increasing exponential data on the log scale with panel 2 x 2 as the target.~~

In Section 2 we describe the participant sample, the graphical task, data generation process, and study design. Section 3 describes the participant data collected and shares results from the statistical analyses of the data using a generalized linear mixed model. We present overall conclusions and discussion of the results in Section 4, and provide an overview of future related papers. The Supplementary Material includes a link to the RShiny data collection applet, participant data used for analysis, and code to replicate the analysis. The results of this study lay the groundwork for further exploration of the implications of using log scales in data visualization.

2 Study Development and Methods

[18]

2.1 Data Generation

In this study, we simulated data from an exponential model to generate the target and null data sets; the models between panels differ in the parameter values selected for the null and target panels. In order to guarantee the simulated data spans the same domain and range of values for each statistical lineup panel, we began with a domain constraint of $x \in [0, 20]$ and a range constraint of $y \in [10, 100]$ with $N = 50$ points randomly assigned throughout the domain. We mapped the randomly generated x values to a corresponding y value based on an exponential model with predetermined parameter values and multiplicative random errors to simulate the response. These constraints assure that participants who select the target panel are doing so because of their visual perception differentiating between curvature or growth rate rather than different starting or ending values.

We simulated data based on a three-parameter exponential model with multiplicative errors:

$$y_i = \alpha \cdot e^{\beta \cdot x_i + \epsilon_i} + \theta \tag{1}$$

$$\text{with } \epsilon_i \sim N(0, \sigma^2).$$

The parameters α and θ were adjusted based on β and σ^2 to guarantee the range and domain constraints are met. The model generated $N = 50$ points $(x_i, y_i), i = 1, ..., N$ where x and y have an increasing exponential relationship. The heuristic data generation procedure is described in Algorithm 1 and Algorithm 2.

Algorithm 1 Lineup Parameter Estimation

- **Input Parameters:** domain $x \in [0, 20]$, range $y \in [10, 100]$, midpoint x_{mid}.
- **Output Parameters:** estimated model parameters $\hat{\alpha}, \hat{\beta}, \hat{\theta}$.

1: ~~Determine the~~ In order to obtain the two middle points (total of four points for estimating three parameters), determine the $y = -x$ line scaled to fit the assigned domain and range.

2: Map the values $x_{mid} - 0.1$ and $x_{mid} + 0.1$ to the $y = -x$ line for the two additional points.

3: From the set of points (x_k, y_k) for $k = 1, 2, 3, 4$, calculate the coefficients from the linear regression model $\ln(y_k) = b_0 + b_1 x_k$ to obtain starting values for $\alpha_0 = e^{b_0}, \beta_0 = b_1, \theta_0 = 0.5 \cdot \min(y)$

4: Using the `nls` function from the base `stats` package in Rstudio [@Rstudio] and the starting parameter values - $\alpha_0, \beta_0, \theta_0$ - fit the nonlinear model, $y_k = \alpha \cdot e^{\beta \cdot x_k} + \theta$ to get estimated parameter values for $\hat{\alpha}, \hat{\beta}, \hat{\theta}$.

Algorithm 2 Lineup Exponential Data Simulation

- **Input Parameters:** sample size $N = 50$, estimated parameters $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$, from Algorithm 1, and standard deviation σ from the exponential curve.
- **Output Parameters:** N points, in the form of vectors $\mathbf{x}$ and $\mathbf{y}$.

1: Generate $\tilde{x}_j, j = 1, ..., \frac{3}{4}N$ as a sequence of evenly spaced points in $[0, 20]$. This ensures the full domain of x is used, fulfilling the constraints of spanning the same domain and range for each parameter combination.

2: Obtain $\tilde{x}_i, i = 1, ..., N$ by sampling $N = 50$ values from the set of $\tilde{x}_j$ values. This guarantees some variability and potential clustering in the exponential growth curve disrupting the perception due to continuity of points.

3: Obtain the final x_i values by jittering $\tilde{x}_i$.

4: Calculate $\tilde{\alpha} = \frac{\hat{\alpha}}{e^{\sigma^2/2}}$. This ensures that the range of simulated values for different standard deviation parameters has an equal expected value for a given rate of change due to the non-constant variance across the domain.

5: Generate $y_i = \tilde{\alpha} \cdot e^{\hat{\beta} x_i + e_i} + \hat{\theta}$ where $e_i \sim N(0, \sigma^2)$.

Table 1: Lineup data simulation final parameters

Curvature Level	x_{mid}	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\sigma}$
High	14.5	0.91	0.88	0.23	9.10	0.25
Medium	13.0	6.86	6.82	0.13	3.14	0.12
Low	11.5	37.26	37.22	0.06	-27.26	0.05

2.2 Parameter Selection

We chose three levels of trend curvature (low curvature, medium curvature, and high curvature). For each curvature level, we simulated 1,000 data sets of (x_{ij}, y_{ij}) points for $i = 1, ..., 50$ increments of x-values and replicated $j = 1, ..., 10$ corresponding y-values per x-value. Each generated x_i point from Algorithm 2 was replicated ten times. We fit a linear regression model on each of the individual data sets and computed the lack of fit statistic (LOF) which measures the deviation of the data from the linear regression model. After obtaining the LOF statistic for each level of curvature, we evaluated the density plots (Fig. 3) to provide a metric for differentiating between the curvature levels and thus detecting the target plot. While the LOF statistic provides a numerical value for discriminating between the difficulty levels, it cannot be directly related to the perceptual discriminability; it serves primarily as an approximation to ensure that we are testing parameters at several distinct curvature levels. Table 1 lists the final parameters used for data simulation.

2.3 Lineup Setup

To generate the small multiple scatter plots for the statistical lineups shown to participants in the study, we simulated a single data set corresponding to curvature level A for the

