

HUMAN PERCEPTION OF EXPONENTIALLY INCREASING DATA
DISPLAYED ON A LOG SCALE

by

Emily Anna Robinson

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfilment of Requirements
For the Degree of Doctor of Philosophy

Major: Statistics

Under the Supervision of Professors Susan VanderPlas and Reka Howard

Lincoln, Nebraska

August, 2022

HUMAN PERCEPTION OF EXPONENTIALLY INCREASING DATA DISPLAYED ON A LOG SCALE

Emily Anna Robinson, Ph.D.

University of Nebraska, 2022

Advisers: Susan VanderPlas and Reka Howard

Log scales are often used to display data over several orders of magnitude within one graph. During the COVID-19 pandemic, we have seen both the benefits and the pitfalls of using log scales to display case counts. Three graphical experimental tasks were conducted to evaluate the impact our choice of scale has on human perception of exponentially increasing trends. The first experiment evaluates whether our ability to perceptually notice differences in exponentially increasing trends is impacted by the choice of scale. We conducted a visual inference experiment in which participants were shown a series of lineup plots (consisting of 19 null panels and 1 target panel) and asked to identify the panel that was most different from the others. Our results indicated that when there was a large difference in curvature between the target plot and 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. An exception occurred when identifying a plot with curvature embedded in surrounding plots closely relating 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. The use of visual inference to identify these guidelines suggests that there are *perceptual* advantages to log scales when differences are subtle. Our other experimental tasks

focus on determining whether there are cognitive disadvantages to log scales: do log scales make it harder to make use of graphical information? We conducted a graphical task similar to the New York Times “You Draw It” page to test an individual’s ability to use and make predictions for exponentially increasing data. We asked participants to draw a line using their computer mouse through the increasing exponential trend shown on both scales. In addition to differentiation and prediction of exponentially increasing data, we conduct an experimental task to test an individuals’ ability to translate a graph of exponentially increasing data into real value quantities and extend their estimations by making comparisons. The results of our experimental tasks allow us to provide guidelines for readers to actively choose which of many possible graphics to draw, according to some set of design choices, to ensure that our charts are effective. *(399 words; 350 word limit)*

Acknowledgments

Thank you to all my people!

Table of Contents

List of Figures	vi
List of Tables	vii
1 Literature Review	1
1.1 Introduction to Graphics	1
1.2 Perception and Psychophysics	4
1.3 Graphical Experiments	6
1.4 Logarithmic Scales and Mapping	9
1.5 Underestimation of Exponential Growth	12
1.6 Risk Assessment	13
1.7 Research Objectives	14
2 Perception through lineups	15
2.1 Introduction	15
2.2 Data Generation	17
2.3 Parameter Selection	19
2.4 Lineup Setup	20
2.5 Study Design	21
2.6 Results	22
2.7 Discussion and Conclusion	24

3	Prediction with you draw it	26
3.1	Introduction	26
4	Numerical Translation and Estimation	27
4.1	Introduction	27
5	Conclusion	28
A	The First Appendix	29
B	The Second Appendix	30
	References	31

List of Figures

2.1	Lineup example	17
2.2	Lack of fit statistic density curves	20
2.3	Lineups log(odds) results	24

List of Tables

2.1	Lineup data simulation final parameters	20
-----	---	----

Chapter 1

Literature Review

Editing text colors: [Emily's editing color](#). Emily may also use mostly black text as well. [Susan's editing color](#). [Reka's editing color](#).

1.1 Introduction to Graphics

Advanced technology and computing power has promoted data visualization as a central tool in modern data science (Unwin, 2020). Data visualization is defined as the art of drawing graphical charts in order to display data. Vanderplas, Cook, & Hofmann (2020) describe the process of creating a graphic from a dataset through the use of variable mapping, data transformations, coordinate systems, and aesthetic features. Graphics are useful for data cleaning, exploring data structure, and have been an essential component in communicating information for the last 200 years (Lewandowsky & Spence, 1989; Unwin, 2020; Vanderplas, Cook, & Hofmann, 2020). During the 20th century, companies began utilizing graphics to understand their mechanics and support business decisions; news sources began displaying graphics of weather forecasts as a means to communicate critical information and aid in decision-making (Vanderplas, Cook, & Hofmann, 2020). Today, we encounter data visualizations everywhere, researchers include graphics to communicate their results

in scientific publications and mass media relies on graphics to convey news stories to the public through newspapers, TV, and the Web (Unwin, 2020).

In general, the public holds the self perception that numbers are difficult to understand and that they did not perform well in mathematics during school (Unwin, 2020). In contrary, there tends to be a positive self perception when it comes to graphics as they are viewed more as illustrations and not as critical parts of an argument. While tables can be burdensome to readers, graphics can improve the interpretation and representation of the same data (Uri & Haemer, 1948). Shah, Mayer, & Hegarty (1999) presents an opposing argument claiming there are still difficulties interpreting and explaining quantitative information depicted in graphs. One explanation for the opposing views is that the general public is untrained in the ability required for the evaluation of graphic material. The complexity of graphic devices is directly related to the degree and formality of training necessary for understanding (Haemer & Kelley, 1949).

Although statistical graphics have become widely used and valued in science, business, and in many other aspects of life, as creators of graphics, we are too accepting of them as default without asking critical questions about the graphics we create or view (Unwin, 2020). Vanderplas, Cook, & Hofmann (2020) poses the general question we must ask ourselves, “how effective is this graph at communicating useful information?” Higher quality of technology has influenced the creation, replication, and complexity of graphics as there are an infinitely many number of graphical displays and design choices that can be implemented at faster speeds with more flexibility. The creator of a graphic makes decisions about the variables displayed, the type of graphic, the size of the graphic and the aspect ratio, the colors and symbols used, the scales and limits, the ordering of categorical

variables, and the ordering of variables in multivariate displays (Unwin, 2020). In response to the increasing number of design choices, consistent themes and higher standards are being placed on graphics. Selecting from an extensive list of styles and choices of graphics in order to effectively communicate insights into the data is a challenging task. A consistent concern is the lack of theory of graphics available to build on; better theory should result in better graphics. Creators of graphics need an established set of concepts and terminology to build their graphics from so they can actively choose which of many possible graphics to draw in order to ensure their charts are effective.

