

HUMAN PERCEPTION OF EXPONENTIALLY INCREASING DATA  
DISPLAYED ON A LOG SCALE EVALUATED THROUGH EXPERIMENTAL  
GRAPHICS TASKS

by

Emily Anna Robinson

A DISSERTATION

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfillment 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 EVALUATED THROUGH EXPERIMENTAL  
GRAPHICS TASKS

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. We conducted a series of three graphical studies to evaluate the impact displaying data on the log scale has on human perception of exponentially increasing trends compared to displaying data on the linear scale. Each study was related to a different graphical task, each requiring a different level of interaction and cognitive use of the data being presented. The first experiment evaluated whether our ability to perceptually notice differences in exponentially increasing trends is impacted by the choice of scale. Participants were shown a set of plots and asked to identify which plot appeared to differ most from the other plots. Results indicated the choice of scale changes the contextual appearance of the data leading to slight perceptual advantages for both scales depending on the curvatures of the trend lines being compared. The second study validated a new method, ‘You Draw It’, for testing statistical graphics and introduced an appropriate statistical analysis method for comparing visually fitted trend lines to statistical regression results. This new method was then used to test participants ability to make forecast predictions for exponentially increasing trends on both scales. The results from the analysis showed a clear underestimation of forecasting trends with high exponential growth rates when participants were asked to make predictions on the linear scale;

improvement in forecasts were made when participants were asked to make predictions on the log scale. The third study evaluated graph comprehension as it relates to the contextual scenario of the data shown. Overall, our results suggested that log logic is difficult and that anchoring and rounding biases result in a sacrifice in accuracy in estimates made on the log scale for large magnitudes. The studies conducted in this research relied on graphical tasks of varying complexity to help us understand the perceptual and cognitive advantages and disadvantages of displaying exponentially increasing data on the log scale. The results are instrumental in establishing guidelines for making design choices about scale which result in data visualizations effective at communicating the intended result.

COPYRIGHT

© 2022, Emily Anna Robinson

## Acknowledgments

Thank you to all my people!

## Table of Contents

<b>List of Figures</b>	viii
<b>List of Tables</b>	xii
<b>1 Literature Review</b>	1
1.1 Motivation and Background . . . . .	1
1.2 Misleading Graphics . . . . .	2
1.3 Graphical Frameworks . . . . .	8
1.4 Heuristics and “Good” Graphics . . . . .	10
1.5 Testing Statistical Graphics . . . . .	12
1.6 Task Complexity . . . . .	17
1.7 Graph Comprehension . . . . .	18
1.8 Logarithmic Scales and Mapping . . . . .	23
1.9 Underestimation of Exponential Growth . . . . .	28
1.10 Research Objectives . . . . .	32
<b>2 Perception through lineups</b>	34
2.1 Introduction . . . . .	34
2.2 Visual Inference . . . . .	34
2.3 Data Generation . . . . .	36
2.4 Parameter Selection . . . . .	37

2.5	Lineup Setup . . . . .	39
2.6	Study Design . . . . .	41
2.7	Results . . . . .	42
2.8	Discussion and Conclusion . . . . .	44
<b>3</b>	<b>Prediction with ‘You Draw It’</b>	<b>47</b>
3.1	Introduction . . . . .	47
3.2	Study Design . . . . .	55
3.3	Eye Fitting Straight Lines in the Modern Era . . . . .	56
3.4	Prediction of Exponential Trends . . . . .	64
3.5	Discussion and Conclusion . . . . .	71
<b>4</b>	<b>Numerical Translation and Estimation</b>	<b>74</b>
4.1	Introduction . . . . .	74
4.2	Study Design . . . . .	76
4.3	Data Generation . . . . .	78
4.4	Results . . . . .	81
4.5	Discussion and Conclusion . . . . .	116
<b>5</b>	<b>Conclusion</b>	<b>118</b>
<b>A</b>	<b>You Draw It Setup with Shiny</b>	<b>123</b>
<b>B</b>	<b>Exponential Prediction Interactive Plots</b>	<b>125</b>
<b>C</b>	<b>Scratchwork participant comparison</b>	<b>130</b>
	<b>References</b>	<b>132</b>

## List of Figures

1.1	91-DIVOC new daily case counts as of July 2021 . . . . .	2
1.2	COVID-19 risk level map as of July 2020 . . . . .	3
1.3	Datawrapper daily chart views during COVID-19 . . . . .	3
1.4	Misleading graphical form examples . . . . .	4
1.5	Misleading 'chartjunk' examples . . . . .	5
1.6	Misleading bad axes . . . . .	6
1.7	Misleading graphic multiple violations example . . . . .	7
1.8	COVID-19 reporting . . . . .	7
1.9	Graphic flowchart . . . . .	9
1.10	Statistical Atlas of the United States (1870) . . . . .	10
1.11	Logarithmic scales ASME guidelines . . . . .	12
1.12	Eells (1926) component parts diagrams . . . . .	14
1.13	Spence (1990) task display . . . . .	15
1.14	Amer (2005) cost volume profit graph . . . . .	19
1.15	Comprehension of heights, Curcio (1987) . . . . .	20
1.16	Linear scale verses log scale . . . . .	24
1.17	Covid-19 deaths (log scale) as of March 23, 2020 . . . . .	26
1.18	Covid-19 case counts (linear scale) as of June 30, 2020 . . . . .	26
1.19	Covid-19 case counts (log scale) as of June 30, 2020 . . . . .	27
1.20	Kindergarten example of mapping numbers 1-10 along a number line . . . . .	28

1.21 Log scale comic . . . . .	29
1.22 Graphs viewed in Menge (2018) survey . . . . .	31
2.1 Lineup examples . . . . .	36
2.2 Lineup parameter selection . . . . .	39
2.3 Lineup curvature combinations . . . . .	40
2.4 Lineups log(odds) results . . . . .	45
3.1 Subjective Judgement in Statistical Analysis (1951) parallel probits . . . . .	50
3.2 Eye Fitting Straight Lines (1981) data sets . . . . .	51
3.3 New York Times 'You Draw It' feature . . . . .	52
3.4 SVG vs raster . . . . .	53
3.5 'You Draw It' task plot example . . . . .	54
3.6 Eye Fitting Straight Lines in the Modern Era simulated data example . . . . .	59
3.7 OLS vs PCA regression lines . . . . .	60
3.8 Eye Fitting Straight Lines in the Modern Era feedback data example . . . . .	61
3.9 Eye Fitting Straight Lines in the Modern Era LMM results . . . . .	65
3.10 Eye Fitting Straight Lines in the Modern Era GAMM results . . . . .	66
3.11 Exponential prediction spaghetti plot . . . . .	69
3.12 Exponential prediction GAMM results . . . . .	72
4.1 Estimation simulated data . . . . .	80
4.2 Estimation word cloud . . . . .	82
4.3 Elementary Q1 sketch . . . . .	84
4.4 Elementary Q1 density in year 10 . . . . .	86
4.5 Estimated population spaghetti plot: data set 1 . . . . .	88
4.6 Estimated population spaghetti plot: data set 2 . . . . .	89

4.7	Estimated population: common responses for year 40 . . . . .	93
4.8	Elementary Q2 sketch . . . . .	94
4.9	Elementary Q2 density . . . . .	96
4.10	Intermediate Q1 sketch . . . . .	99
4.11	Intermediate Q1 density (data set 1) . . . . .	100
4.12	Intermediate Q1 density (data set 2) . . . . .	101
4.13	Intermediate Q1 common responses . . . . .	104
4.14	Intermediate Q2 sketch . . . . .	107
4.15	Intermediate Q2 observed plot . . . . .	109
4.16	Intermediate Q2 density . . . . .	111
4.17	Intermediate Q2 common responses . . . . .	112
4.18	Intermediate Q3 sketch . . . . .	114
4.19	Intermediate Q3 density . . . . .	115
A.1	Interactive plot development . . . . .	124
B.1	Exponential prediction plot (low; 50%; linear) . . . . .	125
B.2	Exponential prediction plot (low; 50%; log) . . . . .	126
B.3	Exponential prediction plot (low; 75%; linear) . . . . .	126
B.4	Exponential prediction plot (low; 75%; log) . . . . .	127
B.5	Exponential prediction plot (high; 50%; linear) . . . . .	127
B.6	Exponential prediction plot (high; 50%; log) . . . . .	128
B.7	Exponential prediction plot (high; 75%; linear) . . . . .	128
B.8	Exponential prediction plot (high; 75%; log) . . . . .	129
C.1	Estimation showed work density plot comparison . . . . .	131

## List of Tables

2.1 Lineup data simulation final parameters . . . . .	38
3.1 Eye Fitting Straight Lines in the Modern Era simulation model parameters	57
4.1 Estimation questions . . . . .	78
C.1 Estimation showed work summary . . . . .	130

## Chapter 1

### Literature Review

#### 1.1 Motivation and Background

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 (Rost, 2020); mass media routinely showed charts to share information with the public about the progression of the pandemic (Romano, Sotis, Dominion, & Guidi, 2020). Fagen-Ulmschneider (2020) began the 91-DIVOC project to explore the global growth of COVID-19 through interactive graphics updated daily. The interactive graphics allowed viewers to explore the current status of COVID-19 by selecting their desired regions, axes, axis scale, and measure of interest (for example, case count, death count, and vaccine count); Fig. 1.1 (Fagen-Ulmschneider, 2020) shows the new confirmed COVID-19 cases per day, normalized by population, as of July 2021. Other graphics displayed COVID-19 data as maps (Fig. 1.2) with color indicating the severity and risk in each US county (“Risk levels,” 2021). People began seeking out graphical displays of COVID-19 data as a direct result of these pieces of work (Rost, 2020); providing increased and ongoing exposure to these graphics over time. Fig. 1.3 illustrates the increased

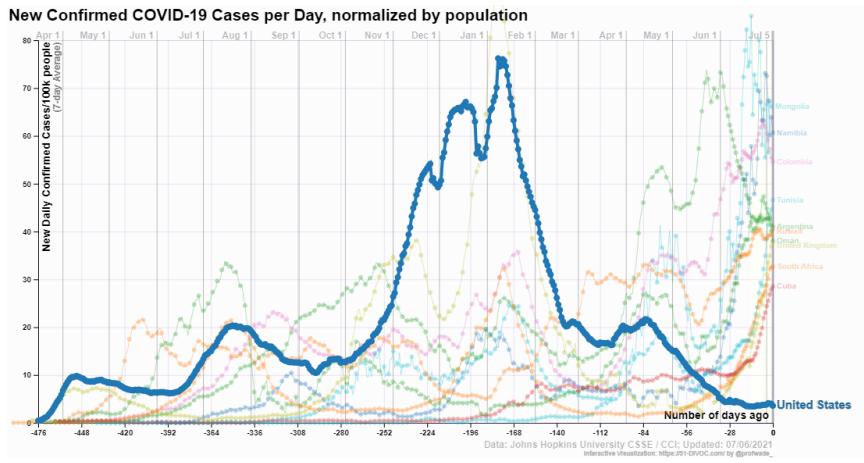


Figure 1.1: New daily COVID-19 case counts as of July 2021 shown in the 91-DIVOC dashboard.

views Datawrapper, a user-friendly web tool used to create basic interactive charts, had during the COVID-19 pandemic (Rost, 2020). 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 sharing information with the public, creators of graphics were faced with design choices in order to ensure their charts were effective at accurately communicating the current status of the pandemic. In order to make educated decisions when designing a chart, we need guidelines as established through experimentation to ensure the graphic is effective at communicating the intended result.

## 1.2 Misleading Graphics

There are many ways in which plots may inaccurately display the data and be ineffective or misleading in sharing information and results (Szafir, 2018).

Misleading charts might have (1) bad form such as 3D pie charts or a plot type that is unsuitable for the type of data (Fig. 1.4) (Reddit: r/dataisugly, n.d.;

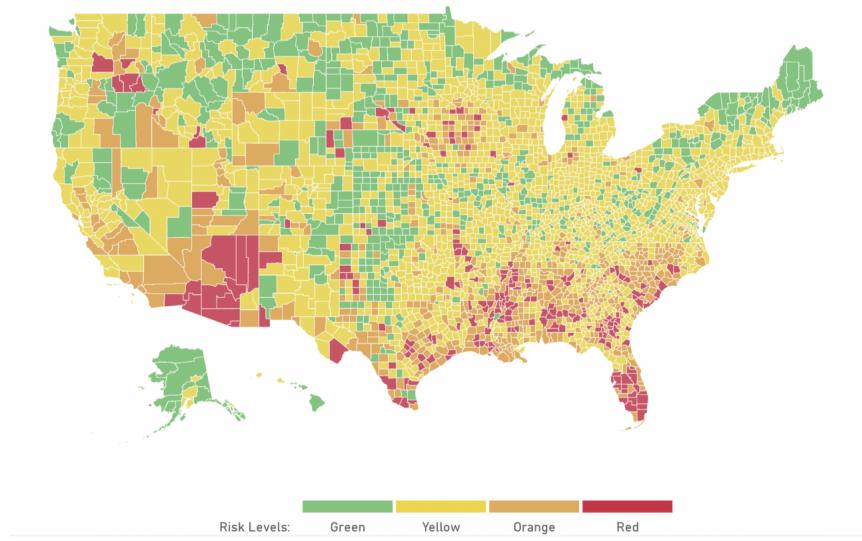


Figure 1.2: COVID-19 risk level map as of July 2020.

### More views than ever

Making it easier to create beautiful, meaningful visualizations has been Datawrapper's mission from day one. The COVID-19 health crisis reminded us all how important such visualizations can be:

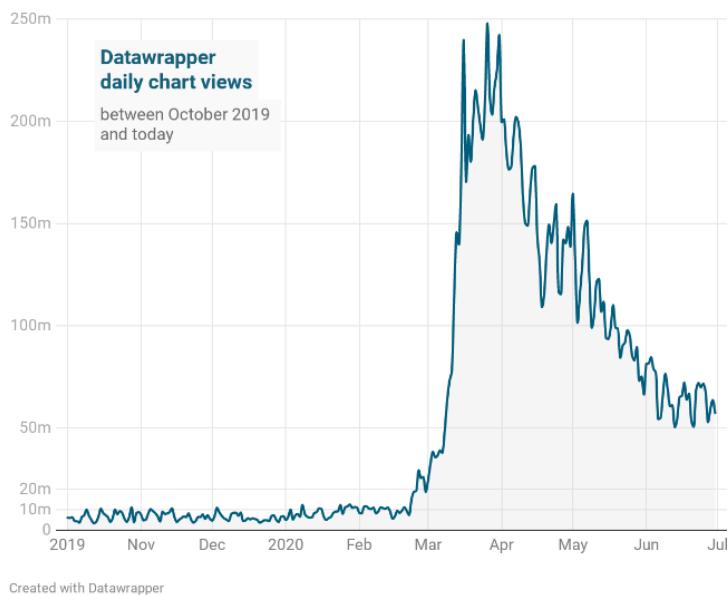


Figure 1.3: Datawrapper daily chart views during COVID-19.

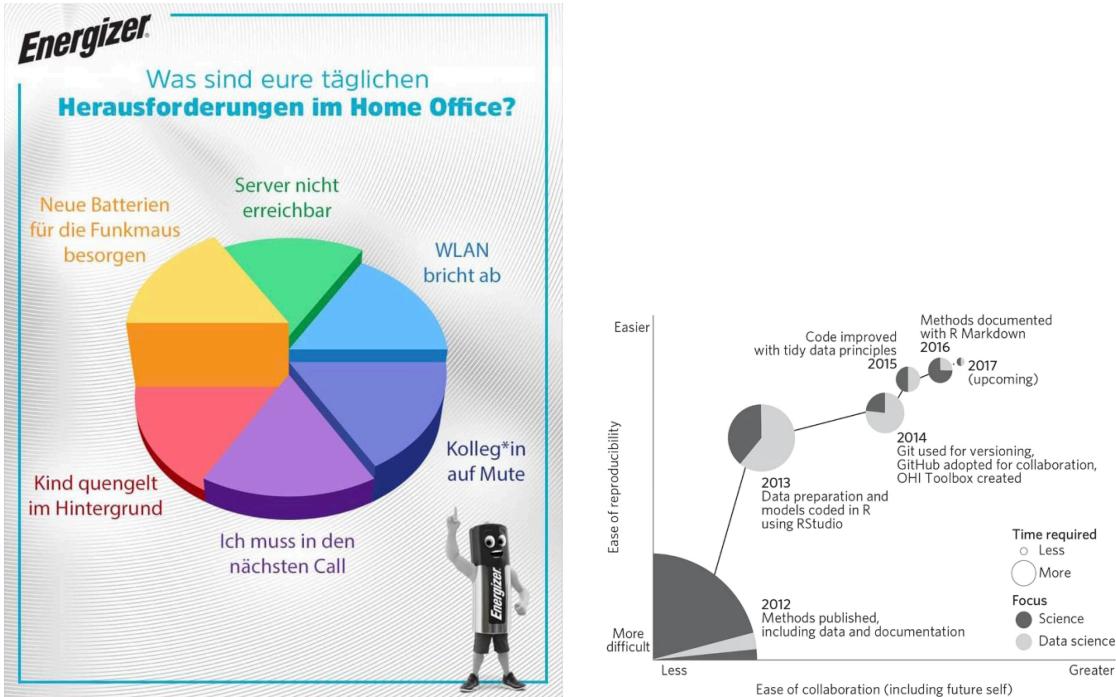


Figure 1.4: These figures display information in bad form which results in an ineffective chart.