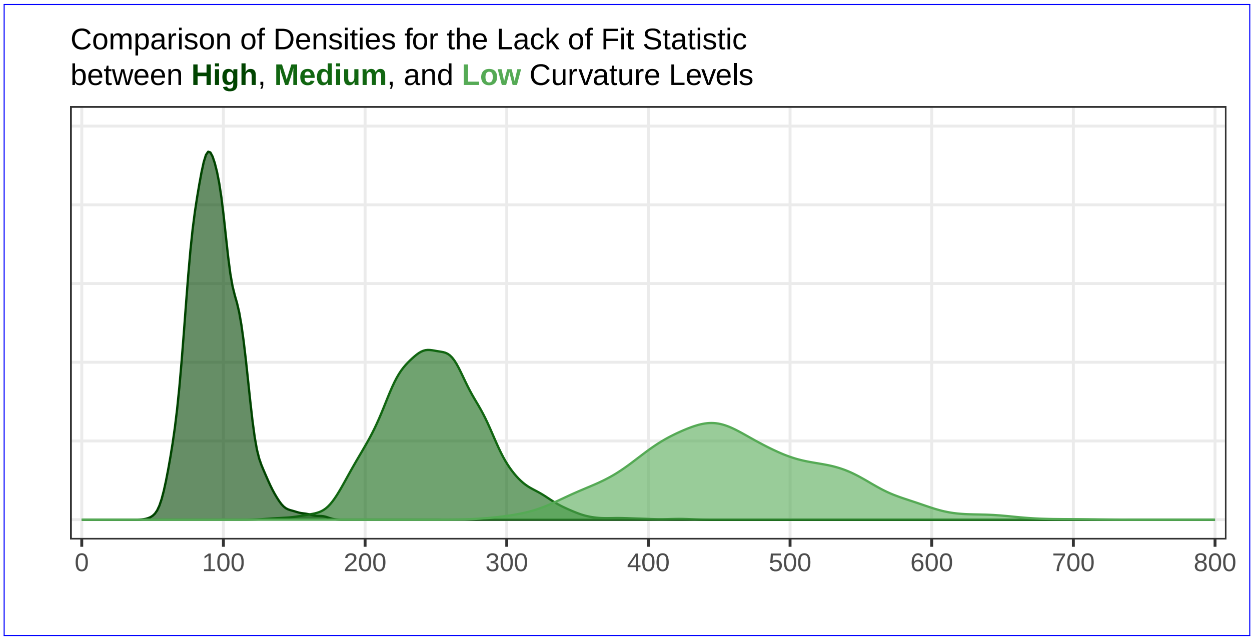
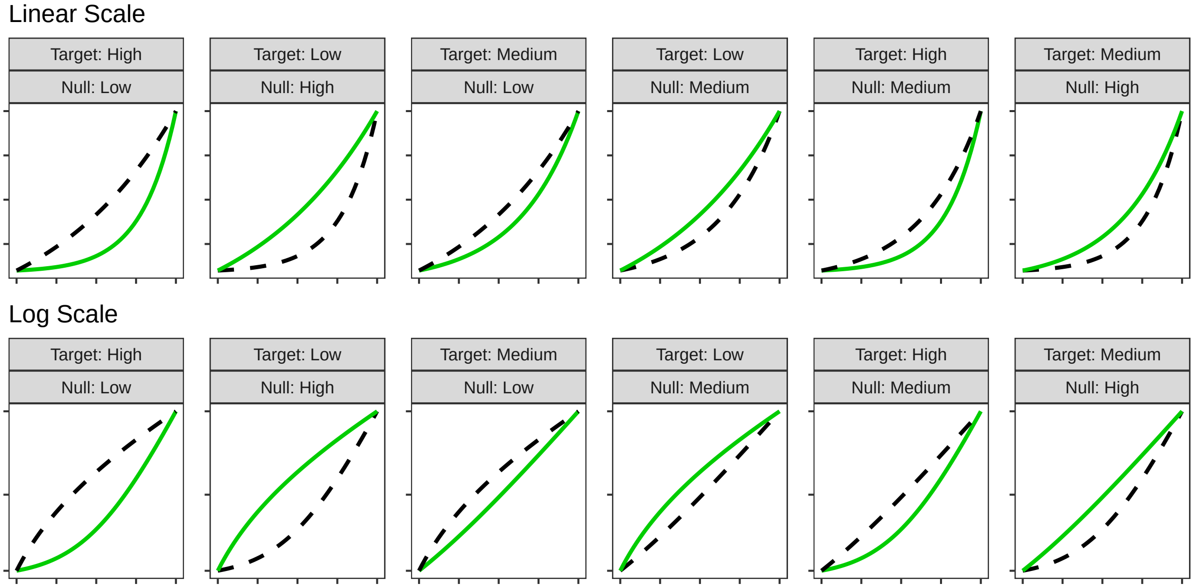


Figure 3: Density plot of the lack of fit statistic showing separation of difficulty levels: obvious curvature, noticable curvature, and almost linear.

target plot and multiple data sets corresponding to curvature level B for the null plots. The `nullabor` package in R [4] randomly assigned the target plot to one of the panels surrounded by panels containing null plots. ~~For example, the statistical lineup randomly embeds a target plot with simulated data following an increasing exponential curve with high curvature within null plots with simulated data following an increasing exponential trend with low curvature.~~ The target and null panels ~~spanned~~ span a similar domain and range due to the implemented constraints when simulating the data; the rationale for this decision is based on preattentive feature perception [27] and is discussed in detail in the appendix. There were a total of six lineup curvature combinations; Fig. 4 illustrates the six lineup curvature combinations (top: linear scale; bottom: log scale) where the ~~green~~ solid line indicates the curvature level designated to the target plot while the ~~black~~ dashed line indicates the curvature level assigned to the null plots. Two sets of each lineup curvature combination were simulated (a total of twelve test data sets) and plotted on



~~Thumbnail plots illustrating~~

Figure 4: Thumbnail plots illustrating the six curvature combinations displayed on both scales (linear and log). The solid line indicates the curvature level to be identified as the target plot from amongst a set of null plots with the curvature level indicated by the dashed line.

both the linear scale and the log scale (24 test lineup plots). In addition, three curvature combinations generated homogeneous "Rorschach" lineups, where all panels were from the same distribution. Each participant evaluated one "Rorschach" lineup~~, but for simplicity, we did not describe these evaluations~~ , ~~and we left their analysis to a later date.~~ . Results from the "Rorschach" evaluations indicate null panel selections were distributed relatively evenly with multiple candidates for the most interesting panel. We display and further discuss the "Rorschach" evaluation results in the appendix.

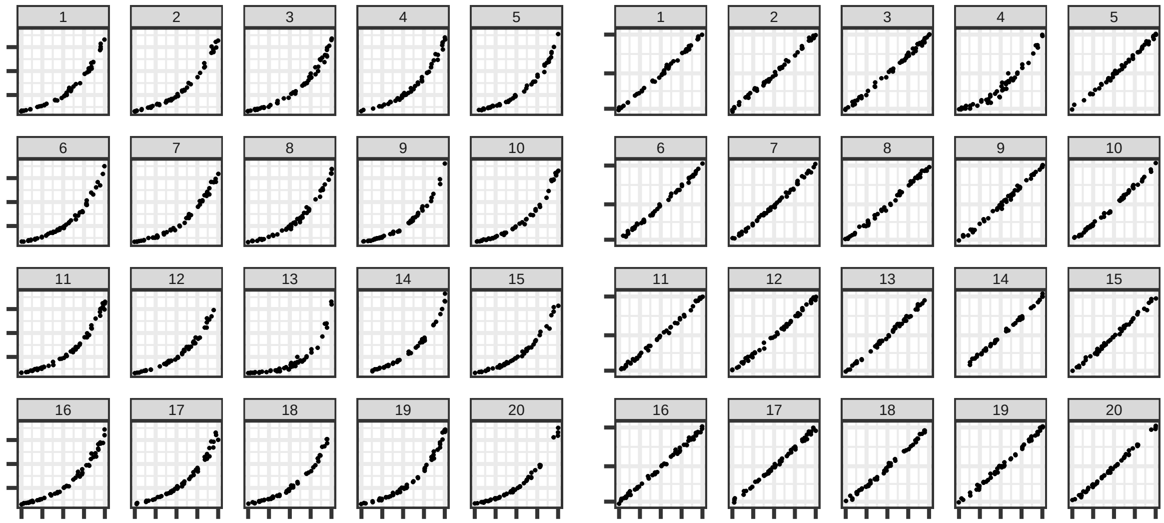


Figure 5: The lineup plot on the ~~six curvature combinations displayed~~ left displays increasing exponential data on ~~both scales (~~a linear ~~and log). The green solid line indicates the curvature level to be identified~~ scale with panel (13 as the target. The lineup plot ~~from amongst a set of null plots with~~ on the ~~curvature level indicated by~~ right displays increasing exponential data on the ~~black dashed line~~log scale with panel 4 as the target.

Fig. 5 presents examples of statistical lineups with the target data simulated with exponential parameters corresponding high curvature and the surrounding null panels simulated with parameters for low curvature. The statistical lineup on the left presents increasing exponential data with displayed on a linear scale with panel 13 as the target panel. The lineup on the right shows increasing exponential data plotted on a log scale with panel 4 as the target panel.

2.4 Study Design and Implementation

We used Prolific, a survey site that connects researchers to study participants, to recruit participants above the age of majority (18+ in most regions; 19+ in certain U.S. states) in their region; we did not request a representative sample, previous literature suggests there are minor effects of demographics on the outcome of graphical experiments involving lineups [21, 16]. Following the completion of the current statistical lineup study, participants sequentially completed two additional graphical experiments related to the perception of logarithmic scales (to be discussed at a later date), and we compensated them for their participation in the series of all three studies.