Many efforts have been made to provide guidelines for graphical designs including Wilkinson’s Grammar of Graphics (Wilkinson, 2012). These guidelines provide the ground work necessary for data plots to be depicted and interpreted as statistics (Buja et al., 2009; Majumder, Hofmann, & Cook, 2013). Vanderplas, Cook, & Hofmann (2020) define a statistic as, “a functional mapping of a variable or set of variables.” The grammar of graphics constructs visual statistics through the use of “tidy data,” characterized as a data set in which each variable is in its own column, each observation is in its own row, and each value is in its own cell (Wickham & Grolemund, 2016). The grammar allows variables, as designated in columns, to be mapped to different elements of the graphic such as the axes, colors, shapes, or facets. Software, such as Hadley Wickham’s ggplot2 (Wickham, 2011), aims to implement these guidelines recommended through the grammar of graphics.

Later, it is illustrated how the structure of “tidy data” and the construction of graphics as statistics aid themselves for easy experimentation which allows researchers to compare the effectiveness and understand the perception of different types of charts (Vanderplas, Cook, & Hofmann, 2020). It is important to consider

the purpose and motivation behind the generation of the chart as well as the complexity and intended audience you intend to view the chart. A general guideline when generating graphics is to keep it familiar in order to not intimidate your audience and to encourage further interaction from the viewers (Unwin, 2020). Despite past attempts to improve the use of graphics in science, Gordon & Finch (2015) evaluated 97 graphs for overall quality, based on five principles of graphical excellence, and found there is still an astonishing lack in the quality of graphics. More startling is the fact that the source of the graphic from an applied science or a statistics graphic had no effect on the quality of the graphic. The improvement in graphics depends on both a better definition of variables, units of measurements, scales, and other graphical elements as well as a typical use of grid lines on an accepted set of graphical forms. With the support of changes in software defaults, future work must be done in order to implement the academic research being conducted in graphics into practice for both academics and non-academics in order to achieve a higher standard of the graphics being presented (Vanderplas, Cook, & Hofmann, 2020).

1.2 Perception and Psychophysics

In order to develop guiding principles for generating graphics effective in communication, we must first understand the basic mechanics of the human perceptual system and the biases we are vulnerable to (Goldstein & Brockmole, 2017). The perceptual process is complex and involved [\[INSERT CYCLE PICTURE HERE?\]](#). This process is broken down into sensation - involving simple processes that occur right at the beginning of a sensory system - and perception - involving higher-order mechanisms and identified with more complex processes. The

perceptual process begins when there is a stimulus in the environment and light is reflected and focused back into the viewer's eyes. Within the eye, the light reflected is transformed and focused by the eye's optical system and an image is formed on the receptors of the viewer's retina. It is important to note that everything a person perceives is based not on direct contact with stimuli but on representations of stimuli that are formed on the receptors and the resulting activity in the person's nervous system. Once light is reflected and focused, our visual receptors respond to the light and transform the light energy into electrical energy through a process called transduction. Signals from the receptors are then transmitted through the retina, to the brain, and then within the brain where perception - what do you see? - and recognition - what is it called? - occur. After recognition, viewers take some sort of motor action - for example, move closer to the object. The perceptual process is not direct and instead takes on more of a cyclic nature where a person may go through many iterations of stimuli, perception, recognition, and action before the final image is identified and understood.

When perception occurs, we first experience the preattentive stage in which we observe color, shape, size, and other basic information about the stimuli being perceived. Preattentive perception effects are automatically processed within the first 500 milliseconds of viewing and do not depend on sustained cognitive attention (Vanderplas, Cook, & Hofmann, 2020). Following the preattentive stage, direct attention is required for additional processing to allow us to draw connections between components that assist in our interpretation of the stimuli. When viewing a chart or graph, most insights we gain are due to the cognitive processes that occur after attention is focused on specific aspects of the graph. The knowledge a viewer brings to the situation influences their cognitive evaluation of the stimuli.

The relationship between physiology and perception can provide us information about how graphics may be understood and interpreted. According to a cognitive analysis, graph interpretation involves (a) relatively simple pattern perception and association processes in which viewers can associate graphic patterns to quantitative referents and (b) more complex and error-prone inferential processes in which viewers must mentally transform data (Shah, Mayer, & Hegarty, 1999). Pyschophysics, the branch of psychology that deals with the relationships between physical stimuli (e.g. light) and mental phenomena, aims to provide explanations of this relationship and point out human perceptual biases. By examining both behavior and physiology together, we are able to understand the mechanisms responsible for perception.

1.3 Graphical Experiments

One way in which we determine the relationship between behavior and physiology is through the use of graphical tests. These tests may take many forms: identifying differences in graphs, reading information off of a chart accurately, using data to make correct real-world decisions, or predicting the next few observations. All of these types of tests require different levels of use and manipulation of the information presented in the chart. The initial push to develop classification and recommendation systems for charts was grounded on heuristics rather than on experimentation (Vanderplas, Cook, & Hofmann, 2020). Request were made for the validation of the perception and utility of statistical charts through graphical experiments. Initial experiments struggled with methodological issues (Croxtton & Stein, 1932; Croxtton & Stryker, 1927; Eells, 1926) with most early experimentation stemming from pyschophysics research on the perception of size and shape

(Teghtsoonian, 1965). While a typical psychophysical experiment focuses on whether an effect is detectable and whether the magnitude of the effect can be accurately estimated, these early experiments instead depended on speed and accuracy for plot evaluation (Lewandowsky & Spence, 1989; Spence, 1990; Teghtsoonian, 1965).