Stewart-Lowndes et al., 2017), (2) include too much “chartjunk”, resulting in clutter or displaying useless data (Fig. 1.5) (Dongarra, Meuer, Strohmaier, et al., 1997; Twitter: @GraphCrimes, 2022), or (3) have bad axes such as a mismatch between scale and context or just plain bad math (Fig. 1.6) (Lauren, 2018; Villa, 2022). Fig. 1.7 demonstrates how a chart can be misleading in more than one way; the map of pet ownership violates guidelines by mapping too many variables to visual encodings which results in clutter and bad form (Reddit: r/CrappyDesign, n.d.).

Baumer, Kaplan, & Horton (2021) shares an example of a misleading graphic in the news shown in May 2020 when Georgia published a graphical display of COVID-19 cases (Fig. 1.8). This graphic was highly misleading in communicating the state of the pandemic due to the ordering along the *x*-axis. Notice the case count for April 17<sup>th</sup> appears to the right of April 19<sup>th</sup>, and that the order of the

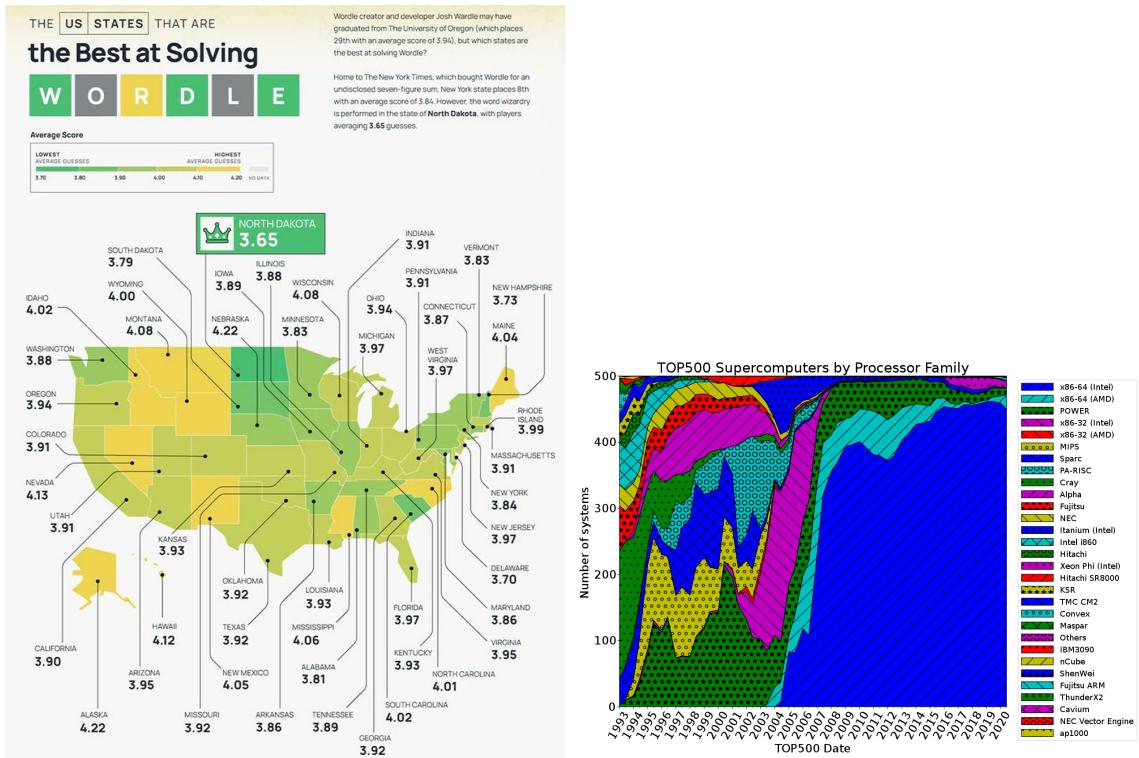


Figure 1.5: These figures provide examples of how too much clutter can be misleading.

counties has been selected so that the case counts are monotonically decreasing for each day of reporting. The appearance of this graphic leads viewers to believe COVID cases are decreasing. Shortly after the graphic was released, the governor's office made a statement that in future charts, chronological order would be used to display time due to public demand.

Misleading charts are not only found in mass media, but graphics displayed in academic research and science are still falling short of the standards. Gordon & Finch (2015) evaluated 97 graphs for overall quality, based on five principles of graphical excellence including: (1) show the data clearly (2) use simplicity in design (3) use good alignment on a common scale for quantities to be compared (4) keep the visual encoding transparent (5) use graphical forms consistent with principles 1 and 4. The authors rated 39% of the 97 graphs sampled as poor, indicating there is

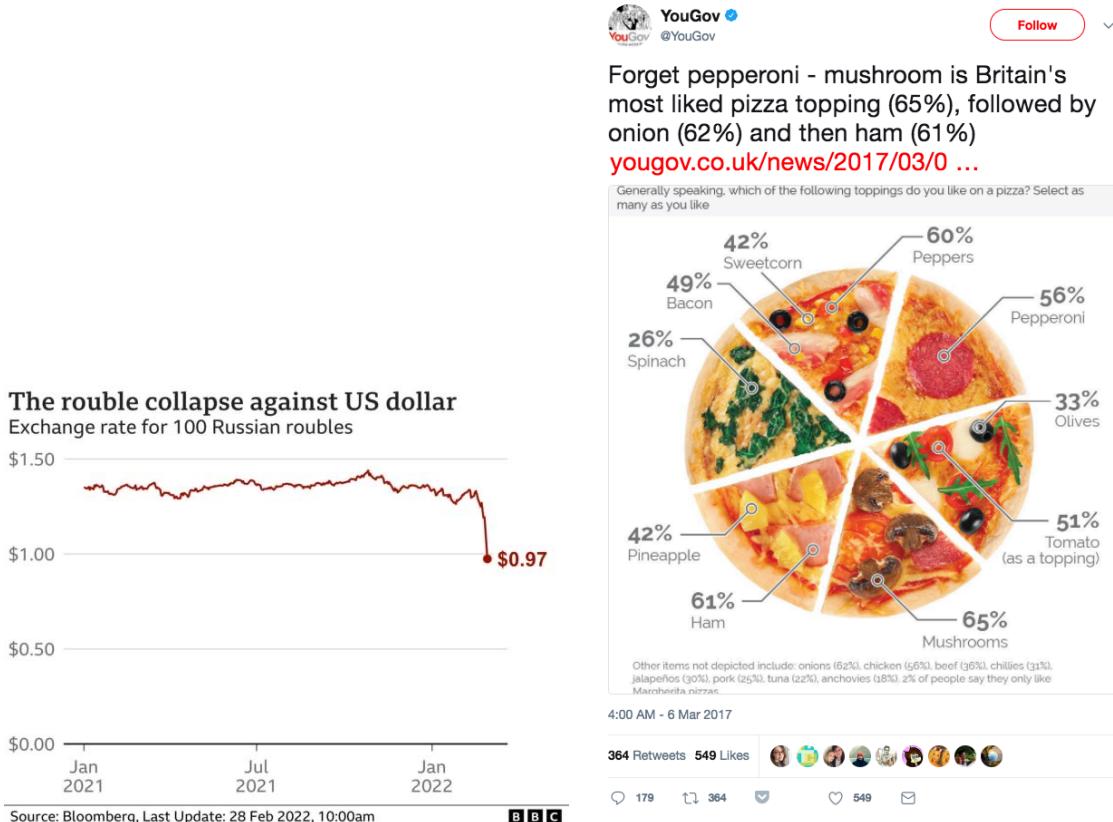


Figure 1.6: These figures are misleading in their axis selections and bad math. (Left) Selecting a baseline  $y$ -axis value of 0 minimizes the large decline in the rate of the rouble in 2022. (Right) The probabilities sum to greater than 100%.

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 graphic from statistics had no effect on the quality of the graphic.

Although statistical graphics have become widely used and valued in science, business, and in many other aspects of life, we may be too accepting of easy-to-create, default data displays, using them without critically questioning the data and/or how effective the display is at displaying the data. Attempts to improve the creation and use of charts have been ongoing since the early 20<sup>th</sup> century (Croxton & Stein, 1932; Croxton & Stryker, 1927; Eells, 1926; Von Huhn,

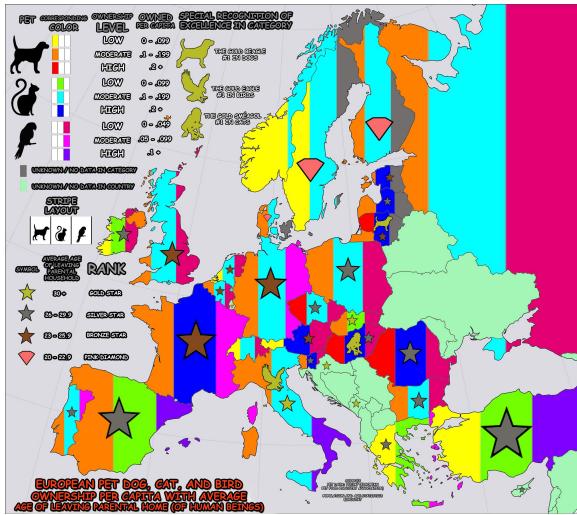


Figure 1.7: This figure demonstrates how a chart can be misleading in more than one way; the map of pet ownership violates guidelines by mapping too many variables to visual encodings which results in clutter and bad form.

#### Top 5 Counties with the Greatest Number of Confirmed COVID-19 (Reproduction of Figure)

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.

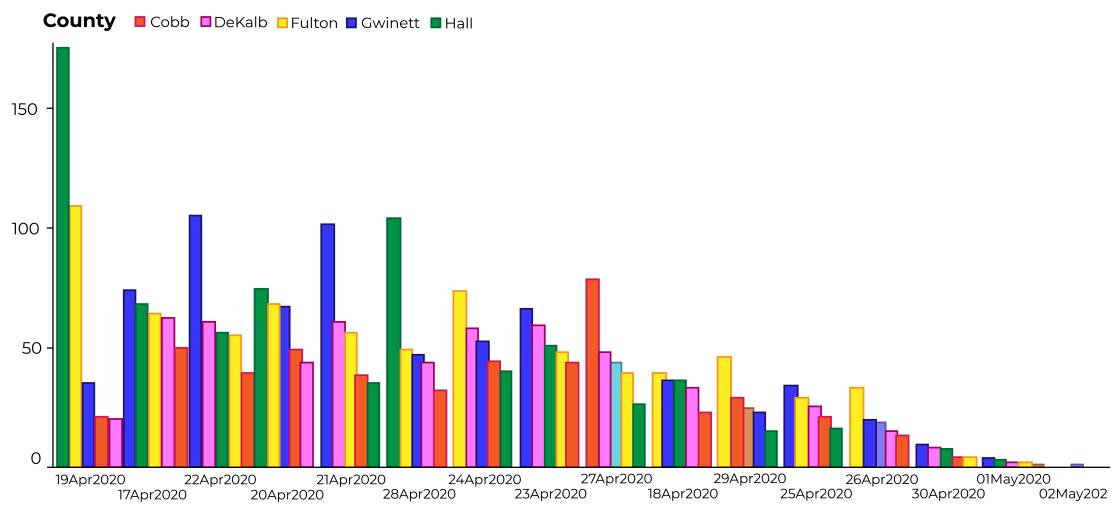


Figure 1.8: Misleading graphic example displaying COVID-19 cases in Georgia as of May 2020. Notice the case count for April 17<sup>th</sup> appears to the right of April 19<sup>th</sup>, and that the order of the counties has been selected so that the case counts are monotonically decreasing for each day of reporting.

1927). We address the ways in which researchers use experimentation to test statistical graphics in order to establish guidelines and improve graphics in section 1.4. In efforts to achieve a higher standard of the graphics being presented, work must be done to implement the academic research being conducted in graphics into practice. For example, better definitions of variables, units of measurements, scales, and other graphical elements is necessary in order to improve the overall quality of graphics. Changes in software defaults such as the originally set number of bins in a bar chart can help support the improvement of graphs in both statistics and the applied science.

### 1.3 Graphical Frameworks

A consistent concern is the lack of theory of graphics available to build on; better theory should result in better data visualizations. In order to improve the construction and distribution of charts, we need an established set of concepts and terminology so we can actively choose which of many possible graphics to draw in order to ensure our charts are effective at communicating the intended result. Many efforts have been made to provide frameworks and classification systems for graphical designs; the most useful for our purposes is Wilkinson’s Grammar of Graphics (Wilkinson, 2013). The grammar of graphics serves as the fundamental framework for data visualization with the notion that graphics are built from the ground up by specifying exactly how to create a particular graph from a given data set. Visual representations are constructed through the use of “tidy data” which is 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). Graphics are viewed as a mapping from variables in a data set

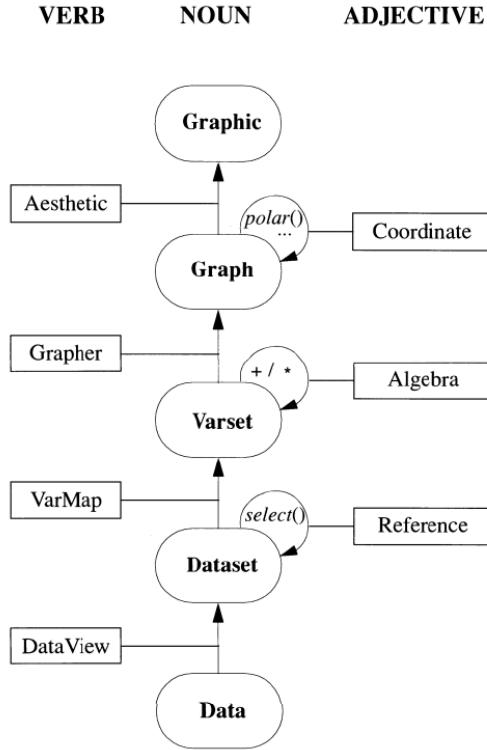


Figure 1.9: The flowchart illustrates the process of creating a graphic from a data set through the use of variable mapping, data transformations, coordinate systems, and aesthetic features.

(or statistics computed from the data) to visual attributes such as the axes, colors, shapes, or facets on the canvas in which the chart is displayed. Fig. 1.9 illustrates the process of creating a graphic from a data set through the use of variable mapping, data transformations, coordinate systems, and aesthetic features (Vanderplas, Cook, & Hofmann, 2020). Software, such as `ggplot2` (Wickham, 2016), aims to implement the framework of creating charts and graphics using the layered framework of the grammar of graphics.

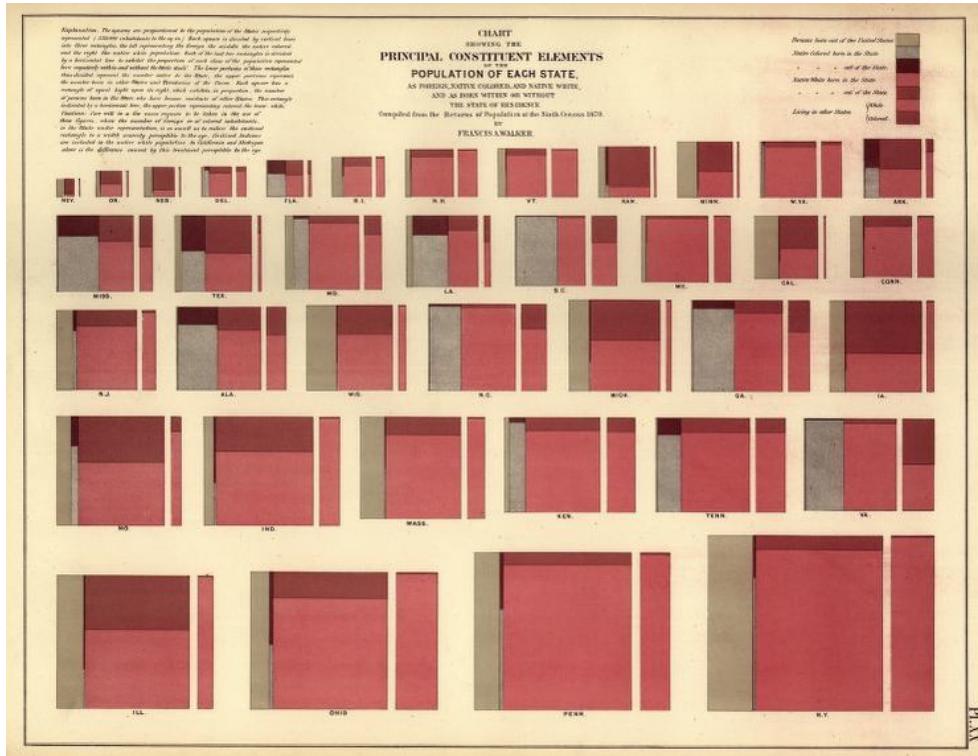


Figure 1.10: The Statistical Atlas of the United States (1870) produced high quality graphics. This figure displays the population of each state where square size represents the proportion of the states population separated into three regions representing the origin and race of the population. The rectangle shown to the right represents the proportion of residents born in the state who have become residents of other states.

## 1.4 Heuristics and “Good” Graphics

Charts have been an essential component in communicating information for the last 200 years (Lewandowsky & Spence, 1989). Some of these charts succeeded in effectively showing the data in order for viewers to extract meaningful information. For example, during 1870, 1880, and 1890, the “Statistical Atlas of the United States” (Walker, 1874) produced high quality and engaging graphics (Fig. 1.10). These data visualizations were created without the use of modern technology, demonstrating that exceptional graphics were achievable before the use of computers.