We showed each participant 13 lineup plots (12 test and one Rorschach). At the start of the study, we randomly assigned participants to one of the two replicate data sets for each of the six unique lineup curvature combinations. Participants evaluated the lineup plot corresponding to the linear and log scales for each assigned test data set. For the additional Rorschach lineup plot, we randomly assigned participants to one data set shown on either the linear or the log scale. We randomized the order in which participants saw the assigned 13 lineup plots.

Fig. 6 presents a screenshot of the study homepage, which served as an introduction to participants and guided them through the series of three graphical experiments. The first experiment, discussed in this paper, utilized statistical lineups to investigate the effect of logarithmic scales on perceptual sensitivity. The second experiment, incorporated an interactive 'You Draw It' feature introduced by **(author?)** [1] and employed in the study by **(author?)** [19], to examine the effect of logarithmic scales on prediction. The third experiment focused on numerical estimation and cognitive understanding of logarithmic scales.

Participants completed the series of graphical tests using an R Shiny application [6] accessible at `https://emily-robinson.shinyapps.io/perception-of-statistical-graphics-log/`. The code used to create the study application is available on GitHub at `https://`

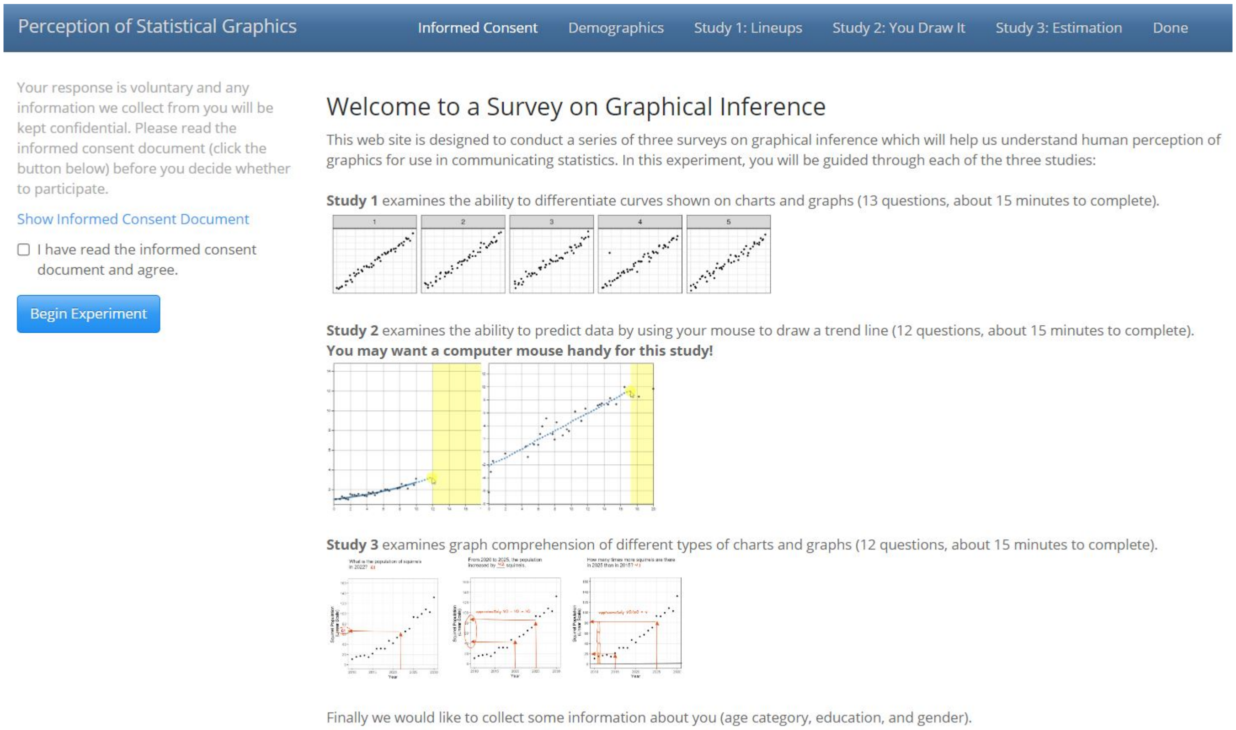


Figure 6: Screenshot of the study applet homepage guiding participants through the series of three graphical experiments related to the perception of logarithmic scales.

github.com/earobinson95/perception-of-statistical-graphics-log. Before completing the lineup study, the 'You Draw It' study, and the estimation study, participants were prompted to provide their demographic information. The series of studies received an IRB exemption (IRB #20200720178EX) from the University of Nebraska - Lincoln.

The statistical linenup study guided participants through the series of 13 lineup plots and asked them to identify the plot which appeared to be most different from the others (Fig. 7). In each lineup evaluation, participants justified their choice ~~and~~ in a select all that apply format (Clustering, Different range, Different shape, Outlier(s), and Other) as shown on the left side of the applet screenshot. Additionally, participants provided their level of confidence in their choice by answering "How certain are you?" on a five-point likert scale from very certain to very uncertain. This graphical task aimed to test an individual's

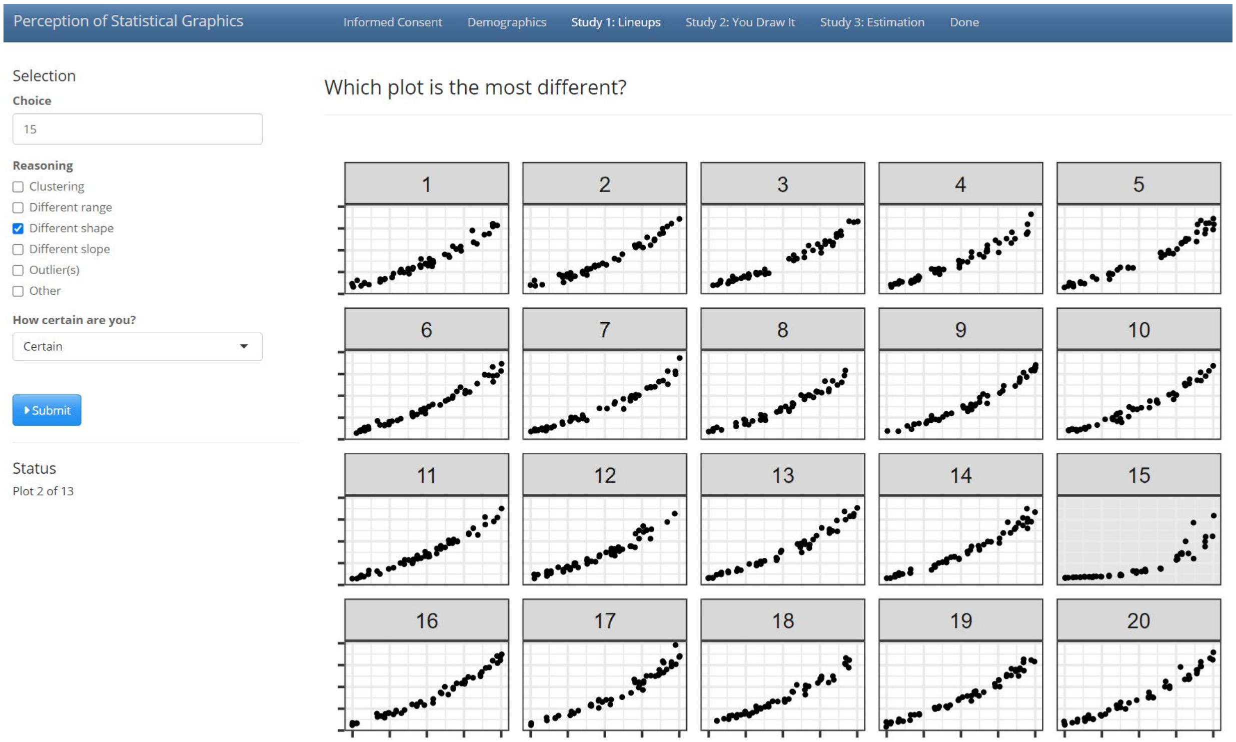


Figure 7: Screenshot of an example trial participants see when completing the lineup study. The applet guided participants through 13 lineup plots and asked them to identify the plot which appeared to be the most different from the others.

ability to perceptually differentiate exponentially increasing trends with differing levels of curvature on both the linear and log scales.