Cognitive psychologists and statisticians made progress by conducting experiments to identify perceptual errors associated with different styles of graphics and charts (Cleveland & McGill, 1985; Spence, 1990). These later experiments relied on similar methodology as early studies by relying on participants directly reading information from the charts to provide a quantitative estimate or answering a predefined question; as with the early studies, accuracy and response time being evaluated (Amer, 2005; Broersma & Molenaar, 1985; Cleveland & McGill, 1984; Dunn, 1988; Peterson & Schramm, 1954; Tan, 1994). Cleveland & McGill (1984) provide a basis for perceptual judgment, still utilized today, by examining six basic stimuli: position along a common scale, position along nonaligned scales, length, angle, slope, and area. Other experiments established the notion that redesigning graphs can result in the improvement of the viewer's interpretation (Shah, Mayer, & Hegarty, 1999). This is done by relying on gestalt principles to minimize the inferential processes and maximize the pattern association processes required to interpret relevant information. The viewer must first encode the visual array by identifying meaningful visual features, such as a straight line slanting downward. Next, the viewer must classify the quantitative measures and relationships in which those visual features illustrate, such as a decreasing linear relationship between x and y . The last step involves translating the quantitative measures and relationships to the variables defined in the data set, such as a population decreasing over years. These studies establish the process in which viewers interact with charts by first

perceptually observing the visual features and later translating to cognitive processing of the information depicted by those features (Shah & Carpenter, 1995).

In recent years, there have been advancements in the methodology used to investigate the effectiveness of statistical charts (Vanderplas, Cook, & Hofmann, 2020). Some of the new methods, such as the lineup protocol (Buja et al., 2009), utilize the grammar of graphics designation of a data plot as a statistic through the functional mapping of variable(s). This allows the data plot to be tested similar to other statistics, by comparing the actual data plot to a set of plots with the absence of any data structure we can test the likelihood of any perceived structure being significant (Vanderplas, Cook, & Hofmann, 2020). While the methodology of these recent experiments differs from earlier studies, the focus is still placed on the initial perception and graph comprehension with a relatively small amount of work conducted to understand the effect of design choices on higher cognitive processes such as learning or analysis (Green & Fisher, 2009). Most recent graphics experiments have utilized tools such as Amazon Turk, Prolific, Reddit, and other crowd sourcing websites to evaluate the psychophysics and patterns associated with design choices (VanderPlas & Hofmann, 2017).

- informal and formal graphical exploration. informal: changes are made to the graph and the iterative versions are compared to determine what information is easily acceptable formal: an experiment is designed and participants are tested in a controlled manner. Vanderplas, Cook, & Hofmann (2020)
- There are limits to what one can test using direct estimation: it is generally preferable to test only very straightforward assessments of the content of a chart or graph, to fit within a simple experimental paradigm. Some studies of graphs utilize psychophysics methodology to assess data visualizations.

- Psychophysics and direct observation studies are limited by the questions that are asked; Vanderplas, Cook, & Hofmann (2020)
- Other methods: Thinking aloud, Eye Tracking, etc. Vanderplas, Cook, & Hofmann (2020)

1.4 Logarithmic Scales and Mapping

We have recently experienced the impact graphics and charts have on a large scale through the SARSNCOV-2 pandemic (COVID-19). At the beginning of 2020, we saw an influx of dashboards being developed to display case counts, transmission rates, and outbreak regions (GmbH, 2020); mass media routinely showed charts to share information with the public about the progression of the pandemic (Romano, Sotis, Dominioni, & Guidi, 2020). People began seeking out graphical displays of COVID-19 data as a direct result of these pieces of work (GmbH, 2020); providing increased and ongoing exposure to these graphics over time. Many of these graphics helped guide decision makers to implement policies such as shut-downs or mandated mask wearing, as well as facilitated communication with the public to increase compliance (Bavel et al., 2020). As graphics began to play an important role in the communication of the pandemic, creators of graphics were faced with design choices in order to ensure their charts were effective.

When faced with data which spans several orders of magnitude, we must decide whether to show the data on its original scale (compressing the smaller magnitudes into relatively little area) or to transform the scale and alter the contextual appearance of the data. One common solution is to use a log scale transformation to display data over several orders of magnitude within one graph. Logarithms turn multiplicative relationships additive, showing elasticities and other

proportional changes, and also linearize power laws (Menge et al., 2018). They also have practical purposes, easing the computation of small numbers such as likelihoods and transforming data to fit statistical assumptions. When presenting log scaled data, it is possible to use either un-transformed scale labels (for example, values of 1, 10 and 100 are equally spaced along the axis) or log transformed scale labels (for example, 0, 1, and 2, showing the corresponding powers of 10). We have recently experienced the benefits and pitfalls of using log scales as COVID-19 dashboards displayed case count data on both the log and linear scale (Burn-Murdoch et al., 2020; Fagen-Ulmschneider, 2020). In spring 2020, during the early stages of the COVID-19 pandemic, there were large magnitude discrepancies in case counts at a given time point between different geographic regions (e.g. states and provinces as well as countries and continents). During this time, we saw the usefulness of log scale transformations showing case count curves for areas with few cases and areas with many cases within one chart. As the pandemic evolved, and the case counts were no longer spreading exponentially, graphs with linear scales seemed more effective at spotting early increases in case counts that signaled more localized outbreaks. This is only one recent example of a situation in which both log and linear scales are useful for showing different aspects of the same data; there are long histories of using log scales to display results in ecology, psychophysics, engineering, and physics (Heckler, Mikula, & Rosenblatt, 2013; Menge et al., 2018; Munroe, 2005).