In the following decades, recommendations and guidelines emerged in order to help improve the overall quality of graphics. Wickham (2013) gives a review and critique of the first formal advice for creating good graphics presented by The International Institute of Statistics in 1901 and Andrews (2022) recalls seventeen general suggestions for the visualization of statistical and quantitative data published by The American Society of Mechanical Engineers (ASME) in 1915. Here we take a look at select guidelines from these articles and connect them to work conducted almost 100 years after the recommendations were made.

When creating graphics, early guidelines state to keep symbols to a minimum; Tufte (1985) coins the term “chartjunk” which refers to any of the visual elements in the chart that are unnecessary for the viewer to comprehend the information represented in the plot. The ASME guidelines supported the minimization of “chartjunk” earlier that century by recommending that charts do not show any more coordinate lines than necessary to guide the eye in reading the diagram. Guidelines for the use of lengths over areas or volumes were established in both sets of early recommendations, but not formally tested until the late 20<sup>th</sup> century by W. S. Cleveland & McGill (1987). The International Institute of Statistics included a recommendation for selecting the ratio of the scales such that the slope of the phenomenon corresponds to the tangent of the curve at a 45° angle. Almost 100 years later, W. S. Cleveland, McGill, & McGill (1988) again makes the recommendation to “bank to 45°” and Heer & Agrawala (2006) explores extensions to propose alternate optimization criteria as well as introduce a technique meant to implement the original guideline. Additional suggestions include the proper use of scales such as arrangements, labels, and baselines; the ASME advocates for careful consideration of the limiting lines for curves drawn on logarithmic coordinates

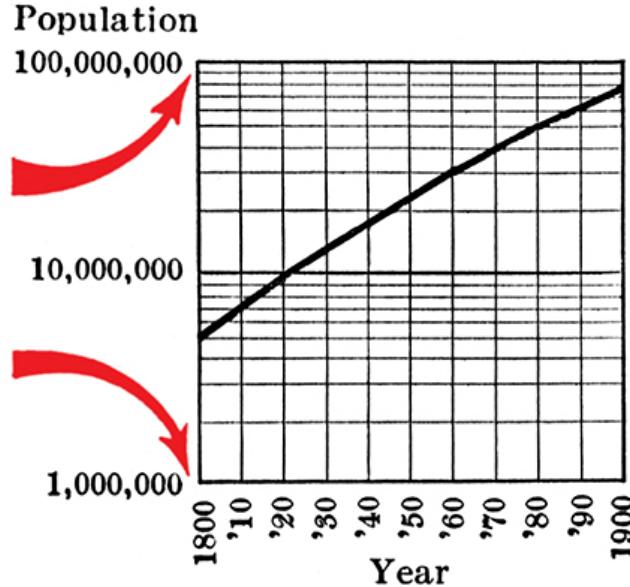


Figure 1.11: The ASME (1915) recommended when curves are drawn on logarithmic coördinates, the limiting lines of the diagram should each be at some power of ten on the logarithmic scales. Note the use of the minor gridline breaks unequally spaced visually, but equally spaced numerically.

(Fig. 1.11).

Helpful suggestions for creating “good” graphics are extremely common (and sometimes conflicting). In the absence of an underlying understanding of how graphics are perceived and used, these suggestions are largely ineffective and sometimes harmful. It is essential to have guidelines established through careful experimentation combined with an understanding of the perceptual and cognitive processes involved in the use of statistical graphics.

## 1.5 Testing Statistical Graphics

One way in which we establish guidelines is through the use of graphical tests (W. Cleveland & McGill, 1984; Lewandowsky & Spence, 1989; Spence, 1990;

VanderPlas & Hofmann, 2015). These tests may take many forms: identifying differences in graphs, accurately reading information off a chart, 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 (Kruskal, 1975; Macdonald-Ross, 1977). Requests were made for the validation of the perception and utility of statistical charts through graphical experiments. Most early experimentation (Croxton & Stein, 1932; Croxton & Stryker, 1927; Eells, 1926) stemmed from psychophysics research on the perception of size and shape. In attempts to understand the human perception and judgment of component parts, Eells (1926) instructed students to think of each circle diagram (Fig. 1.12) as representing 100% and write their best estimate of the percentage of the whole in each sector. Participants were told not to hurry, but to work steadily in order to determine efficiency of judgment. Students were then asked to analyze their mental processes used to make their estimates and indicate the method that best matches: by areas of sectors, by central angles, by arcs on the circumference, by subtending chords. This process was repeated three days later by presenting students the same data represented in bar diagrams (Fig. 1.12). Results of the study led the authors to argue for the use of circle diagrams to show component parts based on both participant accuracy and speed. In response, Croxton & Stryker (1927) evaluated the accuracy of judgment of two types of charts (bars and circles) in efforts to reach a consistent conclusion. During class, students were individually presented pairs of diagrams (without scales) on cards and asked to estimate the percentages displayed

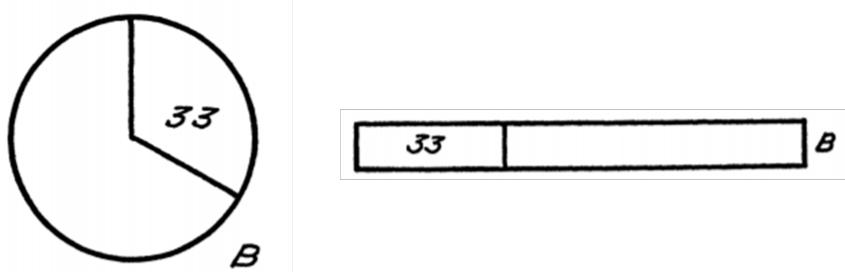


Figure 1.12: Component part diagrams shown to study participants in Eells (1926). Researchers were interested in comparing the partition percentage estimate between circle and bar diagrams.

in the diagram. It was found that the bar was preferable to the circle when shown percentages that deviate from quarters, but that the circle is strongly preferred when shown percentages separating the diagrams into 25% or 50%; this introduces the concept of anchoring discussed further in [section 1.6.3](#).

While a typical psychophysics 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). In attempts to understand the visual psychophysics of simple graphical elements, Spence (1990) presented stimuli (tables, lines - horizontal and vertical, bars, boxes, cylinders, pie charts, and disk charts) to participants on a monitor screen in a computer lab. Participants were asked to use their cursor to position the marker to indicate the proportion to the apparent sizes of the elements (Fig. [1.13](#)). Results found that the table elements (numbers), pie elements, and bar elements led to the most accurate proportion estimates; boxes and disk elements resulted in the least accurate estimates. Measuring the speed at which participants made their judgments, it was found that two- and three- dimensional stimuli (for example, pie charts and box charts) assisted in faster judgment than zero- or one- dimensional stimuli (for

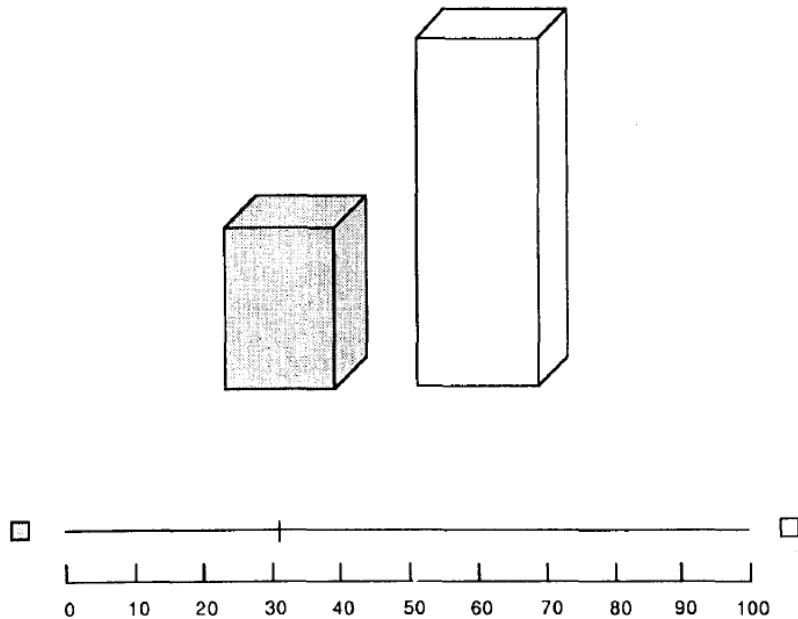


Figure 1.13: Example of a stimuli shown to study participants in Spence (1990). Participants were asked to use their cursor to position the marker to indicate the proportion to the apparent sizes of the elements.

example, tables and lines).

Cognitive psychologists and statisticians made progress by conducting experiments to identify perceptual errors associated with different styles of graphics and charts (W. Cleveland & McGill, 1984, 1985; Shah, Mayer, & Hegarty, 1999). W. Cleveland & McGill (1984) provides a basis for perceptual judgment, still utilized today, by examining six basic plot objects: position along a common scale, position along nonaligned scales, length, angle, slope, and area. In W. Cleveland & McGill (1985), these plot objects are ordered by accuracy performed through graphical-perception tasks; for example, comparisons of angles resulted in more difficult judgments than between lengths of lines. Shah et al. (1999) established the notion that redesigning graphs can result in the improvement of the viewer's interpretation of the data. For example, the use of gestalt principles (Goldstein &

Cacciamani, 2021) such as proximity, similarity, and good continuation can help minimize the inferential processes and maximize the pattern association processes required to interpret relevant information. In section 1.8 we see how the hierarchy of accuracy in plot objects presented in W. Cleveland & McGill (1985) can explain biases in our interpretation and use of graphics.

During the 21<sup>st</sup> century, there have been advancements in the methodology used to investigate the effectiveness of statistical charts (Majumder, Hofmann, & Cook, 2013). Buja et al. (2009a) introduced the lineup protocol in which data plots are depicted and interpreted as statistics. Supported by the grammar of graphics, a data plot can be characterized as a statistic, defined as, “a functional mapping of a variable or set of variables” (Vanderplas et al., 2020). 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. The construction of data plots as statistics allows for easy experimentation, granting researchers the ability to compare the effectiveness of and understand the perception of different types of charts. While the lineup protocol differs from methodology used in earlier studies, the focus is still on initial perception 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). Lineups serve as a powerful tool for testing *perceived* differences by eliminating ambiguous questions. However, the lineup protocol is constrained by the inability to test higher order cognitive skills such as accurately reading information off of a graph or drawing conclusions from the graph, limiting their ability to be used for testing real-world applications.

## 1.6 Task Complexity

In order to understand how our visual system perceives statistical charts, we must first consider the complexity of the graphic and how viewers are interacting with the data and information being displayed (Tory & Moller, 2004). The efficiency in which a viewer extracts data and information from a graphical display is greatly affected by the complexity in the task environment. Cognitive fit refers to a match between the representation of the data and the complexity of the task; the representation and tools should support the task strategies, thus reducing the complexity of the task (Vessey, 1991). Carpenter & Shah (1998) identifies pattern recognition, interpretative processes, and integrative processes as strategies and processes required to complete tasks of varying degrees of complexity. Pattern recognition requires the viewer to encode graphic patterns while interpretive processes operate on those patterns to construct meaning. Integrative processes then relate the meanings to the contextual scenario as inferred from labels and titles. These processes are critical when determining cognitive fit since they provide the link between the graphical representation and task (Vessey, 1991). For example, perceptual differences may be identified through pattern recognition while estimation tasks would require integrative processes. Tory & Moller (2004) argues for multiple visual representations of the data since the users' information needs are dependent on both domain and task. Therefore, we must consider and determine how the viewer is perceiving and interacting with the graphic as this can influence their understanding of the data and information.

## 1.7 Graph Comprehension

Higher order cognitive processes require viewers to translate the visual features into conceptual relations by interpreting titles, labels, and scales. In order to understand how viewers are interpreting and using the data and information displayed on the chart, studies have asked participants to read information directly from a chart and provide a quantitative estimate or answer a predefined question (Amer, 2005; Broersma & Molenaar, 1985; Dunn, 1988; Peterson & Schramm, 1954; Spence, 1990; Tan, 1994). For instance, Amer (2005) demonstrates that visual illusion may bias decision making and graph comprehension, even if the graphs are constructed according to best practice. Participants were presented a cost volume profit graph (Fig. 1.14) with two crossing lines (revenue and cost) and asked to estimate three values: (1) the amount of total revenues on the ordinate corresponding to the endpoint of the total-revenue line plotted on the graph (2) the amount of total costs on the ordinate corresponding to the endpoint of the total-cost line plotted on the graph and (3) the amount of costs/revenues on the ordinate at the break even point—the point where the two lines cross. Results indicate that decision makers may consistently underestimate or overestimate the values displayed on line graphs due to what is called the “Poggendorff illusion.”

### 1.7.1 Questioning

An important consideration in understanding graph comprehension is the questions being asked of the viewer (Graesser, Swamer, Baggett, & Sell, 2014). Low level questions address the content and interpretation of explicit material while deeper questions require inference, application, and evaluation of the information being presented. Three levels of graph comprehension have emerged from

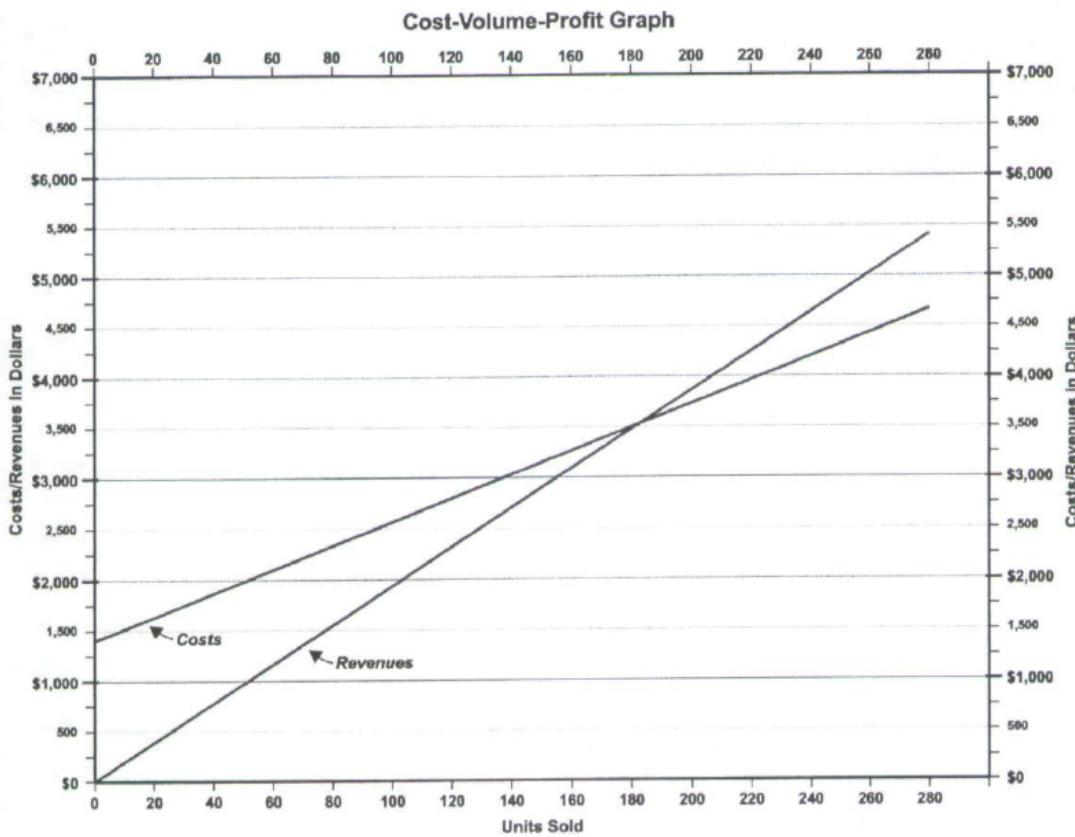


Figure 1.14: Participants in Amer (2005) were shown this plot in order to test their graph comprehension and identify visual biases. This figure illustrates the 'Poggendorff illusion' which results in visual biases of underestimation and overestimation.

mathematics education research (Curcio, 1987; Friel, Curcio, & Bright, 2001; Glazer, 2011; Jolliffe, 1991; R. Wood, 1968). The three behaviors related to graph comprehension involve (1) literal reading of the data (elementary level), (2) reading between the data (intermediate level), and (3) reading beyond the data (advanced level). Curcio (1987) aligns two multiple choice questions with each level of comprehension related to a graph showing the height of four children in centimeters (Fig. 1.15). Two literal items required the viewer to read the data, title, or axis label in order to answer, “What does this graph tell you?” or “How tall was xxx?” Comparison items required comparisons and the use of mathematical concepts to answer, “Who was the tallest?” and “How much taller was x than y?” Lastly,

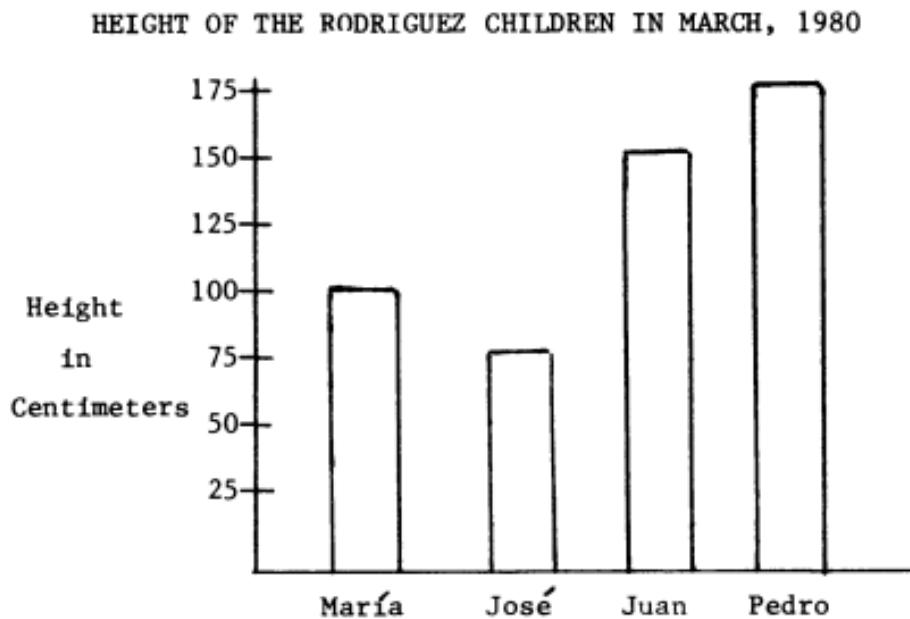


Figure 1.15: This plot was used in Curcio (1987) along with questions related to the heights of the four children in order to better understand graph comprehension skills.

extension items required an extension, prediction, or inference such as, “If  $x$  grows 5 centimeters and  $y$  grows 10 centimeters by Sept. 1981, who will be taller and by how much? In Friel et al. (2001), several studies were reviewed and their questions were placed in the taxonomy of skills required for answering questions at each level. In addition to the graph’s visual features and questioning, it is important for researchers to give careful consideration to the context of the graphic on the viewers comprehension.