2.5 Statistical Analysis

Each lineup plot evaluated was assigned a binary value based on the participant response (correct target plot identification = 1, not correct target plot identification = 0). We defined Y_{ijkl} as the event that participant $l = 1, ..., N_{\text{participant}}$ correctly identified the target plot for data set $k = 1, 2$ with curvature combination $j = 1, 2, 3, 4, 5, 6$ plotted on scale $i = 1, 2$. The binary response was analyzed using a generalized linear mixed model (GLMM) following a binomial distribution with a logit link function with a row-column blocking design accounting for the variation due to participant and data set, respectively, as

$$\text{logit } P(Y_{ijkijkl}) = \eta + \delta_i + \gamma_j + (\delta\gamma)_{ij} + s_l + d_k \tag{2}$$

where

- η is the baseline average probability of selecting the target plot
- δ_i is the effect of scale $i = 1, 2$
- γ_j is the effect of curvature combination $j = 1, 2, 3, 4, 5, 6$
- $(\delta\gamma)_{ij}$ is the two-way interaction between the i^{th} scale and j^{th} curvature combination
- $s_l \sim N(0, \sigma^2_{\text{participant}})$ is the random effect for participant characteristics
- $d_k \sim N(0, \sigma^2_{\text{data}})$ is the random effect for data specific characteristics.

We assumed the random effects for data set and participant are independent. Target plot identification was analyzed using a GLMM implemented in `glmer` from the `lme4` R (version 4.2.2) package [2]. We used the `emmeans` R package [12] to calculate the estimated target detection probabilities and obtain odds ratio comparisons between the log and linear scale.

3 Results

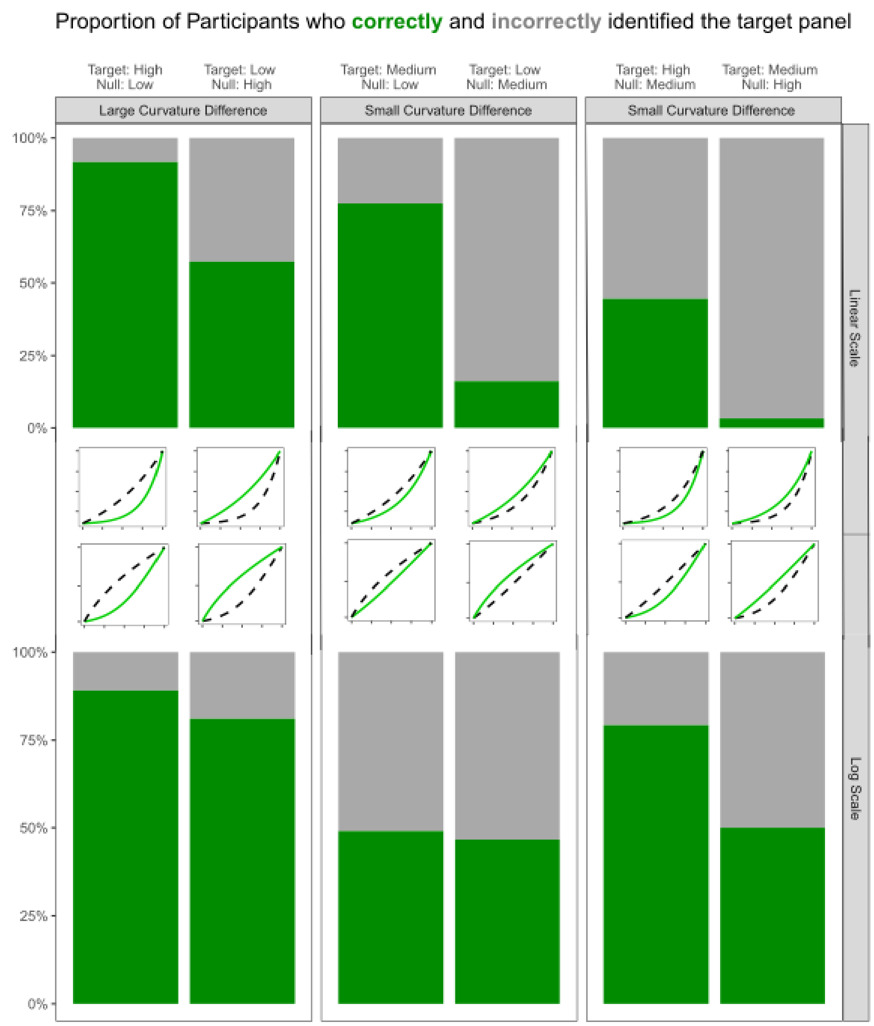
We recruited participants and conducted the study via Prolific, a crowd-sourcing website, in March 2022. The study included a diverse group of participants, with an inner-quartile age range between 23 and 31 and a median age of 26 years old. Among the participants, 59% self-identified as male, 40% as female, and 1% as variant/nonconforming. Additionally, individuals from more than 21 countries participated in the study and 97% of the participants indicated fluency in English. Moreover, 81% of the participants reported having completed some undergraduate courses or higher. During data collection, 325 individuals completed 4,492 individual test lineup evaluations. We included only participants in the final analysis who completed the entire study, which included 311 participants and 3,958 lineup evaluations. Due to server capacity, some participants were required to restart the study, thus resulting in the possibility of more than twelve lineup evaluations per participant. As a whole, participants evaluated each uniquely generated lineup plot between 141 and 203 times (Mean: 164.92, SD: 14.9). Participants correctly identified the target panel in 47% of the 1,981 lineup evaluations made on the linear scale and 65.3% of the 1,977 lineup evaluations made on the log scale.

~~The results from the GLMM indicated a strong interaction between the curvature combination and scale ($\chi^2_5 = 294.443$; p < 0.0001), and the estimated variance due to participant and data set were $\hat{\sigma}^2_{\text{participant}} = 1.19$ (s. e. = 1.09) and $\hat{\sigma}^2_{\text{data}} = 0.433$ (s.e. = 0.66), respectively. Therefore, we concluded that there was low variability in the accuracy of target panel detection between participants and across replications of uniquely simulated data sets. To determine the effect of scale , we compared the estimated accuracy between the log and linear scale within each curvature combination due to the interaction as determined by the GLMM.~~

Fig. 8 shows the observed participant accuracy for each scale and curvature combination scenario. We can see from the observed results that participant accuracy for the linear scale ranged from 3.3% to 91.6% while participant accuracy when identified on the log scale ranges from 46.6% to 89%. On both the log and linear scales, the highest accuracy occurred in lineup plots where the target model and null model had a considerable difference in curvature, and the target plot had more curvature than the null plots (high curvature target plot embedded in low curvature null plots). There was a decrease in accuracy on the linear scale when comparing a target plot with less curvature to null plots with more curvature (medium curvature target plot embedded in high curvature null plots; low curvature target plot embedded in medium curvature null plots; low curvature target plot embedded in high curvature null plots). **(author?)** [3] found that the accuracy of identifying the correct curve type was higher when presented with nonlinear trends, indicating that it is hard to say something is linear (i.e., something has less curvature), but easy to say that it is not linear; our results concur with this observation. Additionally, accuracy increased when data was displayed on the log scale compared to the linear scale in all curvature scenarios with an exception of a medium target curve embedded in low null curves. The thumbnail images below this particular scenario provides support for the results found in **(author?)** [3] and visually demonstrate the opposing perceptual behaviors of the curves for this scenario when displayed on the two different scales. In addition to participant accuracy, we observed that, in general, participants who correctly identified the target plot were more confident across all conditions. We discuss further details regarding selection reasoning and confidence level in the appendix.