Research suggests our perception and mapping of numbers to a number line is logarithmic at first, but transitions to a linear scale later in development, with formal mathematics education (Dehaene, Izard, Spelke, & Pica, 2008; Siegler & Braithwaite, 2017, 2017; Varshney & Sun, 2013). With basic training, members of

remote cultures with a basic vocabulary and minimal education understood the concept that numbers can be mapped into a spacial space (Dehaene, Izard, Spelke, & Pica, 2008). There was a gradual transition from logarithmic to linear scale as the mapping of whole number magnitude representations transitioned from a compressed (approximately lograithmic) distribution to an approximately linear one. These results indicate the universal and cultural-dependent characteristics of the sense of number. Assuming there is a direct relationship between perceptual and cognitive processes, it is reasonable to assume numerical representations should also be displayed on a nonlinear, compressed number scale. Therefore, if we perceive logarithmically by default, it is a natural (and presumably low effort) way to display information and should be easy to read and understand/use. The idea is compression enlarges the coding space, thus increasing the dynamic range of perception and firing neurons within our visual system (Nieder & Miller, 2003). Similar to the training and education required to transition from lograithmic mapping to linear mapping, there is also necessary training required in the assessment of graphical displays associated with logarithmic scales. Haemer & Kelley (1949) identify semi-logarithmic charts for temporal series as requiring a certain degree of technical training.

Weber's law: the fact that larger numbers require a proportional larger difference in order to remain equally discriminate. Dehaene, Izard, Spelke, & Pica (2008) we do not notice absolute changes in stimuli; we notice relative changes. Sun, Wang, Goyal, & Varshney (2012) Weber–Fechner law. It states that perceived intensity P is logarithmic to the stimulus intensity S (above a minimal threshold of perception S_0) Sun, Wang, Goyal, & Varshney (2012)

1.5 Underestimation of Exponential Growth

Early studies explored the estimation and prediction of exponential growth, finding that growth is underestimated when presented both numerically and graphically but that numerical estimation is more accurate than graphical estimation for exponential curves (Wagenaar & Sagaria, 1975). One way to improve estimation of increasing exponential trends is to provide immediate feedback to participants about the accuracy of their current predictions (Mackinnon & Wearing, 1991). While prior contextual knowledge or experience with exponential growth does not improve estimation, instruction on exponential growth reduces the underestimation: participants adjust their initial starting value but not their perception of growth rate (Jones, 1977; Wagenaar & Sagaria, 1975). Our inability to accurately predict exponential growth might also be addressed by log transforming the data, however, this transformation introduces new complexities; most readers are not mathematically sophisticated enough to intuitively understand logarithmic math and translate that back into real-world effects. In Menge et al. (2018), ecologists were surveyed to determine how often ecologists encounter log scaled data and how well ecologists understand log scaled data when they see it in the literature. Participants were presented two relationships displayed on linear-linear scales, log-log scales with untransformed values, or log-log scales with log transformed values. The authors propose three types of misconceptions participants encountered when presented data on log-log scales: ‘hand-hold fallacy,’ ‘Zeno’s zero fallacy,’ and ‘watch out for curves fallacies.’ These misconceptions are a result of linear extrapolation assuming that a line in log-log space represents a line instead of the power law in linear-linear space. The study found that in each of these scenarios, participants were confident in their incorrect responses, indicating

incorrect knowledge rather than a lack of knowledge. The ‘hand-hold falacy’ stems from the misconception that steeper slopes in log-log relationships are steeper slopes in linear-linear space. In fact, it is not only the slope that matters, but also the intercept and the location on the horizontal axis since a line in log-log space represents a power law in linear-linear space (i.e. linear extrapolation). Emerging from ‘Zeno’s zero fallacy’ is the misconception that positively sloped lines in log-log space can imply a non-zero value of y when x is zero. This is never true as positively sloped lines in log-log space actually imply that $y = 0$ when $x = 0$. This misconception again is a result of linear extrapolation assuming that a line in log-log space represents a line instead of the power law in linear-linear space. The last misconception, ‘watch out for curves fallacies’ encompasses three faults: (1) lines in log-log space are lines in linear-linear space, (2) lines in log-log space curve upward in linear-linear space, and (3) curves in log-log space have the same curvature in linear-linear space. Linear extrapolation is again responsible for the first and third faults while the second fault is a result of error in thinking that log-log lines represent power laws (which are exponential relationships), and all exponential relationships curve upward; this is only true when the log-log slope is greater than 1. Menge et al. (2018) found that in each of these scenarios, participants were confident in their incorrect responses, indicating incorrect knowledge rather than a lack of knowledge.

1.6 Risk Assessment

1.7 Research Objectives

In my research, I conduct three graphical experimental tasks to evaluate the impact our choice of scale (log/linear) has on human perception of exponentially increasing trends. The first experiment evaluates whether our ability to perceptually notice differences in exponentially increasing trends is impacted by the choice of scale. I conducted a visual inference experiment in which participants were shown a series of lineup plots (consisting of 19 null panels and 1 target panel) and asked to identify the panel that was most different from the others. The other experimental tasks focus on determining whether there are cognitive disadvantages to log scales: do log scales make it harder to make use of graphical information? I conducted a graphical task similar to the New York Times “You Draw It” page to test an individual’s ability to use and make predictions for exponentially increasing data. Participants were asked to draw a line using their computer mouse through the increasing exponential trend shown on both scales. In addition to differentiation and prediction of exponentially increasing data, an experimental task was conducted to test an individuals’ ability to translate a graph of exponentially increasing data into real value quantities and extend their estimations by making comparisons. The results of the three experimental tasks provide guidelines for readers to actively choose which of many possible graphics to draw, according to some set of design choices, to ensure that their charts are effective.

Chapter 2

Perception through lineups

2.1 Introduction

To lay a foundation for future exploration of the use of log scales, we begin with the most fundamental ability to identify differences in charts; this does not require that participants understand exponential growth, identify log scales, or have any mathematical training. Instead, we are simply testing the change in perceptual sensitivity resulting from visualization choices.