### 1.7.2 Estimation Strategies

While not exclusive to extracting numerical values from charts, mathematics education research places an emphasis on quantitative estimation skills (Hogan & Brezinski, 2003). Three modes of estimation are taught as part of the mathematics curriculum in schools: numerosity, measurement, and computational estimation.

Numerosity estimation requires the estimation of the number of items in a group or array; for example, guessing the number of M&M's in a jar. Measurement estimation requires participants to provide an estimated value related to an object; for instance, an estimated length of a string or weight or a box. Computational estimation is the third mode which refers to estimated answers to computations as a way to avoid exact calculations. These estimates may be presented in either algorithmic form or a contextual scenario with words. A longer history of quantitative estimation can be found in psychometric literature in which estimation tasks appeared in early psychometric studies of mental abilities.

In efforts to develop estimation skills, research has been conducted to evaluate strategies for estimating tasks. Common strategies related to measurement estimation, involve reference point estimation, benchmark estimation, unit iteration, and guess and check. Joram, Gabriele, Bertheau, Gelman, & Subrahmanyam (2005) was interested in the relationship among strategy use and accuracy of students' representations of standard measurement units and measurement accuracy. In this study, students were asked to estimate the lengths of two objects and explain their process. The researchers used talk aloud protocols to prompt students to communicate their estimation strategies to an interviewer; results found that students who used a reference point had a more accurate representation of standard units and estimates of length than students who did not use a reference point. Gail Jones, Gardner, Taylor, Forrester, & Andre (2012) examined the effect of scale (metric verses English) and task context on the accuracy of measurement estimation for linear distances. The study showed that students were less accurate in estimating metric units as compared to English units and that estimation accuracy was highly dependent on task context. Forrester, Latham, & Shire (1990) argues

that estimation, approximating, and measuring are key components in the intuitive understanding of dimension and scale necessary to manipulate information and interact effectively with our environment. Without open ended conversations, our research demonstrates how the use of graphical tasks of varying complexities conducted through an online system can provide insight about the tactics and procedures used to extract meaning from a chart.

### 1.7.3 Estimation Biases

Certain biases including anchoring and rounding to multiples of five or ten arise in open-ended estimation tasks. When it comes to understanding graphics, anchoring is prominent in both graphical representations and data extraction tasks (Tan & Benbasat, 1990). Anchoring bias refers to an individual using easily observed visual cues such as grid lines or “anchors” when extracting information such as the  $x$  or  $y$  value on a chart (Tan, 1994 ; Godlonton, Hernandez, & Murphy, 2018). In addition to  $x$ -value and  $y$ -value anchoring, entity anchoring refers to anchoring on group information within a data set. Rounding errors occur out of natural human preference to provide rounded figures even if a precise estimate is desired or requested (Myers, 1954). Schneeweiss, Komlos, & Ahmad (2010) outlines distortion in results as a consequence of rounding and suggests the use of corrections when conducting statistical regression analyses on data prone to rounding.

Scale and axis labels are other critical factors in estimation accuracy. Dunham & Osborne (1991) argue that if there is not proper attention given to the scale when using a line graph, there is a potential for issues when interpreting asymmetric scales and when choosing appropriate scales for the graphic. Beeby & Taylor (1973) found that when asked to read data from line graphs, viewers consistently misread

the  $y$ -axis scale; when alternate grid lines were labeled, the unlabeled grid lines were read as halves. This misrepresentation is highlighted for asymmetric scales where spatial distance does not necessarily equate to numerical or quantitative difference. The choice of scale can change the shape of a graph, thus creating a conceptual demand for the viewer when constructing a mental image of the graph (Leinhardt, Zaslavsky, & Stein, 1990).

## 1.8 Logarithmic Scales and Mapping

A major issue we encountered in the creation of COVID-19 plots was how to display data from a wide range of values. 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. Exponential curves are a common source of data in which smaller magnitudes are compressed into a smaller area; Fig. 1.16 presents an exponential curve displayed on both the linear and log scale illustrating the use of the log scale when displaying data which spans several magnitudes. Logarithms convert multiplicative relationships (for example, 1 & 10 displayed 10 units apart and 10 & 100 displayed 90 units apart) to additive relationships (for example, 1 & 10 and 10 & 100 both equally spaced along the axis), showing proportional relationships and linearizing power functions (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

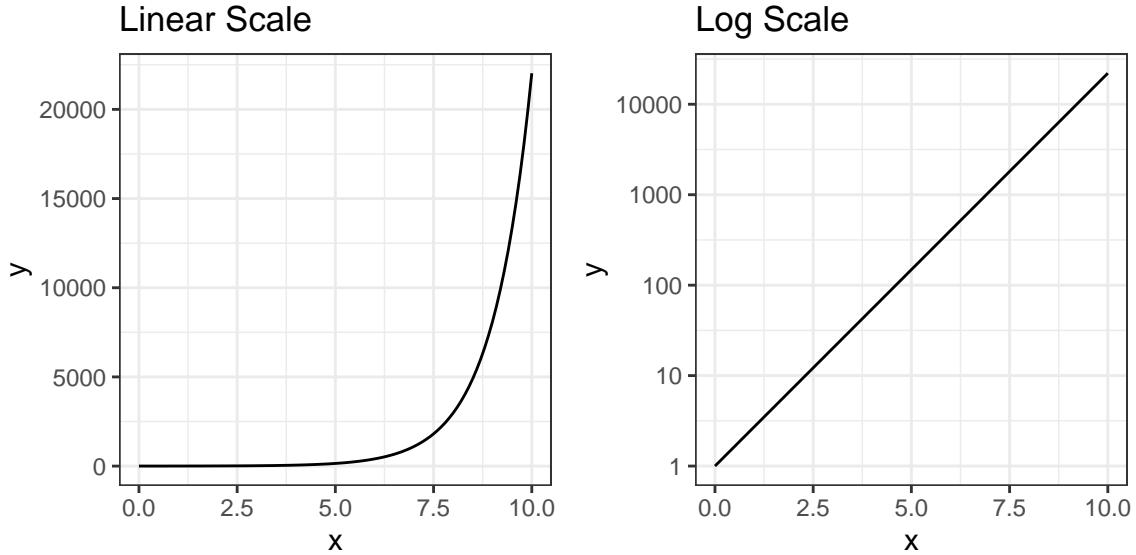


Figure 1.16: These plots present an exponential curve displayed on both the linear and log scale and illustrate the use of the log scale when displaying data which spans several magnitudes.

spaced along the axis) or log transformed scale labels (for example, 0, 1, and 2, showing the corresponding powers of 10).

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 (for example 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. The usefulness of log scales in comparing deaths attributed to COVID-19 between countries as of March 2020 is illustrated in Fig. 1.17; the diagonal reference lines provide a visual aid useful for interpretation (Burn-Murdoch et al., 2020). 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. In Fig. 1.18 and Fig. 1.19, the daily case counts as of June 30, 2020 are displayed on both the linear and log scales respectively (Burn-Murdoch et al., 2020). The effect of the linear scale (Fig. 1.18) appears to evoke a stronger reaction from the public than the log scale (Fig. 1.19) as daily case counts are clearly rising rapidly during the summer wave. 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(Fagen-Ulmschneider, 2020).

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). In Waddell (2005), comparisons were made between the linear and logarithmic scales for the relationship between dosage and carcinogenicity in rodents. Results favored the use of logarithmic scales for doses in order to put the relative doses into perspective whereas using a linear scale to administer doses to animals with the same chemicals to which humans are exposed does not provide useful, comparative information. Given the widespread use of logarithmic scales, it is important to understand the implications of their use in order to provide guidelines for best use.

When we first learn to count, we begin counting by ones (for example, 1, 2, 3, etc.), then by tens (for example, 10, 20, 30, etc.), and advancing to hundreds (for example, 100, 200, 300, etc.), following the base10 order of magnitude system (for example, 1, 10, 100, etc.). 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, &

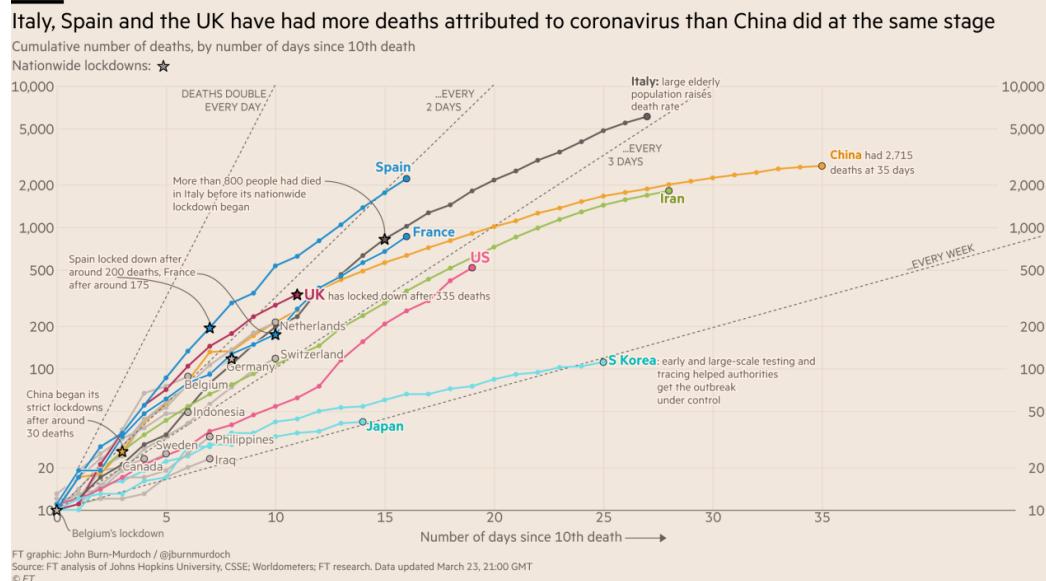


Figure 1.17: Covid-19 deaths (log scale) as of March 23, 2020.

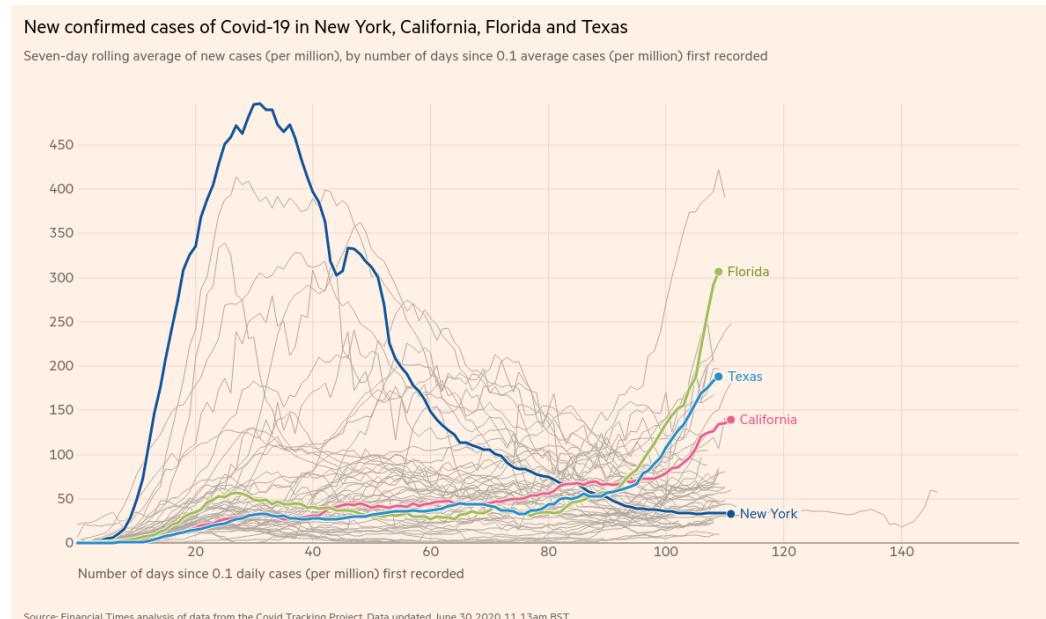


Figure 1.18: Covid-19 case counts (linear scale) as of June 30, 2020.

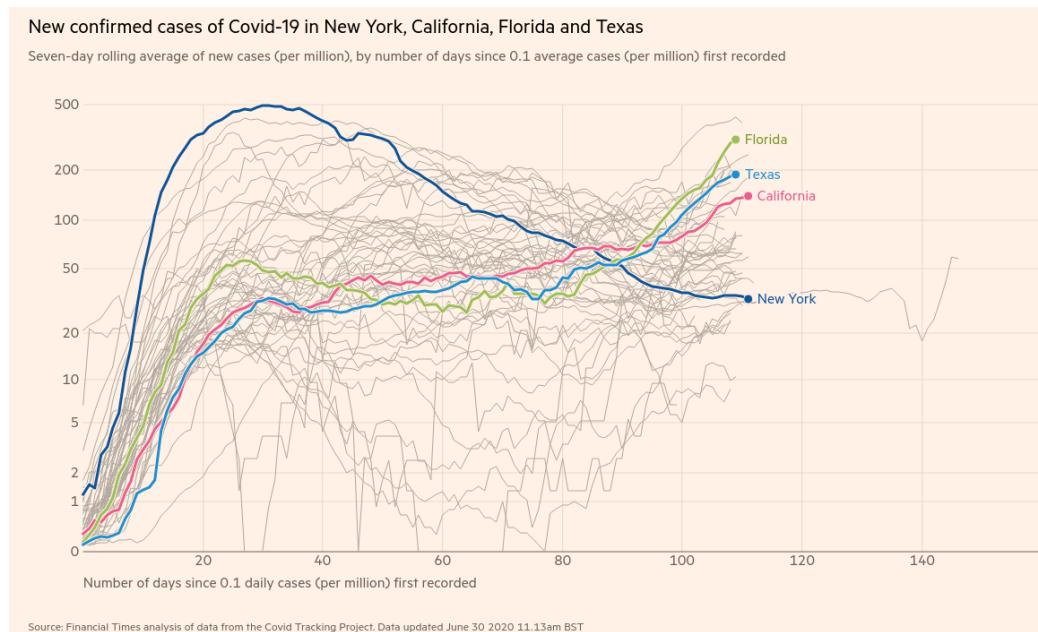


Figure 1.19: Covid-19 case counts (log scale) as of June 30, 2020.

Pica, 2008; Siegler & Braithwaite, 2017, 2017; Varshney & Sun, 2013). For example, a kindergartner asked to place numbers one through ten along a number line would place three close to the middle, following the logarithmic perspective (Varshney & Sun, 2013); Fig. 1.20 demonstrates how a kindergartner might map numbers along a number line. Dehaene et al. (2008) found that 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; for example, numbers can be mapped to a number line or numbers can be mapped onto a clock. There was a gradual transition from logarithmic to linear scale as the mapping of whole number magnitude representations transitioned from a compressed (approximately logarithmic) distribution to an approximately linear one. These results indicate the universal and cultural-dependent characteristics of the sense of number. Regardless of training, our visual system is still vulnerable to biases related to our perception of different stimuli such as weight, light, or sound. Known as *Weber's law*, it was

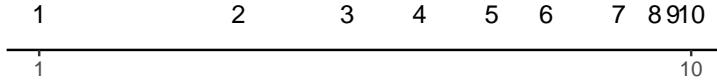


Figure 1.20: Kindergarten example of mapping numbers 1-10 along a number line.

established that we do not notice absolute changes in stimuli, but instead that we notice the relative change (Fechner, 1860). The *Weber-Fechner law* extended the discovery and states the relationship between the perceived intensity is logarithmic to the stimulus intensity when observed above a minimal threshold of perception.

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 logarithmic 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.