The results from the GLMM indicated a strong interaction between the curvature combination and scale ($\chi^2_5 = 294.443$; $p < 0.0001$), and the estimated variance due to participant and data set were $\hat{\sigma}^2_{\text{participant}} = 1.19$ (s.e. $= 1.09$) and $\hat{\sigma}^2_{\text{data}} = 0.433$ (s.e. $=$ 0.66), respectively. Therefore, we concluded that there was low variability in the accuracy of target panel detection between participants and across replications of uniquely simulated data sets. To determine the effect of scale, we compared the estimated accuracy between the log and linear scale within each curvature combination due to the interaction as determined by the GLMM.

Fig. 9 displays the estimated (log) odds ratio of successfully identifying the target panel



2

Figure 8: The observed accuracy for identifying the target panel for each curvature combination scenario with the accuracy for the linear scale shown on the top and the accuracy for the log scale shown on the bottom. The thumbnail figures below each plot display the curvature combination as shown in Fig. 4 on both scales.

on the log scale compared to the linear scale. The thumbnail figures to the right of the plot illustrate the curvature combination on the linear (left thumbnail) and log base ten (right thumbnail) scales associated with the y-axis label. The choice of scale had no impact if the curvature differences were substantial and the target plot had more curvature than the null plots (high curvature target plot embedded in low curvature null plots). However, presenting data on the log scale makes us more sensitive to slight changes in curvature (low or high curvature target plot embedded in medium curvature null plots; medium curvature target plot embedded in high curvature null plots) and apparent differences in curvature when the target plot had less curvature than the null plots (low curvature target plot embedded in high curvature null plots). An exception occurred when identifying a plot with curvature embedded in null plots close to a linear trend (medium curvature target panel embedded in low curvature null panels). The results indicate that participants were more accurate at detecting the target panel on the linear scale than on the log scale. When examining this curvature combination, the same perceptual effect occurred as we previously saw, but in a different context of scales. On the linear scale, participants were perceptually identifying a convexly curved trend from close to a linear trend, whereas after the logarithmic transformation, participants were perceptually identifying a trend close to linear from a concavely curved trend (Fig. 4). This result again supports the claim that it is easy to identify a curve in a group of lines but harder to identify a line in a group of curves [3].

4 Discussion and Conclusion

This work aims to provide foundational research to support the principles that guide design decisions in scientific visualizations of exponential data. In this study, we explored the

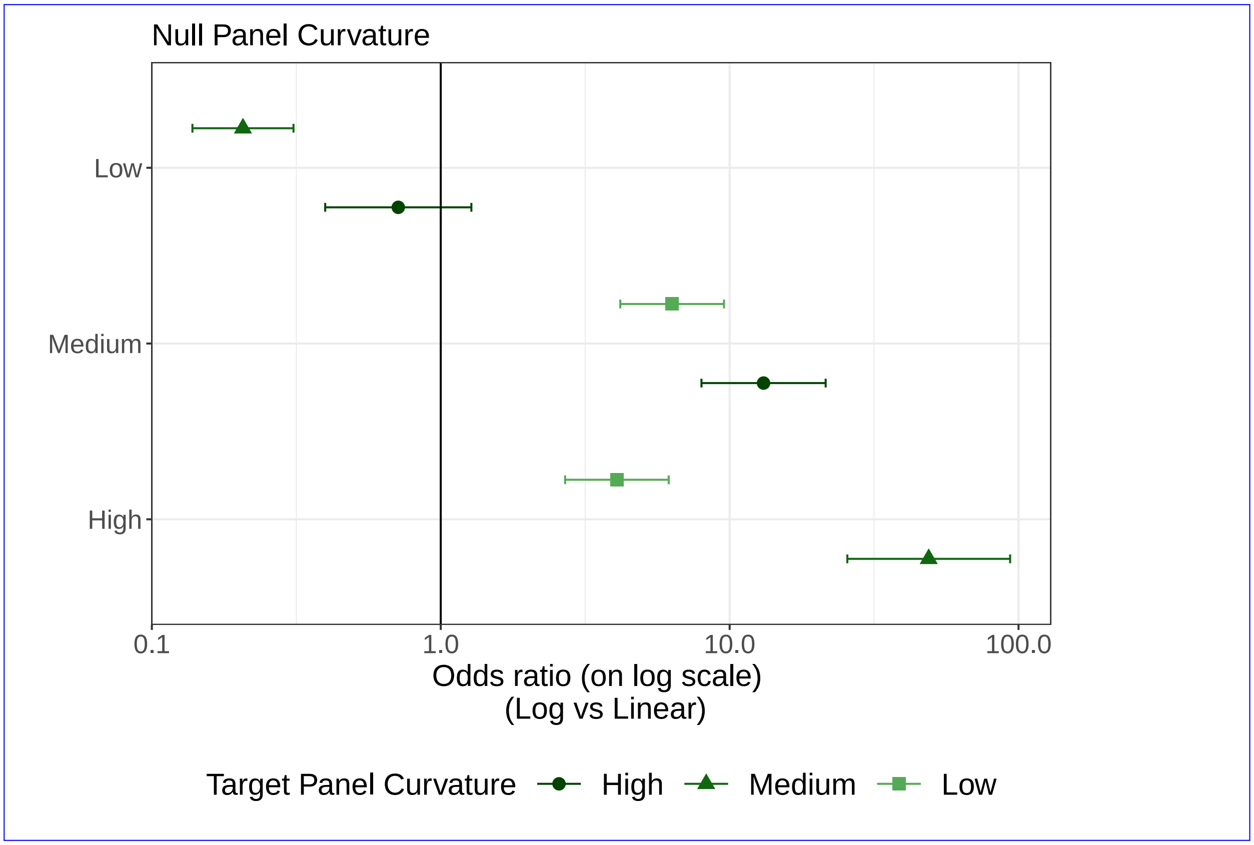


Figure 9: Estimated (log) odds ratio of successfully identifying the target panel on the log scale compared to the linear scale. The y-axis indicates the ~~the~~ model parameters used to simulate the null plots with the target plot model parameter selection designated by shape and shade~~of green~~. The thumbnail figures on the right display the curvature combination as shown in Fig. 4 on both scales (linear - left, log - right).

use of linear and log scales to determine whether the choice of scale impacts our ability to notice differences in exponentially increasing trends. The results indicated that when there was a considerable difference in curvature between the target plot and null plots and the target plot had more curvature than the null plots, the choice of scale had no impact, and participants accurately differentiated between the two curves on both the linear and log scale. However, displaying exponentially increasing data on a log scale improved the accuracy of differentiating between models with slight curvature differences or apparent curvature differences when the target plot had less curvature than the null plots. An exception occurred when identifying a plot with curvature embedded in surrounding plots perceived close to a linear trend, indicating that it is easy to identify a curve in a group of lines but much harder to identify a line in a group of curves. Using visual inference to identify these guidelines suggests that there are perceptual advantages to log scales when differences are subtle. It is worth noting that our study focused specifically on data simulated with a three-parameter exponential model and such conclusions may not be broadly applicable to functions resulting in curvature. This scenario, while fairly specific, lays the perceptual groundwork for more investigation into the use of log scales with exponential data. Now that we know how curvature can be distinguished, it's easier to conduct follow up studies that cover more scenarios and use different graphical testing methods.

We conducted this study as the first in a series of three graphical tests to understand the perceptual and cognitive implications of using log scales to display exponentially increasing data. In our next two papers in this series, we will investigate whether using log scales presents cognitive disadvantages, such as making it harder to utilize graphical information. These studies serve as an example of multi-modal graphical testing, examining different levels of engagement and interaction with graphics in order to produce nuanced, specific guidelines for graphical design. By testing graphics in situations similar to how they are used in practice, we can gain additional insight into graphical perception and improve visual communication of scientific results.