A statistic is a numerical function which summarizes the data; by this definition, graphs are visual statistics. To evaluate a graph, we have to run our statistic through a visual evaluation - a person. If two different methods of presenting data result in qualitatively different results when evaluated visually, then we can conclude that the visual statistics are significantly different. Recent graphical experiments have utilized statistical lineups to quantify the perception of graphical design choices [VanderPlas and Hofmann, 2017, Hofmann et al., 2012, Loy et al., 2016]. Statistical lineups provide an elegant way of combining perception and statistical hypothesis testing using graphical experiments [Wickham et al., 2010, Majumder et al., 2013, Vanderplas et al., 2020b]. ‘Lineups’ are named after the

‘police lineup’ of criminal investigations where witnesses are asked to identify the criminal from a set of individuals. Similarly, a statistical lineup is a plot consisting of smaller panels; the viewer is asked to identify the plot of the real data from among a set of decoy null plots. A statistical lineup typically consists of 20 panels - 1 target panel and 19 null panels (Figure 1). If the viewer can identify the target panel embedded within the set of null panels, this suggests that the real data is visually distinct from data generated under the null model. Crowd sourcing websites such as Amazon Mechanical Turk, Reddit, and Prolific allow us to collect responses from multiple viewers. In this paper, we use statistical lineups to test our ability to differentiate between exponentially increasing curves with differing levels of curvature, using linear and log scales.

- Explicit graphical tests, as we have referred to them, are tests where the user is directed to assess a specific feature of a plot or answer a specific question. Vanderplas, Cook, & Hofmann (2020) implicit graphical test, the user must identify both the purpose and function of the plot and use that information to evaluate the plots as shown. Visual inference (pg 14) problems. Vanderplas, Cook, & Hofmann (2020) Implicit graphical tests approach the problem of spurious plot relationships at the level of the data, leveraging the human visual system to conduct a suite of visual tests for features such as outliers, clusters, linear and nonlinear relationships. The advantage to implicit testing is that lineups do not require a specification of a feature of interest in the testing framework, Vanderplas, Cook, & Hofmann (2020)

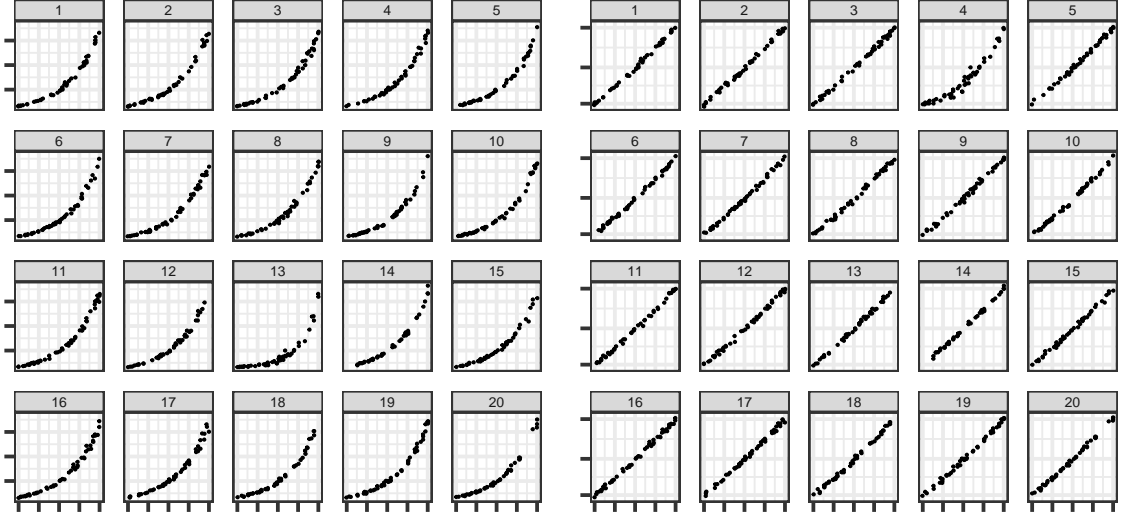


Figure 2.1: Lineup example

2.2 Data Generation

In this study, both the target and null data sets were generated by simulating data from an exponential model; the models differ in the parameters selected for the null and target panels. In order to guarantee the simulated data spans the same domain and range of values, we implemented 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 and mapped to the y-axis using the exponential model with the selected parameters. These constraints provide some assurance that participants who select the target plot 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:

$$y_i = \alpha \cdot e^{\beta \cdot x_i + \epsilon_i} + \theta \quad (2.1)$$

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

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

Algorithm 2.1.1: Parameter Estimation

Input Parameters: domain $x \in [0, 20]$, range $y \in [10, 100]$, midpoint x_{mid} .

Output: estimated model parameters $\hat{\alpha}, \hat{\beta}, \hat{\theta}$

1. 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 two additional points.
3. From the set points (x_k, y_k) for $k = 1, 2, 3, 4$, obtain the coefficients from the linear model $\ln(y_k) = b_0 + b_1 x_k$ to obtain starting values -
 $\alpha_0 = e^{b_0}, \beta_0 = b_1, \theta_0 = 0.5 \cdot \min(y)$
4. Using the `nls()` function from the `stats` package in 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 obtain estimated parameter values - $\hat{\alpha}, \hat{\beta}, \hat{\theta}$.

Algorithm 2.1.2: Exponential Simulation

Input Parameters: sample size $N = 50$, estimated parameters $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$, σ standard deviation from the exponential curve.

Output Parameters: N points, in the form of vectors \mathbf{x} and \mathbf{y} .

1. Generate $\tilde{x}_j, j = 1, \dots, N \cdot \frac{3}{4}$ 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 $\tilde{x}_i, i = 1, \dots, N$ by sampling $N = 50$ values from the set of \tilde{x}_j values.
This gaurantees 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 devaition 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 $e_i \sim N(0, \sigma^2)$.