## 1.9 Underestimation of Exponential Growth

In addition to biases which result from the use of log scales, there is a general misinterpretation of exponential growth; Fig. 1.21 (Von Bergmann, 2021) illustrates how individuals in public health interpret exponential growth distinctly different

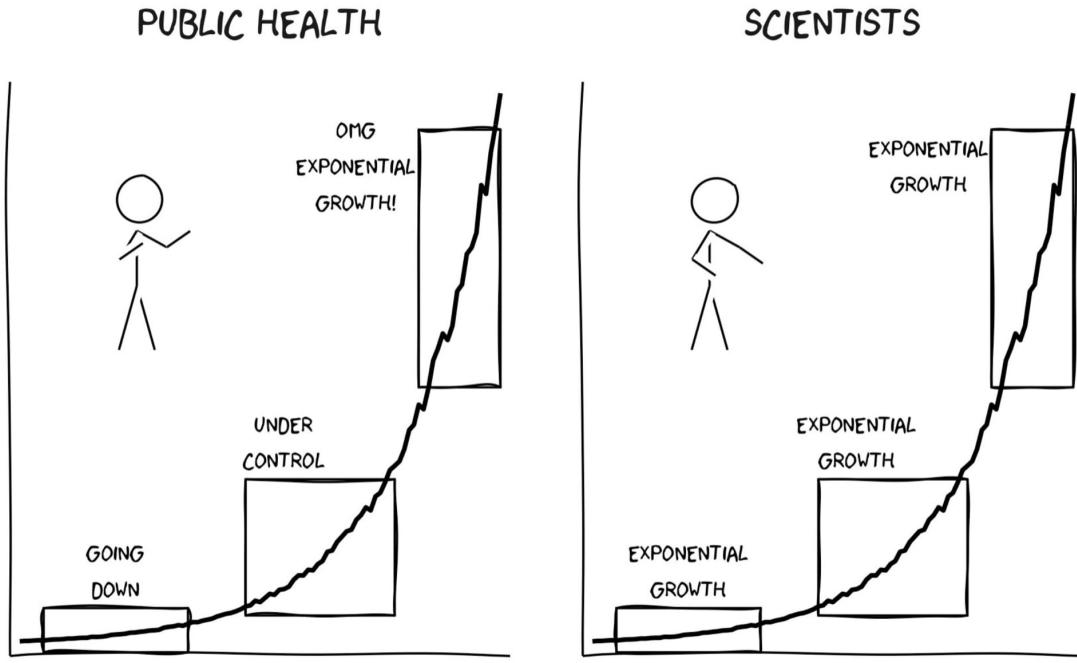


Figure 1.21: Comic illustrating the general misinterpretation of exponential growth.

from scientists during early, middle, and late stages of growth. Exponential growth is often misjudged in early stages, appearing to have a small growth rate. As exponential growth continues, the middle stage appears to be growing, but not at an astounding rate, appearing more quadratic. It is not until late stages of exponential growth when it is quite apparent that there is exponential growth occurring. This misinterpretation can lead to decisions made under inaccurate understanding causing future consequences.

Early studies explored the estimation and prediction of exponential growth and found that growth is underestimated when presented both numerically and graphically (Wagenaar & Sagaria, 1975). The hierarchy of plot objects found in W. Cleveland & McGill (1985) can provide a possible explanation for the underestimation that occurs in exponentially increasing trends; the exponential trend can be thought of as a series of tangential angles leading to less accurate

judgement of the next points. Results from Wagenaar & Sagaria (1975) indicated that numerical estimation is more accurate than graphical estimation for exponential curves. Experimental studies were conducted in order to determine strategies to improve the accuracy of estimation of exponential growth (Gregory Jones, 1977; MacKinnon & Wearing, 1991; Wagenaar & Sagaria, 1975). There was no improvement in estimation found when participants had contextual knowledge or experience with exponential growth, but instruction on exponential growth reduced the underestimation; participants adjusted their initial starting value but not their perception of the growth rate (Gregory Jones, 1977; Wagenaar & Sagaria, 1975). MacKinnon & Wearing (1991) found that estimation was improved by providing immediate feedback to participants about the accuracy of their current predictions.

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 three relationships displayed on linear-linear scales, log-log scales with untransformed values, or log–log scales with log transformed values (Fig. 1.22). 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 (which is an exponential relationship) in linear-linear space.

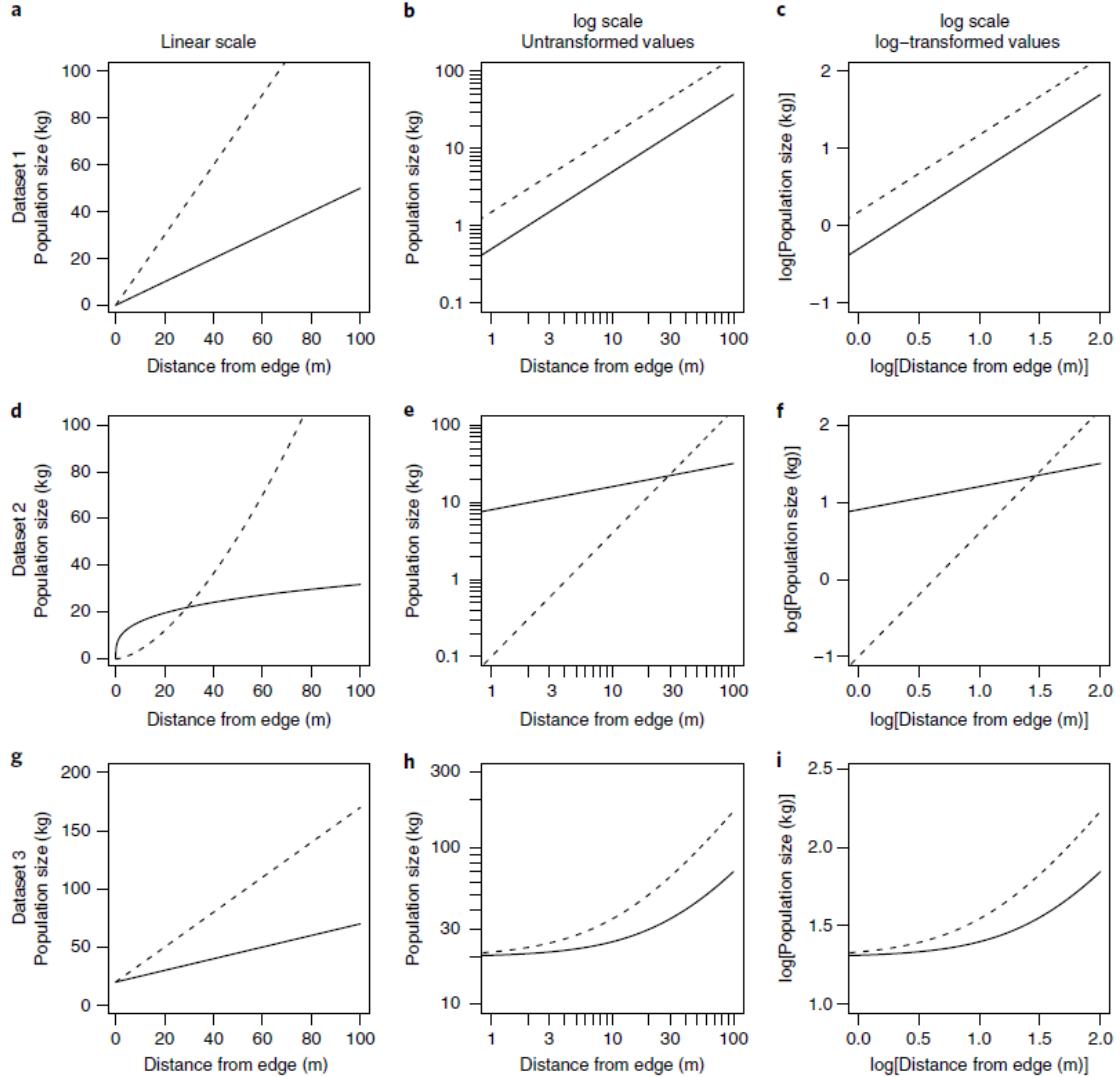


Figure 1.22: Graphs presented to participants in Menge (2018). Three relationships were displayed on the linear-linear scales, log-log scales with transformed values, or log-log scales with log transformed values. These figures demonstrate misconceptions participants encountered when presented data on the log-log scales.

The ‘hand-hold fallacy’ stems from the misconception that steeper slopes in log-log relationships are steeper slopes in linear-linear space, illustrated in Fig. 1.22 d-f. 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 (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, illustrated in Fig. 1.22 a-c and d-f. 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, illustrated in Fig. 1.22 d-f, (2) lines in log-log space curve upward in linear-linear space, illustrated in Fig. 1.22 d-f, and (3) curves in log-log space have the same curvature in linear-linear space, illustrated in Fig. 1.22 g-i. 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, and all exponential relationships curve upward; this is only true when the log-log slope is greater than one. 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.

## 1.10 Research Objectives

In this research, we conducted a series of three graphical studies to evaluate the impact displaying data on the log scale has on human perception of exponentially increasing trends compared to displaying data on the linear scale. Each study was

related to a different graphical task, each requiring a different level of interaction and cognitive use of the data being presented. The first experiment evaluated 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 lineups and asked to identify the plot that differed most from the surrounding plots. 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? To test an individual's ability to make predictions for exponentially increasing data, participants were asked to draw a line using their computer mouse through an exponentially increasing trend shown on both the linear and log scale. In addition to differentiation and prediction of exponentially increasing data, an estimation 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. Combined, the three studies provide a comprehensive evaluation of the impact of displaying exponentially increasing data on a log scale as it relates to perception, prediction, and estimation. The results of these studies help us to make recommendations and provide guidelines for the use of log scales.

## 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. Identifying differences 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. The study in this chapter is conducted through visual inference and the use of statistical lineups (Buja et al., 2009a) to differentiate between exponentially increasing curves with differing levels of curvature, using linear and log scales.

#### 2.2 Visual Inference

In section 1.4, we explained how a data plot can be evaluated and treated as a visual statistic, a numerical function which summarizes the data. To evaluate a graph, the statistic (data plot) must be run through a visual evaluation - a person. We can conclude the visual statistics are significantly different if two different methods of presenting data result in qualitatively different results when evaluated

visually. Recent graphical experiments have utilized statistical lineups to quantify the perception of graphical design choices (Hofmann, Follett, Majumder, & Cook, 2012; Loy, Follett, & Hofmann, 2016; Loy, Hofmann, & Cook, 2017; VanderPlas & Hofmann, 2017). Statistical lineups provide an elegant way of combining perception and statistical hypothesis testing using graphical experiments (Majumder et al., 2013; Vanderplas et al., 2020; Wickham, Cook, Hofmann, & Buja, 2010). ‘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 panel containing the real data from among a set of decoy null plots. Null plots display data under the assumption there is no relationship and can be generated by permutation or simulation. A statistical lineup typically consists of 20 panels - one target panel and 19 null panels. If the viewer can identify the target panel randomly embedded within the set of null panels, this suggests that the real data is visually distinct from data generated under the null model. Fig. 2.1 provides examples of statistical lineups. The lineup plot on the left displays increasing exponential data displayed on a linear scale with panel 13 as the target; the lineup plot on the right displays increasing exponential data on the log base ten scale with panel 4 as the target.

While explicit graphical tests direct the participant to a specific feature of a plot to answer a specific question, implicit graphical tests require the user to identify both the purpose and function of the plot in order to evaluate the plots shown (Vanderplas et al., 2020). Implicit graphical tests, such as lineups, have the advantage of simultaneously visually testing for multiple visual features including outliers, clusters, linear and nonlinear relationships. Responses from multiple

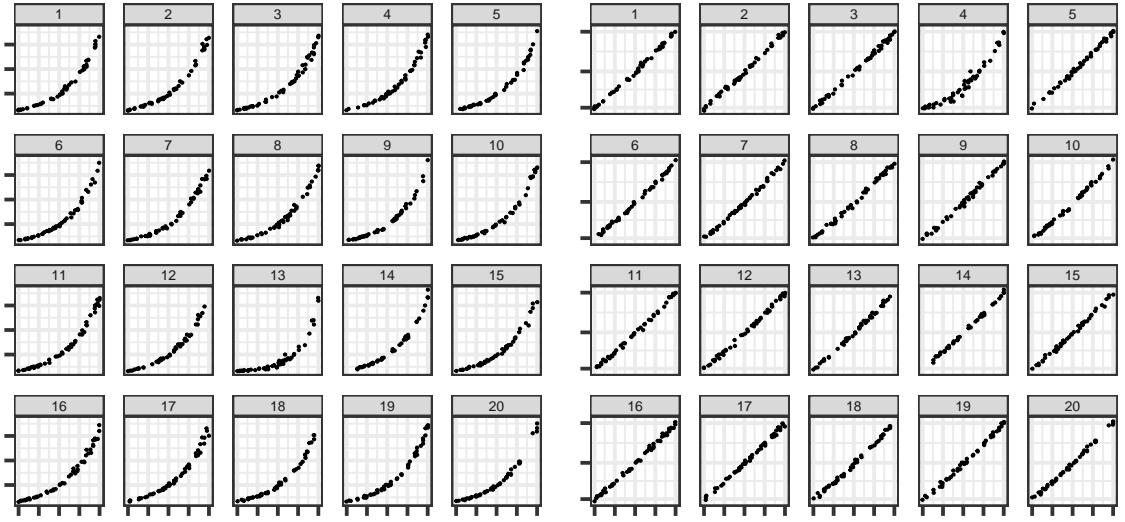


Figure 2.1: The lineup plot on the left displays increasing exponential data on a linear scale with panel  $(2 \times 5) + 3$  as the target. The lineup plot on the right displays increasing exponential data on the log scale with panel  $2 \times 2$  as the target.

viewers are collected through convenience sampling (in informal situations) or crowd sourcing websites such as Prolific, Amazon Mechanical Turk, and Reddit (in more formal situations).

### 2.3 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 begin 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 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.

Data were simulated 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  $\alpha$  and  $\theta$  are adjusted based on  $\beta$  and  $\sigma^2$  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 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  $y = -x$  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 regression 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 base `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 -  $\hat{\alpha}, \hat{\beta}, \hat{\theta}$ .
- 

## 2.4 Parameter Selection

We followed a ‘Goldilocks’ inspired procedure to choose three levels of trend curvature (low curvature, medium curvature, and high curvature). For each curvature level, we simulated 1000 data sets of  $(x_{ij}, y_{ij})$  points for  $i = 1, \dots, 50$   $x$ -values and  $j = 1 \dots 10$  corresponding  $y$ -values per  $x$ -value. Each generated  $x_i$  point from Algorithm 2 was replicated ten times. On each of the individual data sets, we

---

**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  $\sigma$  standard deviation from the exponential curve.
  - **Output Parameters:**  $N$  points, in the form of vectors  $\mathbf{x}$  and  $\mathbf{y}$ .
- 1: Generate  $\tilde{x}_j, j = 1, \dots, \frac{3}{4}N$  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  $\tilde{x}_i, i = 1, \dots, N$  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  $\tilde{\alpha} = \frac{\hat{\alpha}}{e^{\hat{\theta}/2}}$ . 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 2.1: Lineup data simulation final parameters

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

conducted a linear regression model and computed the lack of fit statistic (LOF) which measures the deviation of the data from the linear regression model. The density curves of the LOF statistics for each level of curvature are plotted (Fig. 2.2) 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. Final parameters used for data simulation are shown in Table 2.1.

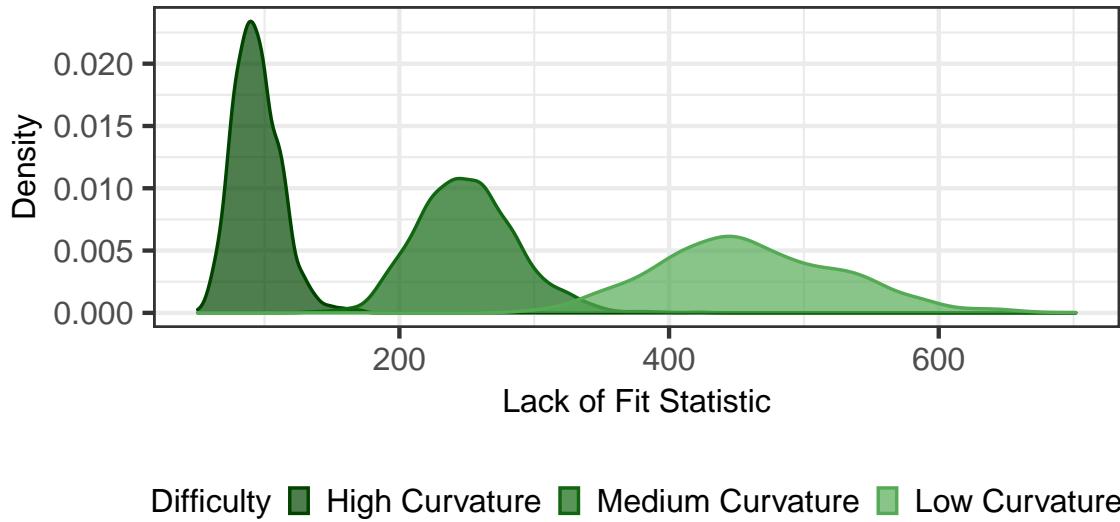


Figure 2.2: Density plot of the lack of fit statistic showing separation of difficulty levels: obvious curvature, noticeable curvature, and almost linear.