Supplementary Material

- **Participant Data:** De-identified participant data collected in the study and used for analyses (lineup-model-data.csv).
- **Data Analysis Code:** The code used to replicate the analysis in this paper (lineups-analysis.qmd).
- **Study Applet Code:** The code used to create the study applet via RShiny can be found on GitHub at `https://github.com/earobinson95/perception-of-statistical-graphics-log`
- **README:** File containing detailed descriptions of the supplementary material. (README.html).

A Exponential growth rates and curvature

Let us consider a simple equation which we could use to simulate data which grows exponentially in x, with optional error term ϵ:

$$y = \exp\{\beta_1 x + \epsilon\} \quad \text{where} \quad \epsilon \sim N(0, \sigma). \tag{A1}$$

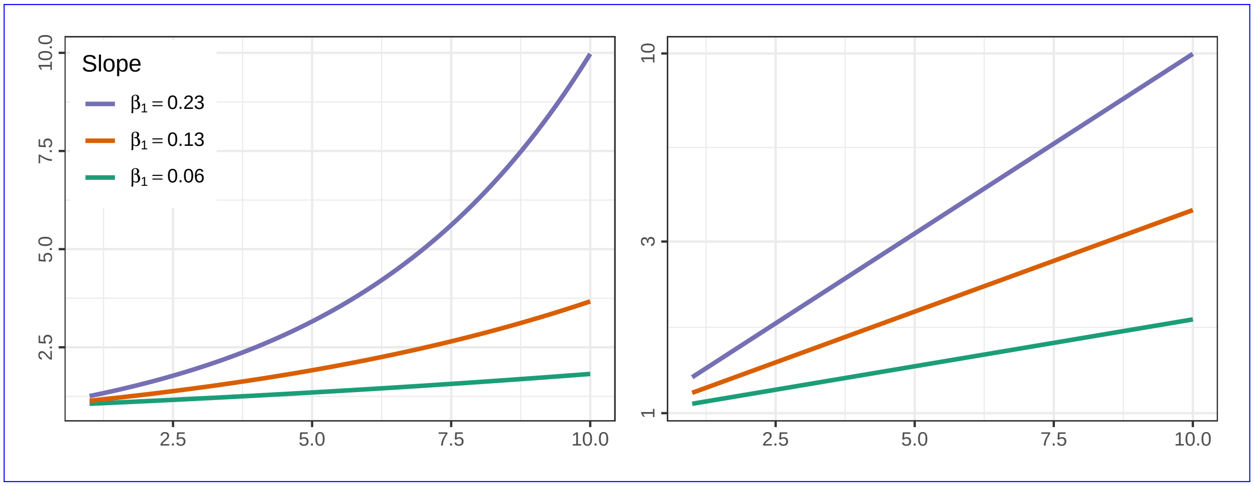


Figure A1: Linear and log scale simple exponential equations as specified in Eq. (A1), with error term excluded. The y-axis endpoints are a primary visual signal when differentiating between the lines in both plots.

It is important in lineup studies to control for extraneous visual signals, and when other visual signals creep in, results of the study can be difficult to interpret [22].

In this particular case, the extraneous visual signals are the endpoints of the lines (the endpoint at $x = 10$ on a linear scale, and both endpoints on the log scale), as shown in Fig. A1. Preattentive features guide attention [27]; one of the first things we do when scanning a graph is to notice the extent of the data within the axes. We must control the endpoints to ensure that participants are using active attention to assess the lineup and draw conclusions, so that the whole visual signal is processed as part of the lineup evaluation; if participants are making decisions off of the endpoints, then we cannot

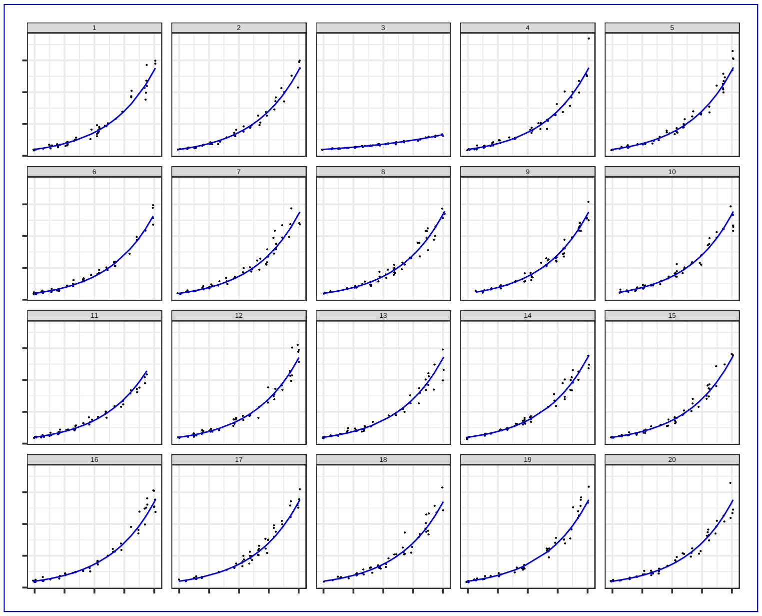


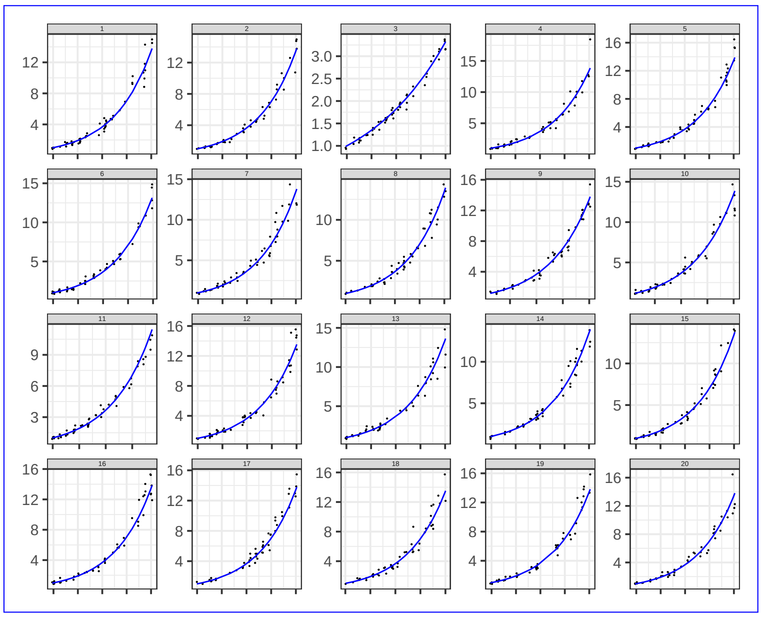
Figure A2: When axis limits are fixed to the most expansive extent in the generated data, the primary signal becomes the extent of the domain which has data, rather than the growth rate itself.

interpret the results to say that they are assessing the rate of exponential growth rather than the end result.

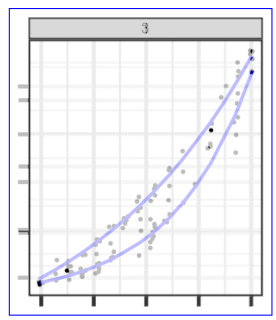
There are several different options for controlling the visual signal in a lineup:

0. Do nothing and set each subplot's limits to the overall maximum limits.
1. allow each subplot to have different axis limits
2. truncate the displayed range of each subplot to the minimum range generated in the lineup
3. add extra parameters to the exponential equation to adjust the range of the data so that it fits within a pre-specified domain, as in Algorithm 2.