2.3 Parameter Selection

For each level of difficulty, we simulated 1000 data sets of (x_{ij}, y_{ij}) points for $i = 1, \dots, 50$ and $j = 1 \dots 10$. Each generated x_i point from *Algorithm 2.1.2* was replicated 10 times. Then the lack of fit statistic (LOF) was computed for each simulated data set by calculating the deviation of the data from a linear line. Plotting the density curves of the LOF statistics for each level of difficulty choice allows us to evaluate the ability of differentiating between the difficulty levels and thus detecting the target plot. In Figure 2.2, we can see the densities of each of the

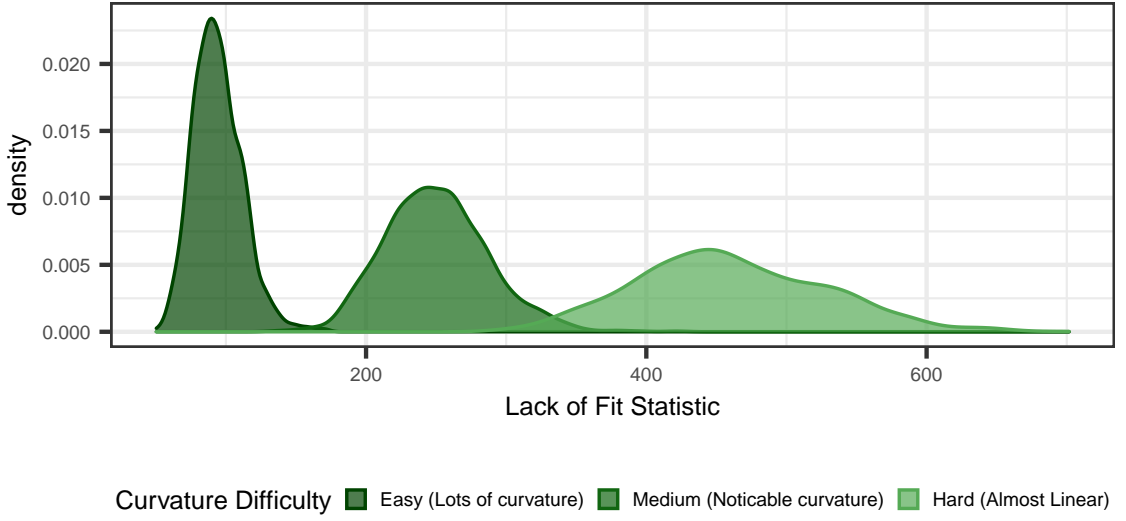


Figure 2.2: Lack of fit statistic density curves

Table 2.1: Lineup data simulation final parameters

	x_{mid}	$\hat{\alpha}$	$\tilde{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\sigma}$
Easy	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
Hard	11.5	37.26	37.22	0.06	-27.26	0.05

three difficulty levels. While the LOF statistic provides us a numerical value for discriminating between the difficulty levels, we cannot directly relate this to the perceptual discriminability; it serves primarily as an approximation to ensure that we are testing parameters at several distinct levels of difficulty.

Final parameter estimates are shown in Table 2.1.

2.4 Lineup Setup

Lineup plots were generated by mapping one simulated data set corresponding to difficulty level A to a scatter plot to be identified as the target plot while multiple simulated data sets corresponding to difficulty level B were individually

mapped to scatter plots for the null plots. For example, a target plot with simulated data following an increasing exponential curve with obvious curvature is embedded within null plots with simulated data following an increasing exponential trend that is almost linear (i.e. Hard Null - Easy Target). By our constraints, the target plot and null plots will span a similar domain and range. There are a total of six (i.e. $3! \cdot 2!$) lineup parameter combinations. Two sets of each lineup parameter combination were simulated (total of 12 test data sets) and plotted on both the linear scale and the log scale (total of 24 test lineup plots). In addition, there are three parameter combinations which generate homogeneous “Rorschach” lineups, where all panels are from the same distribution. Each participant evaluated one of these lineups, but for simplicity, these evaluations are not described in this paper.

2.5 Study Design

Each participant was shown a total of thirteen lineup plots (twelve test lineup plots and one Rorschach lineup plot). Participants were randomly assigned one of the two replicate data sets for each of the six unique lineup parameter combinations. For each assigned test data set, the participant was shown the lineup plot corresponding to both the linear scale and the log scale. For the additional Rorschach lineup plot, participants were randomly assigned one data set shown on either the linear or the log scale. The order of the thirteen lineup plots shown was randomized for each participant.

Participants above the age of majority were recruited from Reddit’s Visualization and Sample Size communities. Since participants recruited on Reddit were not compensated for their time, most participants have an interest in data visualization research. Previous literature suggests that prior mathematical

knowledge or experience with exponential data is not associated with the outcome of graphical experiments (**vanderplasSpatialReasoningData2016?**).

Participants completed the experiment using a Shiny applet (<https://shiny.srvanderplas.com/log-study/>).

Participants were shown a series of lineup plots and asked to identify the plot that was most different from the others. On each plot, participants were asked to justify their choice and provide their level of confidence in their choice. The goal of this experimental task is to test an individuals ability to perceptually differentiate exponentially increasing trends with differing levels of curvature on both the linear and log scale.

2.6 Results

Participant recruitment through Reddit occurred over the course of two weeks during which 58 individuals completed 518 unique test lineup evaluations. Previous studies have found that results do not differ on lineup-related tasks between Reddit and e.g. Amazon Mechanical Turk (VanderPlas & Hofmann, 2017). Participants who completed fewer than 6 lineup evaluations were removed from the study (17 participants, 41 evaluations). The final data set included a total of 41 participants and 477 lineup evaluations. Each plot was evaluated by between 18 and 28 individuals (Mean: 21.77, SD: 2.29). In 67% of the 477 lineup evaluations, participants correctly identified the target panel.

Target plot identification was analyzed using the Glimmix Procedure in SAS 9.4. Each lineup plot evaluated was assigned a value based on the participant response (correct = 1, not correct = 0). The binary response was analyzed using a generalized linear mixed model following a binomial distribution with a logit link

function following a row-column blocking design to account for the variation due to participant and data set respectively. See model details and estimates in [Appendix Reference](#).