## 2.5 Lineup Setup

Lineup plots were generated by mapping one simulated data set corresponding to curvature level A to a scatter plot to be identified as the target panel while multiple simulated data sets corresponding to curvature level B were individually mapped to scatter plots for the null panels. The `nullabor` package in R (Buja et al., 2009b) was used to randomly assign the target plot to one of the panels surrounded by panels containing null plots. For example, a target plot with simulated data following an increasing exponential curve with high curvature is randomly embedded within null plots with simulated data following an increasing exponential trend with low curvature. By the implemented constraints, the target panel and null panels will span a similar domain and range. There are a total of six lineup curvature combinations; Fig. 2.3 illustrates the six lineup curvature combinations (top: linear scale; bottom: log scale) where the green line indicates the curvature level designated to the target plot while the black line indicates the curvature level

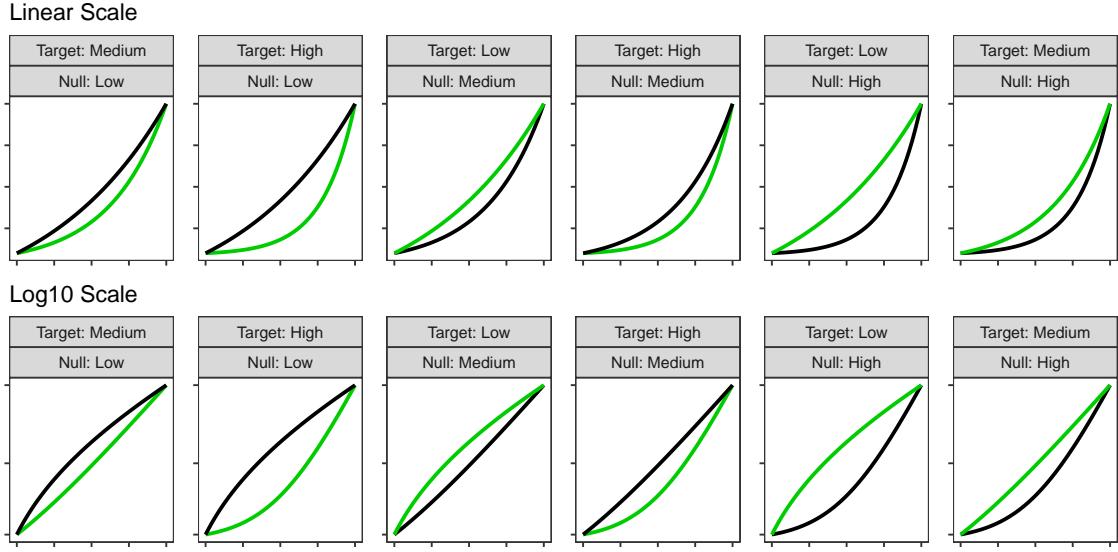


Figure 2.3: Thumbnail plots illustrating the six curvature combinations displayed on both scales (linear and log). The green line indicates the curvature level to be identified as the target plot from amongst a set of null plots with the curvature level indicated by the black line.

assigned to the null plots. Two sets of each lineup curvature combination were simulated (total of twelve 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 curvature 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 chapter and their analysis is left to a later date.

## 2.6 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 curvature 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 in their region were recruited from Prolific, a survey site that connects researchers to study participants. Participants were compensated for their time and participated in all three related graphical studies consecutively. The lineup study in this chapter was completed first in the series of graphical studies. Previous literature suggests that prior mathematical knowledge or experience with exponential data is not associated with the outcome of graphical experiments (VanderPlas & Hofmann, 2015). Participants completed the series of graphical tests using a R Shiny application found [here](#).

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 graphical 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.

## 2.7 Results

Participant recruitment and study deployment was conducted via Prolific, a crowd sourcing website, on Wednesday, March 23, 2022 during which 325 individuals completed 4492 unique test lineup evaluations. Only participants who completed the lineup study were included in the final data set which included a total of 311 participants and 3958 lineup evaluations. Each plot was evaluated between 141 and 203 times (Mean: 164.92, SD: 14.9). Participants correctly identified the target panel in 47% of the 1981 lineup evaluations made on the linear scale and 65.3% of the 1977 lineup evaluations made on the log scale.

Target plot identification was analyzed using the `glmer` function in the `lme4` R package (Bates, Mächler, Bolker, & Walker, 2014). Estimates and odds ratio comparisons between the log and linear scales were calculated using the `emmeans` R package (Lenth, 2021). Each lineup plot evaluated was assigned a binary value based on the participant response (correct = 1, not correct = 0). Define  $Y_{ijkl}$  to be the event that participant  $l = 1, \dots, N_{\text{participant}}$  correctly identifies 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 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

$$\text{logit } P(Y_{ijk}) = \eta + \delta_i + \gamma_j + \delta\gamma_{ij} + s_l + d_k \quad (2.2)$$

where

- $\eta$  is the baseline average probability of selecting the target plot

- $\delta_i$  is the effect of scale  $i = 1, 2$
- $\gamma_j$  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
- $s_l \sim N(0, \sigma_{\text{participant}}^2)$  is the random effect for participant characteristics
- $d_k \sim N(0, \sigma_{\text{data}}^2)$  is the random effect for data specific characteristics.

We assume that random effects for data set and participant are independent.

Results indicate a strong interaction between the curvature combination and scale ( $\chi_5^2 = 294.443$ ;  $p < 0.0001$ ). Variance due to participant and data set were estimated to be  $\hat{\sigma}_{\text{participant}}^2 = 1.19$  (s.e. 1.09) and  $\hat{\sigma}_{\text{data}}^2 = 0.433$  (s.e. 0.66) respectively.

On both the log and linear scales, the highest accuracy occurred in lineup plots where the target model and null model had a large curvature difference and the target plot had more curvature than the null plots (high curvature target plot embedded in low curvature null plots). There is 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). L. 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 (something has less curvature), but easy to say that it is not linear; our results concur with this observation. Fig. 2.4 displays the estimated (log) odds ratio of successfully identifying the target panel on the log scale compared to the linear scale. The

thumbnail figures to the right of the plot illustrate the curvature combination on both the linear (left thumbnail) and log base ten (right thumbnail) scales associated with the  $y$ -axis label. The choice of scale has no impact if curvature differences are large 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 large differences in curvature when the target plot has less curvature than the null plots (low curvature target plot embedded in high curvature null plots). An exception occurs 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 the log scale. When examining this curvature combination, the same perceptual effect occurs as what we previously saw, but in a different context of scales. On the linear scale, participants are perceptually identifying a curved trend from close to a linear trend whereas after the logarithmic transformation, participants are perceptually identifying a trend close to linear from a curved trend. This again supports 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 (L. Best et al., 2007).

## 2.8 Discussion and Conclusion

The overall goal of this chapter is to provide basic research to support the principles used to guide design decisions in scientific visualizations of exponential data. In this study, we explored the use of linear and log scales to determine

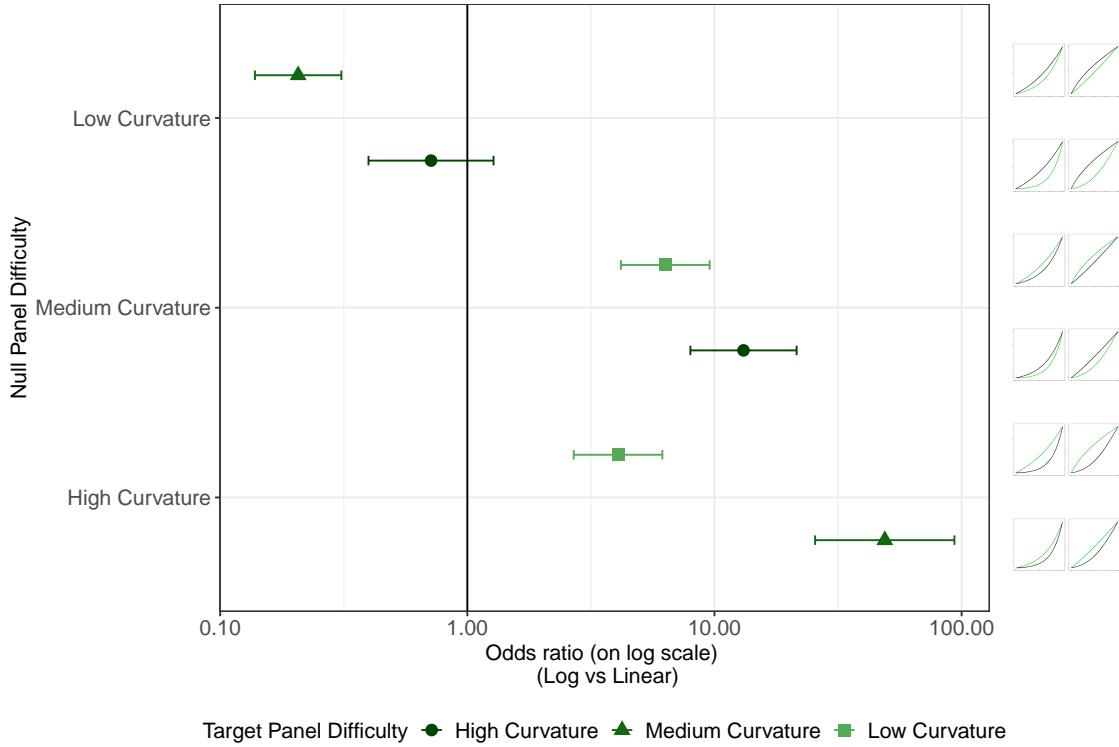


Figure 2.4: Estimated (log) odds ratio of successfully identifying the target panel on the log scale compared to the linear scale. The y-axis indicates the the model parameters used to simulate the null plots with the target plot model parameter selection designated by shape and shade of green. The thumbnail figures on the right display the curvature combination as shown in Fig. 2.3 on both scales (linear - left, log - right).

whether our ability to notice differences in exponentially increasing trends is impacted by the choice of scale. The results indicated that when there was a large 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 large 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 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

In [Chapter 2](#), a base foundation for future exploration of the use of log scales was established by evaluating participants ability to identify differences in charts through the use of lineups. This did not require that participants were able to understand exponential growth, identify log scales, or have any mathematical training; instead, it simply tested whether individuals are able to perceptually distinguish different curvature and slopes in a standard scatterplot. This is necessary, but not sufficient, to determine whether individuals are capable of higher-level interaction with statistical data on log and linear scales. In order to determine whether there are cognitive disadvantages to log scales, we utilized interactive graphics to test an individual’s ability to make predictions for exponentially increasing data. In this study, participants were asked to draw a line using their computer mouse through an exponentially increasing trend shown on both the log and linear scales.

##### 3.1.1 A Review of Regression and Prediction

Our visual system is naturally built to look for structure and identify patterns. For instance, points going down from left to right indicates a negative correlation

between the  $x$  and  $y$  variables. In the past, manual methods have been used to compare our intuitive visual sense of patterns to those determined by statistical methods. Initial studies in the 20<sup>th</sup> century explored the use of fitting lines by eye through a set of points (Finney, 1951; Mosteller, Siegel, Trapido, & Youtz, 1981). Common methods of fitting trends by eye involve maneuvering a string, black thread, or ruler until the fit is suitable, then drawing the line through the set of points.

Researchers in Finney (1951) were interested in assessing the effect of stopping iterative maximum likelihood calculations after one iteration. Many techniques in statistical analysis are performed with the aid of iterative calculations such as Newton's method or Fisher's scoring. Guesses are made at the best estimates of certain parameters and these guesses are then used as the basis of a computation which yields a new set of approximation to the parameter estimates; this same procedure is then performed on the new parameter estimates and the computing cycle is repeated until convergence, as determined by the statistician, is reached. The author was interested in whether one iteration of calculations was sufficient in the estimation of parameters connected with dose-response relationships. One measure of interest in dose-response relationships is the relative potency between a test preparation of doses and standard preparation of does; relative potency is calculated as the ratio of two equally effective doses between the two preparation methods. Fig. 3.1 shows a pair of parallel probit responses in a biological assay. The  $x$ -axis is the  $\log_{1.5}$  dose level for four dose levels (for example, doses 4, 6, 9, and 13 correspond correspond to equally spaced values on a logarithmic scale, labeled 0, 1, 2, and 3) and the  $y$ -axis is the corresponding probit response as calculated in Finney & Stevens (1948); circles correspond to the test preparation method while the

crosses correspond to the standard preparation method. For these sort of assays, the does-response relationship follows a linear regression of the probit response on the logarithm of the dose levels; the two preparation methods can be constrained to be parallel (Jerne & Wood, 1949), limiting the relative potency to one consistent value. In this study, twenty-one scientists were recruited via postal mail and asked to “rule two lines” in order to judge by eye the positions for a pair of parallel probit regression lines in a biological assay (Fig. 3.1). The author then computed one iterative calculation of the relative potency based on starting values as indicated by the pair of lines provided by each participant and compared these relative potency estimates to that which was estimated by the full probit technique (reaching convergence through multiple iterations). Results indicated that one cycle of iterations for calculating the relative potency was sufficient based on the starting values provided by eye from the participants.

Thirty years later, Mosteller et al. (1981), sought to understand the properties of least squares and other computed lines by establishing one systematic method of fitting lines by eye. The authors recruited 153 graduate students and post doctoral researchers in Introductory Biostatistics. Participants were asked to fit lines by eye to four sets of points (Fig. 3.2) using an 8.5 x 11 inch transparency with a straight line etched completely across the middle. A latin square design (Anderson & McLean, 1974) with packets of the set of points stapled together in four different sequences was used to determine if there is an effect of order of presentation; results indicated that order of presentation had no effect. Without a formal analysis of the study, the researchers discussed the idea that participants tended to fit the slope of the first principal component (error minimized orthogonally, both horizontal and vertical, to the regression line) over the slope of the least squares regression line

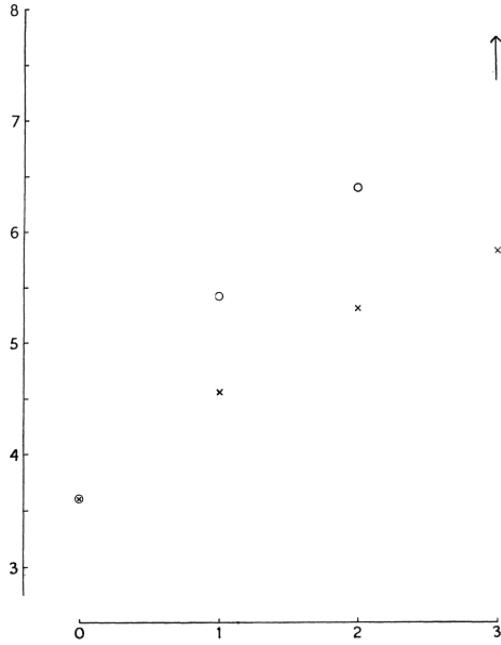
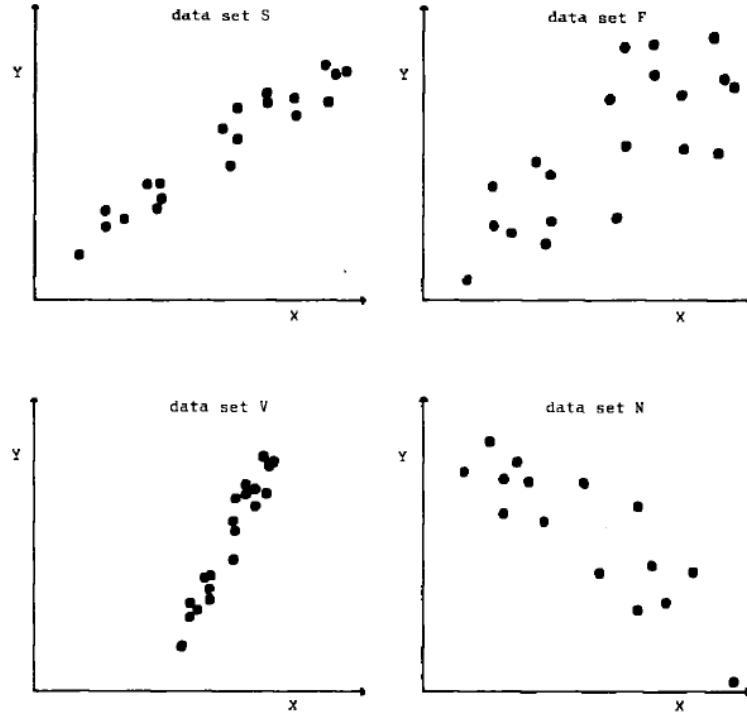


Figure 3.1: Parallel probit responses in a biological assay shown to study participants in Subjective Judgement in Statistical Analysis (1951). The  $x$ -axis is the  $\log_{1.5}$  dose level and the  $y$ -axis is the corresponding probit response; circles correspond to the test preparation method while the crosses correspond to the standard preparation method.

(error minimized vertically to the regression line) (Fig. 3.7).

Recently, Cicciione & Dehaene (2021) conducted a comprehensive set of studies investigating human ability to detect trends in graphical representations from a psychophysical approach. Participants were asked to judge trends, estimate slopes, and conduct extrapolation. To estimate slopes, participants were asked to report the slope of the best-fitting regression line using a track-pad to adjust the tilt of a line on the screen. Results indicated the slopes participants reported were always in excess of the ideal slopes, both in the positive and in the negative direction, and those biases increase with noise and with number of points. This supports the results found in Mosteller et al. (1981) and suggest that participants might use Deming regression (Deming, 1943), which is equivalent to a regression equation



**Figure 1. The Data Sets of S, F, V, and N**

Figure 3.2: Scatter plots of the data shown to study participants in Eye Fitting Straight Lines (1981).

based on the first principal component or principal axes and minimizes the Euclidean distance of points from the line, when fitting a line to a noisy scatter-plot.

While not explicitly intended for perceptual testing, in 2015, the New York Times introduced an interactive feature, called ‘You Draw It’ (Aisch, Cox, & Quealy, 2015; Buchanan, Park, & Pearce, 2017; Katz, 2017), where readers input their own assumptions about various metrics and compare how these assumptions relate to reality. The New York Times team utilizes Data Driven Documents (D3) that allow readers to predict these metrics through the use of drawing a line on their computer screen with their computer mouse. Fig. 3.3 (Katz, 2017) is one such example in which readers were asked to draw the line for the missing years

Since 1990, the number of Americans who have died every year from **car accidents**...