Typically, lineups do not show axis labels or scaling, because the goal is to assess the signal from the data without additional context. In addition, the traditional 20-panel lineup would be quite visually confusing in this circumstance:



(a) Lineup



(b) Overlay of null and signal panels.

Figure A3: When y-axis limits vary by panel, the y-axis labels are the primary visual signal; the secondary signal (when the panels are overlaid) is the curvature of the line.

Fig. A3 demonstrates that when we ignore the y-axis labels and overlay the signal panel with one of the null panels, the remaining difference is the curvature of the line. If we want to keep the standard convention of not including y-axis labels in lineups for simplicity and to reduce plot clutter, this alternative signal seems promising (and as we will show, is approximately equivalent to option 3).

Let us next consider option 2: Crop each panel to the minimum limits of all generated data. The result is shown in Fig. A4. This operation shifts which axis becomes the visually important factor, but doesn't change the problem: previously the issue was the y-axis extent, now it is the x-axis extent. Both of these parameters are implicitly affected by how we generate exponential data. In order to truly assess whether people can discriminate between comparable exponential growth rates graphically, it is more useful to approach the problem from a curvature perspective rather than to artificially limit the data shown.

This issue is a fundamental problem when testing graphics: the test must meet the "goldilocks" standard - not too hard, not too easy, but just right. Both option 0 and option 2 fail this standard.

Visually, both the x and y axes matter equally, even if it might on paper make sense to truncate x in order to control the y axis range.

An additional benefit of controlling endpoints is that it also provides some realism: our goal was to assess the ability to examine exponential growth **rates**, motivated by the COVID-19 pandemic. As each geographic reporting region started with different numbers of cases and had different control policies, the α and θ parameters are reasonable to represent some of these differences while still examining the underlying exponential behavior.

To control these endpoints on both scales then we have to move to an exponential model

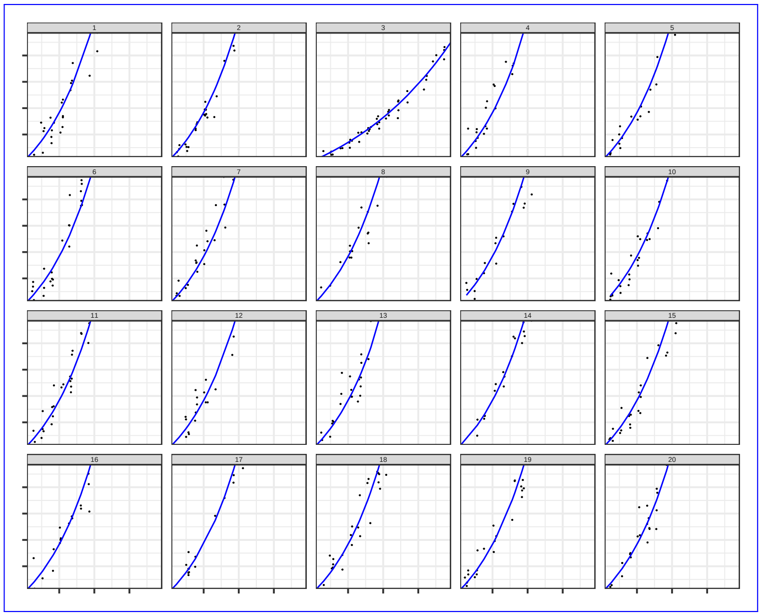


Figure A4: When axis limits are fixed to the most restrictive in all generated data, the primary signal becomes the extent of the x axis which is covered with data, rather than the growth rate itself.

that is a bit more complicated:

$$y = \alpha \cdot \exp\{\beta_1 x + \epsilon\} + \theta \quad \text{where} \quad \epsilon \sim N(0, \sigma). \tag{A2}$$

α and θ were used to make endpoints consistent among the growth rates. As a result, the lineups examine the degree of inflection of the trend rather than the explicit exponential growth rate. An explanation for how the ultimate values for α and θ were determined is provided in Algorithm 2. Fig. A5 uses the parameters in Table 1 and the data generating method described in Algorithm 2. The result is a series of panels which have obvious variation due to the points (a desirable feature) but where the underlying relationship shown in blue has consistent endpoints. As a result, the question becomes whether participants can identify that plot 3 has a different growth rate (as measured by the curvature of the line).

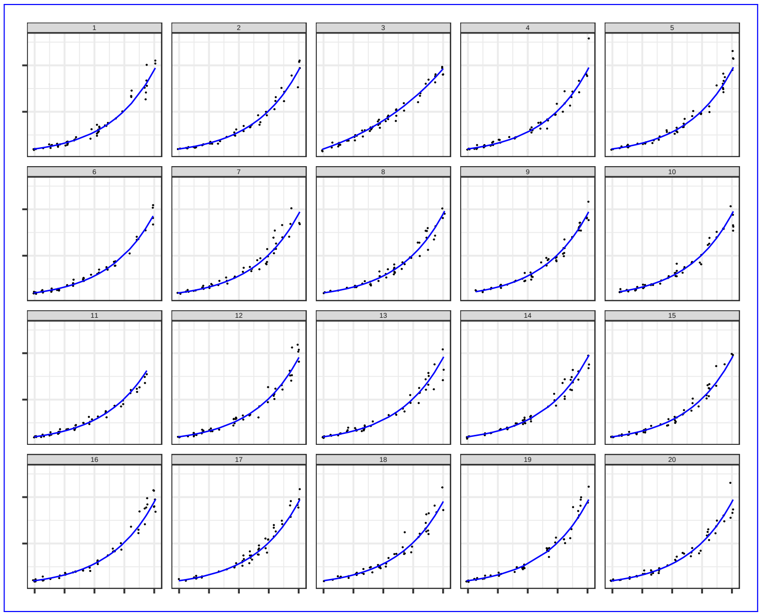


Figure A5: When the data are scaled such that endpoints are similar across all conditions, the primary feature becomes the curvature of the line, while individual plots show random variation due to the scatter around the line.

B Rorschach Lineup Evaluation

In addition to the experimental lineups, we generated three sets of homogeneous "Rorschach" lineups, each featuring panels simulated from the same distribution. Importantly, participants were unaware that there was no designated target panel for identification. Fig. B6 illustrates the selection proportions for each panel on the "Rorschach" lineups corresponding to different curvature combinations and replicated data sets, when presented on both log and linear scales. Notably, the panel number is irrelevant, and the darker shade (bottom) denotes panels chosen more frequently than others.

This approach allows us to explore whether any null panels were significantly more unique, providing insights into the null plot sampling method. The plots exhibit a relatively even distribution, offering multiple candidates for the most interesting panel. It is worth noting that in the low curvature "Rorschach" lineups simulated in data set replication one,