On both the log and linear scales, the highest accuracy occurred in lineup plots where the target model and null model had large curvature differences (Easy Null - Hard Target; Hard Null - Easy Target). There is a decrease in accuracy on the linear scale when comparing a target plot with less curvature to null plots with more curvature (Easy Null - Medium Target; Medium Null - Hard Target). Best, Smith, & Stubbs (2007) found that accuracy of identifying the correct curve type was higher when nonlinear trends were presented 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. Overall, there are no significant differences in accuracy between curvature combinations when data is presented on a log scale indicating participants were consistent in their success of identifying the target panel on the log scale. Figure 2.3 displays the estimated (log) odds ratio of successfully identifying the target panel on the log scale compared to the linear scale. The choice of scale has no impact if curvature differences are large (Hard Null - Easy Target; Easy Null - Hard Target). However, presenting data on the log scale makes us more sensitive to slight changes in curvature (Medium Null - Easy Target; Medium Null - Hard Target; Easy Null - Medium Target). An exception occurs when identifying a plot with curvature embedded in null plots close to a linear trend (Hard Null - Medium Target), again supporting the claim that it is easy to identify a curve in a bunch of lines but much harder to identify a line in a bunch of curves (Best, Smith, & Stubbs, 2007).

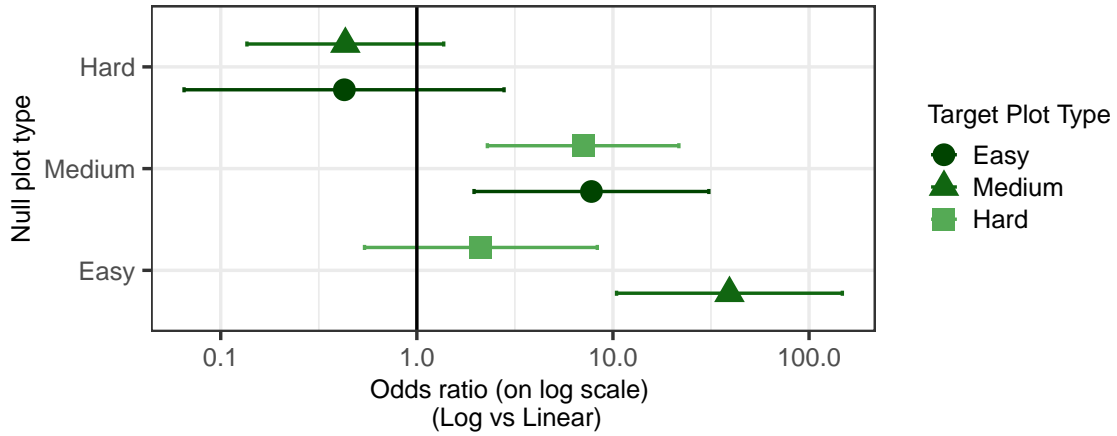


Figure 2.3: Lineups log(odds) results

2.7 Discussion and Conclusion

The overall goal of this paper is to provide basic research to support the principles used to guide design decisions in scientific visualizations of exponential data. In this study, we explore the use of linear and log scales to determine whether our ability to notice differences in exponentially increasing trends is impacted by the choice of scale. Our results indicated that when there was a large difference in curvature between the target plot and 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. An exception occurred when identifying a plot with curvature embedded in surrounding plots closely relating 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. The use of visual inference to identify these guidelines suggests that there are *perceptual* advantages to log scales when differences are subtle. What remains to be seen is whether there are cognitive disadvantages to log scales: do log scales

make it harder to make use of graphical information?

Chapter 3

Prediction with you draw it

3.1 Introduction

- advocate smoothing of scatterplots to assist in detecting the shape of the point cloud in situations where the error in the data is substantial, or where the density of points changes along the abscissa Cleveland & McGill (1984)

Chapter 4

Numerical Translation and Estimation

4.1 Introduction

- Such complex inferential processes involve quantitatively transforming the information in the display (e.g., mentally transforming from a linear to logarithmic scale or calculating the difference between two or more data points; Cleveland, 1984, 1985).

Chapter 5

Conclusion

Appendix A

The First Appendix

Appendix B

The Second Appendix

References

- 10 Amer, T. (2005). Bias due to visual illusion in the graphical presentation of accounting information. *Journal of Information Systems*, 19(1), 1–18.
- Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., ... Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460–471.
<http://doi.org/10.1038/s41562-020-0884-z>
- Best, L. A., Smith, L. D., & Stubbs, D. A. (2007). Perception of Linear and Nonlinear Trends: Using Slope and Curvature Information to Make Trend Discriminations. *Perceptual and Motor Skills*, 104(3), 707–721.
<http://doi.org/10.2466/pms.104.3.707-721>
- Broersma, H., & Molenaar, I. (1985). Graphical perception of distributional aspects of data. *Computational Statistics Quarterly*, 2(1), 53–72.
- Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D. F., & Wickham, H. (2009). 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.
<http://doi.org/10.1098/rsta.2009.0120>

- Burn-Murdoch, J., Nevitt, C., Tilford, C., Rininsland, A., Kao, J. S., Elliott, O., ... Stabe, M. (2020). Coronavirus tracked: Has the epidemic peaked near you? *Coronavirus chart: see how your country compares*. Financial Times.
Retrieved from <https://ig.ft.com/coronavirus-chart/?areas=eur>
- Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods.
Retrieved from <http://euclid.psych.yorku.ca/www/psy6135/papers/ClevelandMcGill1984.pdf>
- Cleveland, W. S., & McGill, R. (1985). Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science, New Series*, 229(4716), 828–833. Retrieved from <http://www.jstor.org/stable/1695272>
- Croxtan, F. E., & Stein, H. (1932). Graphic comparisons by bars, squares, circles, and cubes. *Journal of the American Statistical Association*, 27(177), 54–60.
- Croxtan, F. E., & Stryker, R. E. (1927). Bar charts versus circle diagrams. *Journal of the American Statistical Association*, 22(160), 473–482.
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in western and amazonian indigene cultures. *Science*, 320(5880), 1217–1220.
- Dunn, R. (1988). Framed rectangle charts or statistical maps with shading: An experiment in graphical perception. *The American Statistician*, 42(2), 123–129.
- Eells, W. C. (1926). The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association*, 21(154),

119–132.