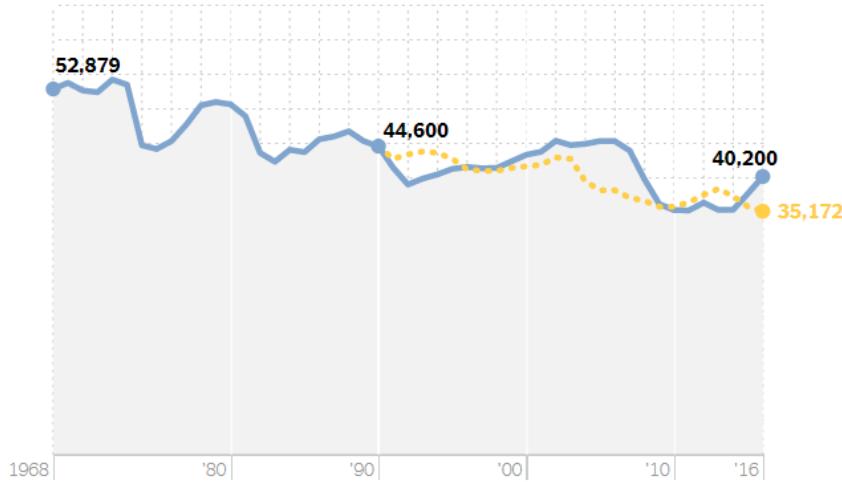


Figure 3.3: New York Times 'You Draw It' feature; readers were asked to use their mouse to draw the line for the missing years in order to provide what they estimated to be the number of Americans who have died every year from car accidents, since 1990.

providing what they estimated to be the number of Americans who have died every year from car accidents, since 1990. After the reader completed drawing the line, the actual observed values were revealed and the reader was able to check their estimated knowledge against the actual reported data.

### 3.1.2 Data Driven Documents

Major news and research organizations such as the New York Times, FiveThirtyEight, Washington Post, and the Pew Research Center create and customize graphics with Data Driven Documents (D3). In June 2020, the New York Times released a front page displaying figures that represent each of the 100,000 lives lost from the COVID-19 pandemic until that point in time (Barry et al., 2020); this visualization was meant to bring about a visceral reaction and resonate with

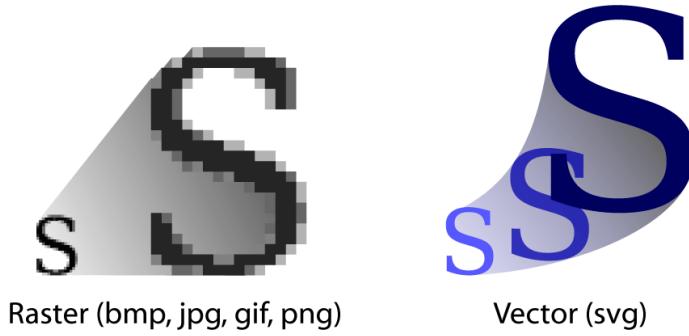


Figure 3.4: SVG vs raster

readers. During 2021 March Madness, FiveThirtyEight created a roster-shuffling machine which allowed readers to build their own NBA contender through interactivity (R. Best & Boice, 2021). Data Driven Documents (D3) is an open-source JavaScript based graphing framework created by Mike Bostock during his time working on graphics at the New York Times. The grammar of D3 includes elements such as circles, paths, and rectangles with choices of attributes and styles such as color and size. Data Driven Documents depend on Extensible Markup Language (XML) to generate graphics and images by binding objects and layers to the plotting area as Scalable Vector Graphics (SVG) in order to preserve the shapes rather than the pixels (Fig. 3.4) (Tol, 2021). Advantages of using D3 include animation and allowing for movement and user interaction such as hovering, clicking, and brushing.

A challenge of working with D3 is the environment necessary to display the graphics and images. The `r2d3` package in R provides an efficient integration of D3 visuals and R by displaying them in familiar HTML output formats such as RMarkdown or Shiny applications (Luraschi & Allaire, 2018). The creator of the graphic applies `D3.js` source code to visualize data which has previously been processed within an R setting.

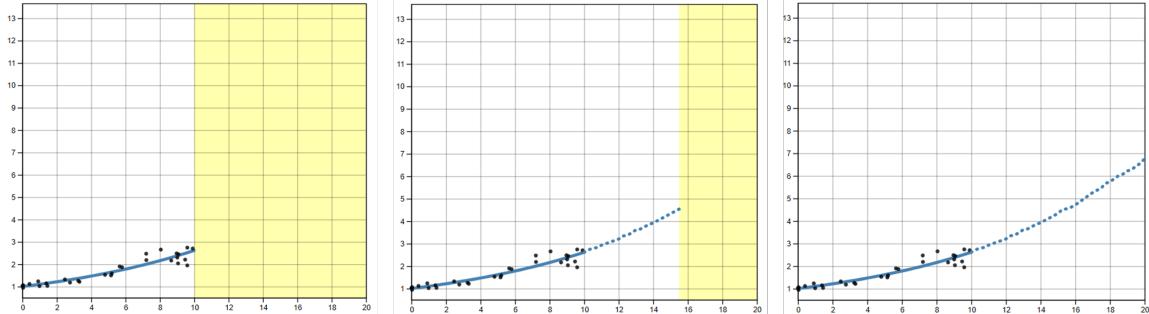


Figure 3.5: Example of a ‘You Draw It’ interactive plot as shown to participants during the study. The first frame shows what the participant saw along with the prompt, *‘Use your mouse to fill in the trend in the yellow box region’*. Next, the yellow box region moved along as the participant drew their trend-line until the yellow region disappeared, indicating the participant had filled in the entire domain.

The example R code illustrates the structure of the `r2d3` function which includes specification of a data frame in R (converted to a JSON file), the D3.js source code file, and the D3 version that accompanies the source code. A default SVG container for layering elements is then generated by the `r2d3` function which renders the plot using the source code. Appendix A outlines the development of the ‘You Draw It’ interactive plots used in this study through the use of `r2d3` and R shiny applications. Fig. 3.5 provides an example of a ‘You Draw It’ interactive plot as was shown to participants during the study. The first frame shows what the participant saw along with the prompt, “Use your mouse to fill in the trend in the yellow box region”. Next, the yellow box region moved along as the participant drew their trend-line until the yellow region disappeared, indicating the participant had filled in the entire domain.

```
r2d3(data = data, script = "d3-source-code.js",
      d3_version = "5")
```

### 3.2 Study Design

This chapter contains two sub-studies; the first aims to establish ‘You Draw It’ as a tool for measuring predictions of trends fitted by eye and a method for testing graphics, the second then applies ‘You Draw It’ to test an individual’s ability to make predictions for exponentially increasing data on the log and linear scale. The first sub-study, referred to as Eye Fitting Straight Lines in the Modern Era, was intended to implement the ‘You Draw It’ feature as a way to measure the patterns we see in data. We validate the ‘You Draw It’ method for testing graphics by replicating the less technological study conducted by Mosteller et al. (1981). Based on previous research, we hypothesize that visual regression tends to mimic principle component or Deming regression rather than an ordinary least squares regression. In order to assess this hypothesis, we introduce a method for statistically modeling the participant drawn lines using generalized additive mixed models (GAMM). The second sub-study, referred to as Prediction of Exponential Trends, uses the established ‘You Draw It’ method to test an individual’s ability to make predictions for exponentially increasing data on both the log and linear scales. We then use the GAMMS to analyze participant drawn lines; a benefit of using a GAMM is the estimation of smoothing splines, allowing for flexibility in the residual trend and analysis of nonlinear trends.

A total of six data sets - four Eye Fitting Straight Lines in the Modern Era and two Prediction of Exponential Trends - are generated for each individual at the start of the experiment. The two simulated data sets corresponding to the simulated data models used in the Prediction of Exponential Trends sub-study are then plotted a total of four times each with different aesthetic and scale choices for a total of eight task plots. Participants in the study are first shown two ‘You Draw It’ practice

plots followed by twelve ‘You Draw It’ task plots. The order of all twelve task plots was randomly assigned for each individual in a completely randomized design where users saw the four task plots from the Eye Fitting Straight Lines in the Modern Era sub-study interspersed with the eight task plots from the Prediction of Exponential Trends sub-study.

The ‘You Draw It’ study in this chapter was completed second in the series of the three graphical studies and took about fifteen minutes for participants to complete drawn trend lines for the twelve ‘You Draw It’ task plots. Participants completed the series of graphical tests using a R Shiny application found [here](#). Participant recruitment and study deployment was conducted via Prolific, a crowd sourcing website, on Wednesday, March 23, 2022 during which a total of 302 individuals completed 1254 unique ‘You Draw It’ task plots for the first sub-study and 309 individuals completed 2520 unique ‘You Draw It’ task plots associated with the second sub-study.

### 3.3 Eye Fitting Straight Lines in the Modern Era

Finney (1951) and Mosteller et al. (1981) use methods such as using a ruler, string, or transparency sheet to fit straight lines through a set of points. This section replicates the study found in Mosteller et al. (1981) and extends this study with formal statistical analysis methods in order to establish ‘You Draw It’ as a tool and method for testing graphics.

#### 3.3.1 Data Generation

All data processing was conducted in R before being passed to the D3.js source code. A total of  $N = 30$  points  $(x_i, y_i), i = 1, \dots, N$  were generated for

Table 3.1: Eye Fitting Straight Lines in the Modern Era simulation model parameters

Parameter Choice	$y_{\bar{x}}$	$\beta_1$	$\sigma$
S	3.88	0.66	1.30
F	3.90	0.66	1.98
V	3.89	1.98	1.50
N	4.11	-0.70	2.50

$x_i \in [x_{min}, x_{max}]$  where  $x$  and  $y$  have a linear relationship. Data were simulated based on linear model with additive errors:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad (3.1)$$

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

The parameters  $\beta_0$  and  $\beta_1$  were selected to replicate Mosteller et al. (1981) with  $e_i$  generated by rejection sampling in order to guarantee the points shown align with that of the fitted line. An ordinary least squares regression was then fit to the simulated points in order to obtain the best fit line and fitted values in 0.25 increments across the domain,  $(x_k, \hat{y}_{k,OLS}), k = 1, \dots, 4x_{max} + 1$ . The data simulation function then outputted a list of point data and line data both indicating the parameter identification,  $x$  value, and corresponding simulated or fitted  $y$  value.

The data simulation procedure is described in Algorithm 3.

Simulated model equation parameters were selected to reflect the four data sets (F, N, S, and V) used in Mosteller et al. (1981) (Table 3.1). Parameter choices F, N, and S simulated data across a domain of 0 to 20. Parameter choice F produced a trend with a positive slope and a large variance while N had a negative slope and a large variance. In comparison, S resulted a trend with a positive slope with a small variance and V yielded a steep positive slope with a small variance over the domain

---

**Algorithm 3** Eye Fitting Straight Lines in the Modern Era Data Simulation
 

---

- **Input Parameters:**  $y_{\bar{x}}$  for calculating the y-intercept,  $\beta_0$ ; slope  $\beta_1$ ; standard deviation from line  $\sigma$ ; sample size of points  $N = 30$ ; domain  $x_{min}$  and  $x_{max}$ ; fitted value increment  $x_{by} = 0.25$ .
  - **Output Parameters:** List of point data and line data each indicating the parameter identification,  $x$  value, and corresponding simulated or fitted  $y$  value.
- 1: Randomly select and jitter  $N = 30$   $x$  values along the domain,  $x_{i=1:N} \in [x_{min}, x_{max}]$ .
  - 2: Determine the  $y$ -intercept,  $\beta_0$ , at  $x = 0$  from the provided slope ( $\beta_1$ ) and  $y$  value at the mean of  $x$  ( $y_{\bar{x}}$ ) using point-slope equation of a line.
  - 3: Generate “good” errors,  $e_{i=1:N}$  based on  $N(0, \sigma)$  by setting a constraint requiring the mean of the first  $\frac{1}{3}N$  errors  $< |2\sigma|$ .
  - 4: Simulate point data based on  $y_i = \beta_0 + \beta_1 x_i + e_i$
  - 5: Obtain ordinary least squares regression coefficients,  $\hat{\beta}_0$  and  $\hat{\beta}_1$ , for the simulated point data using the `lm` function in the `stats` package in base R.
  - 6: Obtain fitted values every 0.25 increment across the domain from the ordinary least squares regression  $\hat{y}_{k,OLS} = \hat{\beta}_{0,OLS} + \hat{\beta}_{1,OLS} x_k$ .
  - 7: Output data list of point data and line data each indicating the parameter identification,  $x$  value, and corresponding simulated or fitted  $y$  value.
- 

of 4 to 16. Fig. 3.6 illustrates an example of simulated data for all four parameter choices intended to reflect the trends seen in Fig. 3.2. Aesthetic design choices were made consistent across each of the interactive ‘You Draw It’ plots; the  $y$ -axis range extended 10% beyond (above and below) the range of the simulated data points to allow for users to draw outside the simulated data set range and minimize participants anchoring their lines to the edges of the graph.

### 3.3.2 Results

In addition to the participant drawn points,  $(x_k, y_{k,drawn})$ , and the ordinary least squares (OLS) regression fitted values,  $(x_k, \hat{y}_{k,OLS})$ , a regression equation with a slope based on the first principal component (PCA) was used to calculate fitted values,  $(x_k, \hat{y}_{k,PCA})$ . For each set of simulated data and parameter choice, the PCA regression slope,  $\hat{\beta}_{1,PCA}$ , and y-intercept,  $\hat{\beta}_{0,PCA}$ , were determined by using the

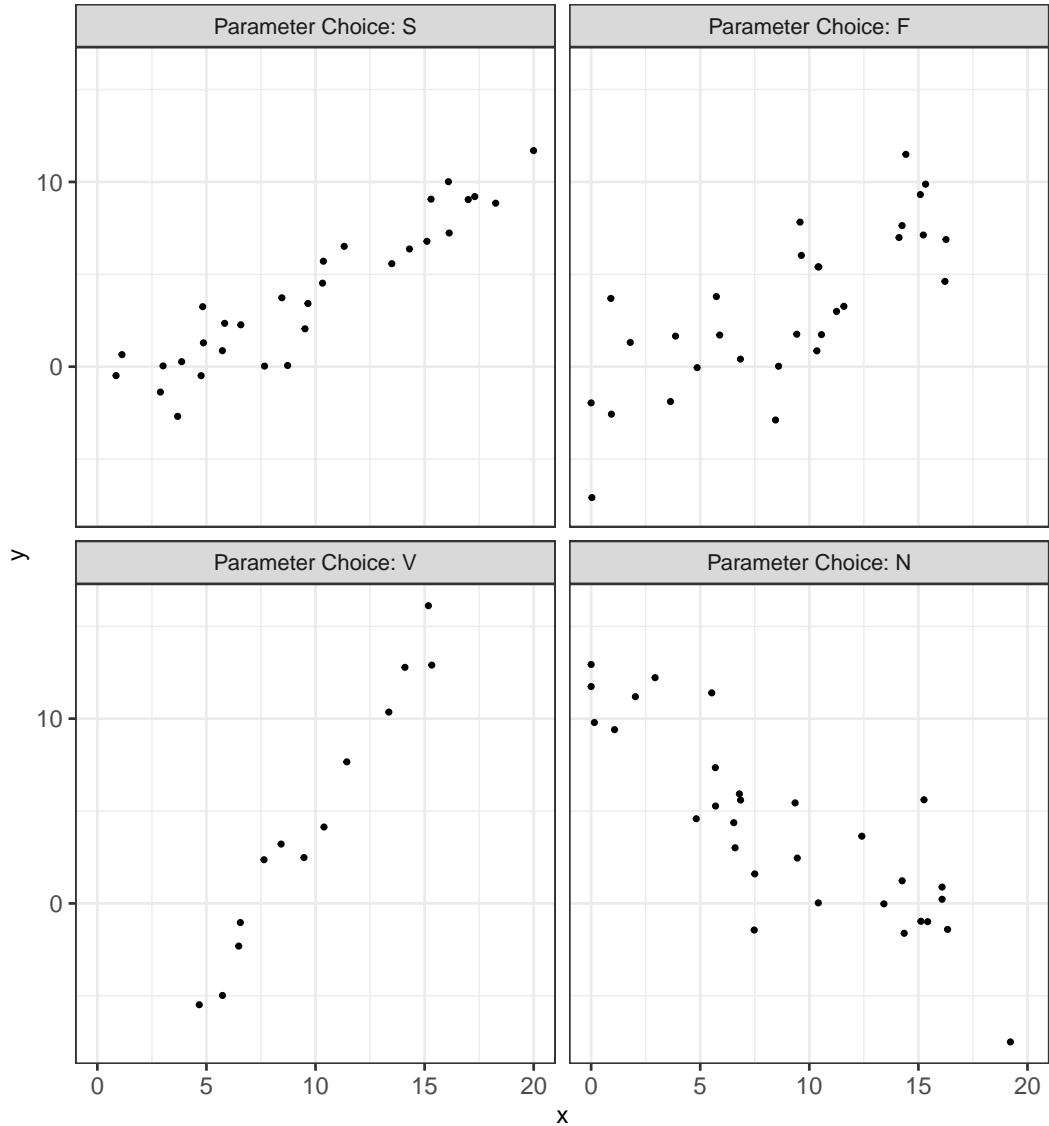


Figure 3.6: Scatter plots of example simulated data in Eye Fitting Straight Lines in the Modern Era sub-study. The four parameter choices were intended to reflect the trends seen in Fig. 3.2.