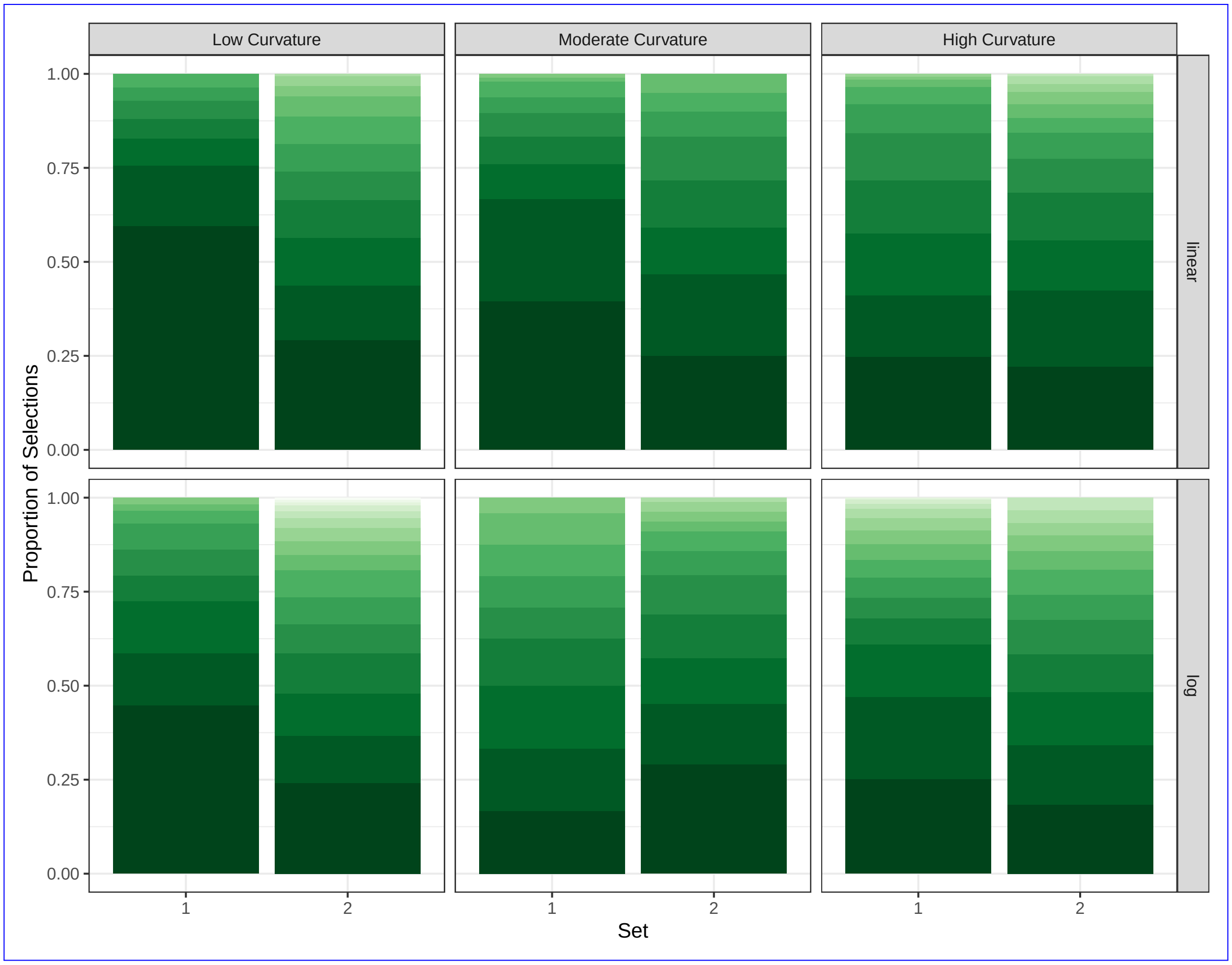


Figure B6: Observed proportions of panel selections for the 'Rorschach' lineups. The darker shade (bottom) denotes panels chosen more frequently than others.

a single panel stands out, with over 50% of participants selecting the same null panel on the linear scale and just under 50% on the log scale. Through a visual examination of "Rorschach" evaluations, we determined that our null sampling method is appropriate.

C Participant confidence and accuracy

In addition to investigating how scale and curvature rate influence participant accuracy, we asked participants to indicate their level of confidence on a five-point Likert scale

Figure C7: Observed proportion of evaluations for each confidence rating (five-point Likert scale) by accuracy. The width of the bars indicates the number of evaluations for correct and incorrect identifications within each curvature combination.

by answering, "How certain are you?" Fig. C7 shows the observed confidence rating proportions for correct and incorrect identifications across the different scale and curvature combination scenarios. Overall, participants who correctly identified the target plot were more confident in their plot choice across all conditions with the highest confidence levels and accuracy in scenarios with large differences in curvature between the target and null panels.

References

[1] Gregor Aisch, Amanda Cox, and Kevin Quealy. You draw it: How family income predicts children's college chances. *The New York Times*, 28, 2015.

[2] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[3] Lisa Best, Laurence Smith, and D Stubbs. Perception of linear and nonlinear trends: Using slope and curvature information to make trend discriminations. *Perceptual and Motor Skills*, 104(3):707–721, 2007.

[4] Andreas Buja, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun-Kyung Lee, Deborah Swayne, and Hadley Wickham. Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4361–4383, 2009.

[5] John Burn-Murdoch, Caroline Nevitt, Cale Tilford, Andrew Rininsland, Joanna Kao, Oliver Elliott, Emma Lewis, Brooke Fox, and Martin Stabe. Coronavirus tracked: Has the epidemic peaked near you?, 2020.

[6] Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. *shiny: Web Application Framework for R*, 2022. R package version 1.7.4.

[7] William Cleveland and Robert McGill. Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716):828–833, 1985.

[8] Wade Fagen-Ulmschneider. 91-divoc, 2020.

[9] Andrew Heckler, Brendon Mikula, and Rebecca Rosenblatt. Student accuracy in reading logarithmic plots: The problem and how to fix it. In *2013 IEEE Frontiers in Education Conference (FIE)*, pages 1066–1071. IEEE, 2013.

[10] Heike Hofmann, Lendie Follett, Mahbubul Majumder, and Dianne Cook. Graphical tests for power comparison of competing designs. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2441–2448, 2012.

[11] Gregory Jones. Polynomial perception of exponential growth. *Perception & Psychophysics*, 1977.

[12] Russell Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2021. R package version 1.6.0.

[13] Adam Loy, Lendie Follett, and Heike Hofmann. Variations of q–q plots: The power of our eyes! *The American Statistician*, 70(2):202–214, 2016.

[14] Adam Loy, Heike Hofmann, and Dianne Cook. Model choice and diagnostics for linear mixed-effects models using statistics on street corners. *Journal of Computational and Graphical Statistics*, 26(3):478–492, 2017.

[15] Andrew MacKinnon and Alexander Wearing. Feedback and the forecasting of exponential change. *Acta Psychologica*, 76(2):177–191, 1991.

[16] Mahbubul Majumder, Heike Hofmann, and Dianne Cook. Validation of visual statistical inference, applied to linear models. *Journal of the American Statistical Association*, 108(503):942–956, 2013.

[17] Duncan Menge, Anna MacPherson, Thomas Bytnerowicz, Andrew Quebbeman, Naomi Schwartz, Benton Taylor, and Amelia Wolf. Logarithmic scales in ecolog-

ical data presentation may cause misinterpretation. *Nature Ecology & Evolution*, 2(9):1393–1402, 2018.

[18] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.

[19] Emily A Robinson, Reka Howard, and Susan VanderPlas. Eye fitting straight lines in the modern era. *Journal of Computational and Graphical Statistics*, pages 1–8, 2022.

[20] Susan Vanderplas, Dianne Cook, and Heike Hofmann. Testing statistical charts: What makes a good graph? *Annual Review of Statistics and Its Application*, 7:61–88, 2020.

[21] Susan VanderPlas and Heike Hofmann. Spatial reasoning and data displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):459–468, 2015.

[22] Susan VanderPlas and Heike Hofmann. Clusters beat trend!? testing feature hierarchy in statistical graphics. *Journal of Computational and Graphical Statistics*, 26(2):231–242, 2017.

[23] Jens Von Bergmann. xkcd exponential: Public health vs scientists, Mar 2021.

[24] William Waddell. Comparisons of thresholds for carcinogenicity on linear and logarithmic dosage scales. *Human & Experimental Toxicology*, 24(6):325–332, 2005.

[25] William Wagenaar and Sabato Sagaria. Misperception of exponential growth. *Perception & Psychophysics*, 18(6):416–422, 1975.

[26] Hadley Wickham, Dianne Cook, Heike Hofmann, and Andreas Buja. Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):973–979, 2010.

[27] Jeremy M Wolfe and Igor S Utochkin. What is a preattentive feature? 29:19–26.