Fagen-Ulmschneider, W. (2020). 91-DIVOC. *An interactive visualization of COVID-19*. University of Illinois. Retrieved from <https://91-divoc.com/pages/covid-visualization/>

GmbH, D. (2020, July). You’ve informed the public with visualizations about the coronavirus. Thank you. *Chartable*. Retrieved from <https://blog.datawrapper.de/datawrapper-effect-corona/index.html>

Goldstein, E. B., & Brockmole, J. R. (2017). Sensation and Perception, 480.

Gordon, I., & Finch, S. (2015). Statistician Heal Thyself: Have We Lost the Plot? *Journal of Computational and Graphical Statistics*, 24(4), 1210–1229. <http://doi.org/10.1080/10618600.2014.989324>

Green, T. M., & Fisher, B. (2009). The personal equation of complex individual cognition during visual interface interaction. In *Workshop on human-computer interaction and visualization* (pp. 38–57). Springer.

Haemer, K. W., & Kelley, T. L. (1949). Presentation Problems: Suiting the Chart to the Audience: Common Graphic Devices Classified According to Ease of Reading. *The American Statistician*, 3(5), 11–11. <http://doi.org/10.1080/00031305.1949.10501608>

Heckler, A. F., Mikula, B., & Rosenblatt, R. (2013). Student accuracy in reading logarithmic plots: The problem and how to fix it. In *2013 IEEE Frontiers in Education Conference (FIE)* (pp. 1066–1071). <http://doi.org/10.1109/FIE.2013.6684990>

- Jones, G. V. (1977). Polynomial perception of exponential growth. *Perception & Psychophysics*, 21(2), 197–198. <http://doi.org/10.3758/BF03198726>
- Lewandowsky, S., & Spence, I. (1989). The Perception of Statistical Graphs. *Sociological Methods & Research*, 18(2-3), 200–242. <http://doi.org/10.1177/0049124189018002002>
- Mackinnon, A. J., & Wearing, A. J. (1991). Feedback and the forecasting of exponential change. *Acta Psychologica*, 76(2), 177–191. [http://doi.org/10.1016/0001-6918\(91\)90045-2](http://doi.org/10.1016/0001-6918(91)90045-2)
- Majumder, M., Hofmann, H., & Cook, D. (2013). Validation of Visual Statistical Inference, Applied to Linear Models. *Journal of the American Statistical Association*, 108(503), 942–956. <http://doi.org/10.1080/01621459.2013.808157>
- Menge, D. N. L., MacPherson, A. C., Bytnerowicz, T. A., Quebbeman, A. W., Schwartz, N. B., Taylor, B. N., & Wolf, A. A. (2018). Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature Ecology & Evolution*, 2(9), 1393–1402. <http://doi.org/10.1038/s41559-018-0610-7>
- Munroe, R. (2005). Log scale. *xkcd: A webcomic of romance, sarcasm, math, and language*. Creative Commons. Retrieved from <https://xkcd.com/1162/>
- Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1), 149–157.
- Peterson, L. V., & Schramm, W. (1954). How accurately are different kinds of graphs read? *Audio Visual Communication Review*, 178–189.

- Romano, A., Sotis, C., Dominioni, G., & Guidi, S. (2020). *The Scale of COVID-19 Graphs Affects Understanding, Attitudes, and Policy Preferences* ({SSRN} {Scholarly} {Paper} No. ID 3588511). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=3588511>
- Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, 124(1), 43.
- Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, 91(4), 690.
- Siegler, R. S., & Braithwaite, D. W. (2017). Numerical Development. *Annual Review of Psychology*, 68(1), 187–213. <http://doi.org/10.1146/annurev-psych-010416-044101>
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 683–692. <http://doi.org/10.1037/0096-1523.16.4.683>
- Sun, J. Z., Wang, G. I., Goyal, V. K., & Varshney, L. R. (2012). A framework for Bayesian optimality of psychophysical laws. *Journal of Mathematical Psychology*, 56(6), 495–501. <http://doi.org/10.1016/j.jmp.2012.08.002>
- Tan, J. K. (1994). Human processing of two-dimensional graphics: Information-volume concepts and effects in graph-task fit anchoring frameworks. *International Journal of Human-Computer Interaction*, 6(4), 414–456.

- Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, 78(3), 392–402.
- Unwin, A. (2020). Why is Data Visualization Important? What is Important in Data Visualization? *Harvard Data Science Review*.
<http://doi.org/10.1162/99608f92.8ae4d525>
- Uri, G., & Haemer, K. (1948). Presentation Problems: Graphic Representation for Multiple Linear Correlation. *The American Statistician*, 2(6), 19–19.
<http://doi.org/10.1080/00031305.1948.10483416>
- Vanderplas, S., Cook, D., & Hofmann, H. (2020). Testing Statistical Charts: What Makes a Good Graph? *Annual Review of Statistics and Its Application*, 7(1), 61–88. <http://doi.org/10.1146/annurev-statistics-031219-041252>
- VanderPlas, S., & Hofmann, H. (2017). Clusters Beat Trend!? Testing Feature Hierarchy in Statistical Graphics. *Journal of Computational and Graphical Statistics*, 26(2), 231–242. <http://doi.org/10.1080/10618600.2016.1209116>
- Varshney, L. R., & Sun, J. Z. (2013). Why do we perceive logarithmically? *Significance*, 10(1), 28–31. <http://doi.org/10.1111/j.1740-9713.2013.00636.x>
- Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, 18(6), 416–422.
<http://doi.org/10.3758/BF03204114>
- Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2), 180–185.
- Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. ” O’Reilly Media, Inc.”.

Wilkinson, L. (2012). The grammar of graphics. In *Handbook of computational statistics* (pp. 375–414). Springer.