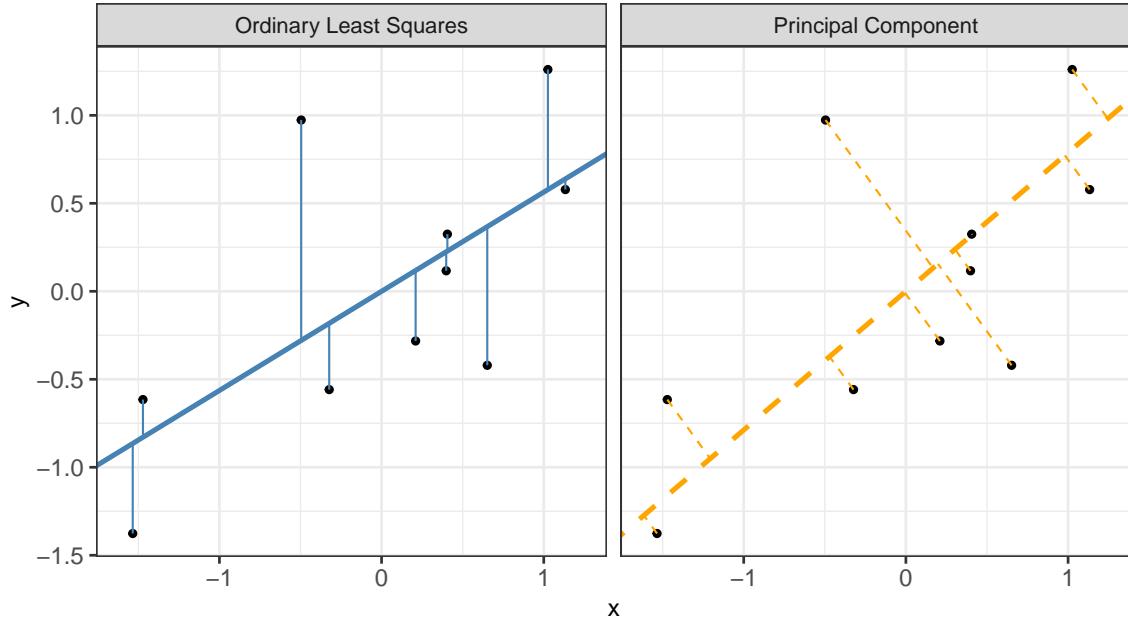


Figure 3.7: Comparison between an OLS regression equation which minimizes the vertical distance of points from the line and a regression equation with a slope calculated by the first principal component which minimizes the smallest distance of points from the line.

`mcreg` function in the `mcr` package in R (Schuetzenmeister & Model, 2021) which implements Deming regression (equivalent to a regression based on the slope of the first principal component). Fitted values,  $\hat{y}_{k,PCA}$  were then obtained every 0.25 increment across the domain from the PCA regression equation,

$\hat{y}_{k,PCA} = \hat{\beta}_{0,PCA} + \hat{\beta}_{1,PCA}x_k$ . Fig. 3.7 illustrates the difference between an OLS regression equation which minimizes the vertical distance of points from the line and a regression equation with a slope calculated by the first principal component which minimizes the smallest distance of points from the line.

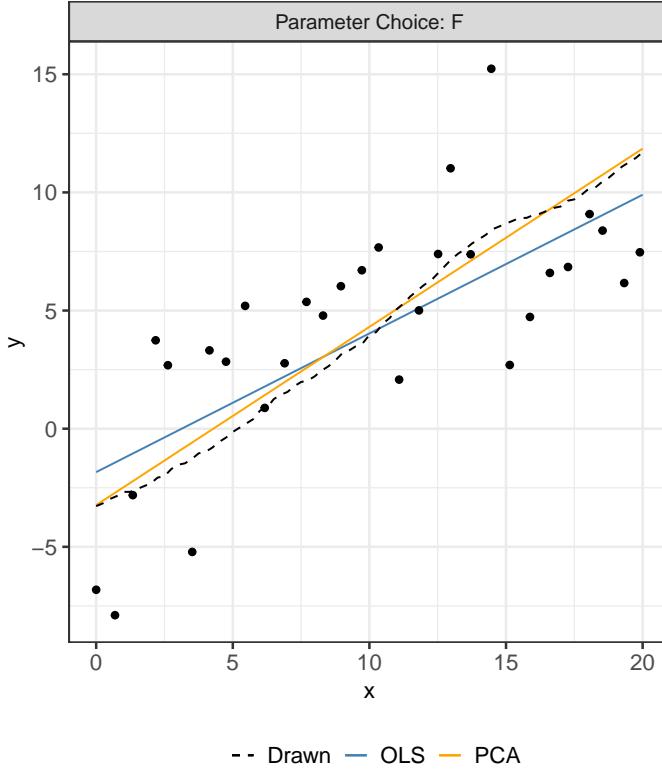


Figure 3.8: Example of three trend lines showing the the OLS fitted, PCA fitted, and participant drawn values overlaid on the simulated data points.

For each participant, the final data set used for analysis contains  $x_{ijk}$ ,  $y_{ijk,drawn}$ ,  $\hat{y}_{ijk,OLS}$ , and  $\hat{y}_{ijk,PCA}$  for parameter choice  $i = 1, 2, 3, 4$ ,  $j = 1, \dots, N_{\text{participant}}$ , and  $x_{ijk}$  value  $k = 1, \dots, 4x_{max} + 1$ . Using both a linear mixed model (LMM) and a generalized additive mixed model (GAMM), comparisons of vertical residuals in relation to the OLS fitted values ( $e_{ijk,OLS} = y_{ijk,drawn} - \hat{y}_{ijk,OLS}$ ) and PCA fitted values ( $e_{ijk,PCA} = y_{ijk,drawn} - \hat{y}_{ijk,PCA}$ ) were made across the domain. Fig. 3.8 displays an example of all three fitted trend lines for parameter choice F.

Using the `lmer` function in the `lme4` package (Bates et al., 2014), a LMM is fit separately to the OLS and PCA residuals, constraining the fit to a linear trend.

Parameter choice,  $x$ , and the interaction between  $x$  and parameter choice were treated as fixed effects with a random participant effect accounting for variation due to participant. The LMM equation for each fit (OLS and PCA) residuals is given by:

$$y_{ijk,drawn} - \hat{y}_{ijk,fit} = e_{ijk,fit} = [\gamma_0 + \alpha_i] + [\gamma_1 x_{ijk} + \gamma_{2i} x_{ijk}] + p_j + \epsilon_{ijk} \quad (3.2)$$

where

- $y_{ijk,drawn}$  is the drawn y-value for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment of  $x$ -value
- $\hat{y}_{ijk,fit}$  is the fitted y-value for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment of  $x$ -value corresponding to either the OLS or PCA fit
- $e_{ijk,fit}$  is the residual between the drawn and fitted y-values for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment of  $x$ -value corresponding to either the OLS or PCA fit
- $\gamma_0$  is the overall intercept
- $\alpha_i$  is the effect of the  $i^{th}$  parameter choice (F, S, V, N) on the intercept
- $\gamma_1$  is the overall slope for  $x$
- $\gamma_{2i}$  is the effect of the parameter choice on the slope
- $x_{ijk}$  is the  $x$ -value for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment
- $p_j \sim N(0, \sigma_{\text{participant}}^2)$  is the random error due to the  $j^{th}$  participant's characteristics
- $\epsilon_{ijk} \sim N(0, \sigma^2)$  is the residual error.

Eliminating the linear trend constraint, the `bam` function in the `mgcv` package (S. Wood, 2003, 2004, 2011, 2017; S. Wood, Pya, & Säfken, 2016) is used to fit a

GAMM separately to the OLS and PCA residuals to allow for estimation of smoothing splines. Parameter choice was treated as a fixed effect with no estimated intercept and a separate smoothing spline for  $x$  was estimated for each parameter choice. A random participant effect accounting for variation due to participant and a random spline for each participant accounted for variation in spline for each participant. The GAMM equation for each fit (OLS and PCA) residuals is given by:

$$y_{ijk,drawn} - \hat{y}_{ijk,fit} = e_{ijk,fit} = \alpha_i + s_i(x_{ijk}) + p_j + s_j(x_{ijk}) \quad (3.3)$$

where

- $y_{ijk,drawn}$  is the drawn y-value for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment of  $x$ -value
- $\hat{y}_{ijk,fit}$  is the fitted y-value for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment of  $x$ -value corresponding to either the OLS or PCA fit
- $e_{ijk,fit}$  is the residual between the drawn and fitted y-values for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment of  $x$ -value corresponding to either the OLS or PCA fit
- $\alpha_i$  is the intercept for the parameter choice  $i$
- $s_i$  is the smoothing spline for the  $i^{th}$  parameter choice
- $x_{ijk}$  is the  $x$ -value for the  $i^{th}$  parameter choice,  $j^{th}$  participant, and  $k^{th}$  increment
- $p_j \sim N(0, \sigma_{\text{participant}}^2)$  is the error due to participant variation
- $s_j$  is the random smoothing spline for each participant.

Fig. 3.9 and Fig. 3.10 show the estimated trends of residuals (vertical deviation of participant drawn points from both the OLS and PCA fitted points) as modeled

by a LMM and GAMM respectively. A random sample of 75 participants was selected to display individual participant residuals behind the overall residual trend. Examining the plots, the estimated trends of PCA residuals (orange) appear to align more parallel and closer to the  $y = 0$  horizontal (dashed) line than the OLS residuals (blue). In particular, this trend is more prominent in parameter choices with large variances (F and N). These results are consistent to those found in Mosteller et al. (1981) indicating participants fit a trend line closer to the estimated regression line with the slope of the first principal component than the estimated OLS regression line. This study established ‘You Draw It’ as a method for graphical testing and reinforced the differences between intuitive visual model fitting and statistical model fitting, providing information about human perception as it relates to the use of statistical graphics.

### 3.4 Prediction of Exponential Trends

The results from the first sub-study validated ‘You Draw It’ as a tool for testing graphics and introduced an appropriate statistical analysis method for comparing visually fitted trend lines to statistical regression results. This sub-study was designed to test an individual’s ability to make predictions for exponentially increasing data on both the log and linear scales, addressing cognitive understanding of log scales. Participants were asked to draw a line using their computer mouse through the exponentially increasing trend shown on both the log and linear scale.

#### 3.4.1 Data Generation

All data processing was conducted in R before being passed to the D3.js source code. A total of  $N = 30$  points  $(x_i, y_i), i = 1, \dots, N$  were generated for

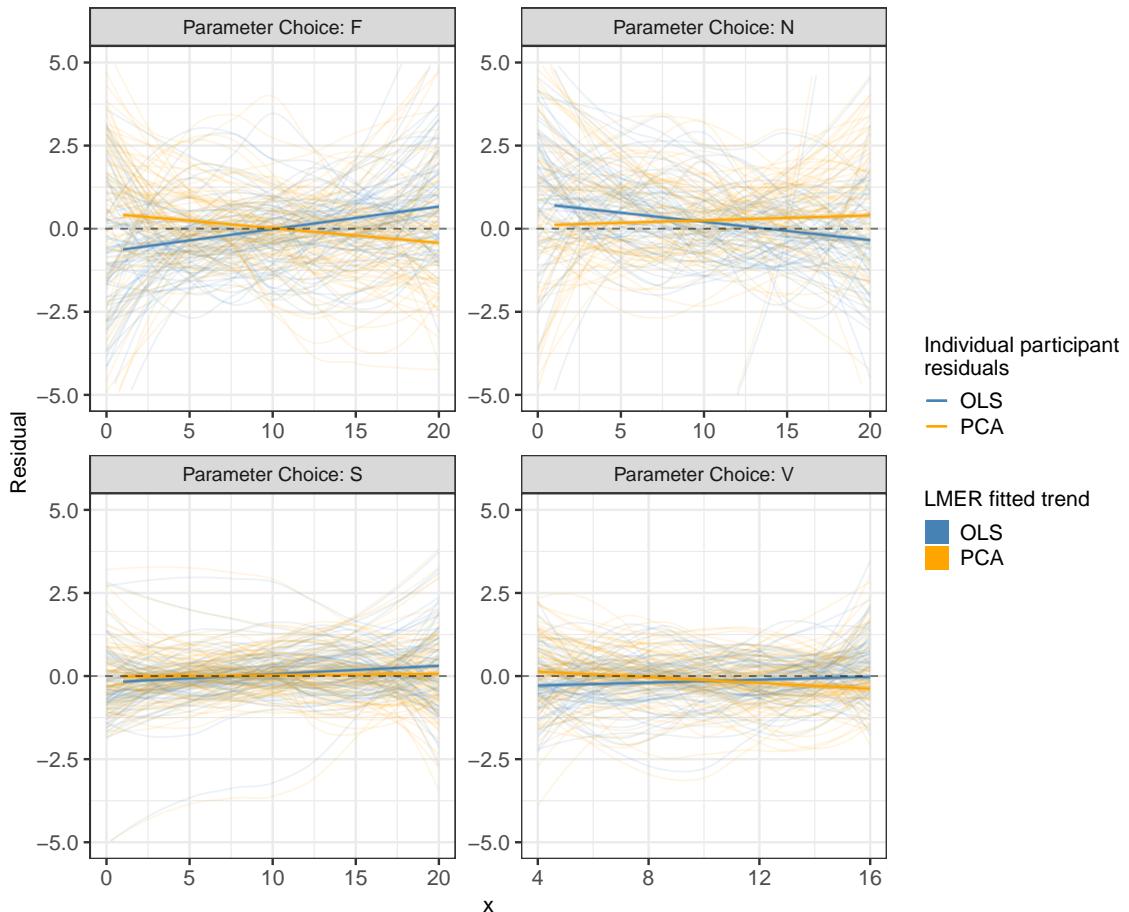


Figure 3.9: Estimated trends of residuals (vertical deviation of participant drawn points from both the OLS (blue) and PCA (orange) fitted points) as fit by the linear mixed model. A random sample of 75 participants was selected to display the individual participant residuals behind the overall trend.

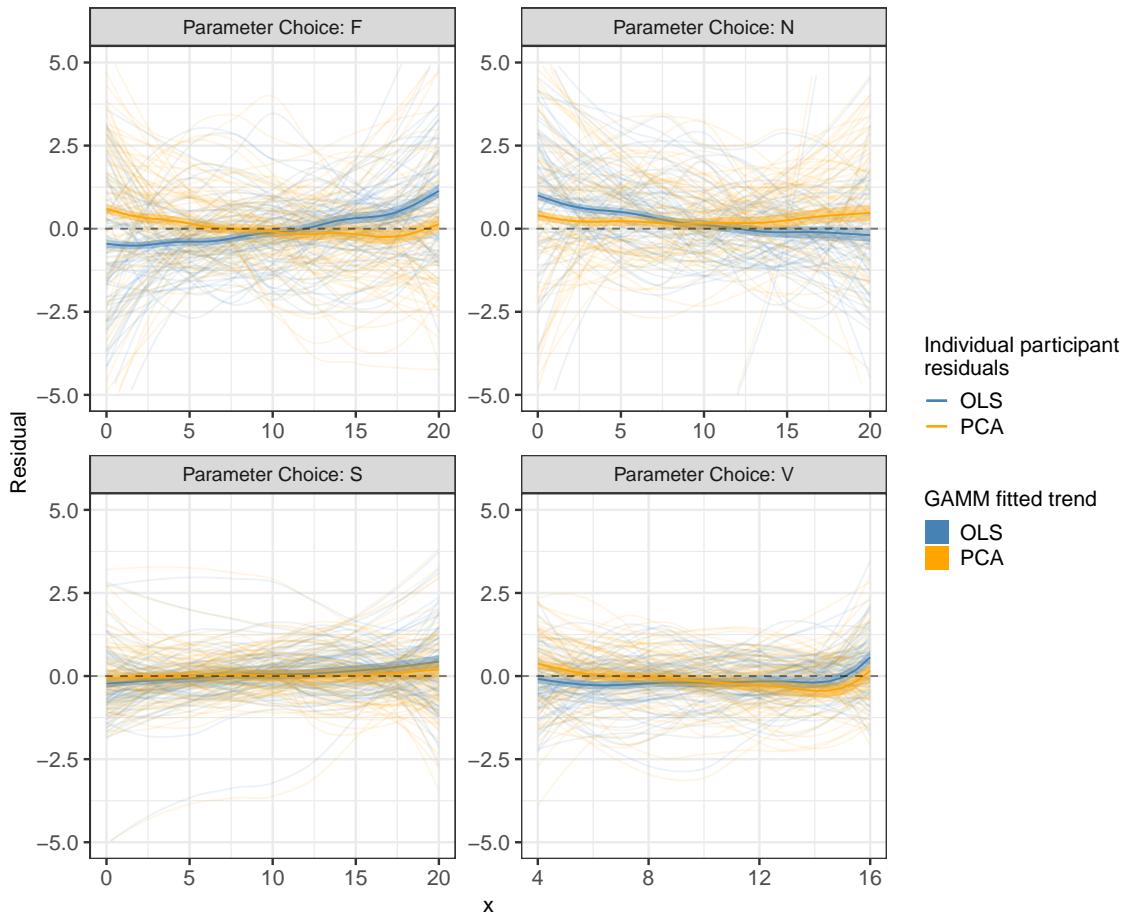


Figure 3.10: Estimated trends of residuals (vertical deviation of participant drawn points from both the OLS (blue) and PCA (orange) fitted points) as fit by the generalized additive mixed model. A random sample of 75 participants was selected to display the individual participant residuals behind the overall trend.

$x_i \in [x_{min}, x_{max}]$  where  $x$  and  $y$  have an exponential relationship. Data were simulated based on a one parameter exponential model with multiplicative errors:

$$y_i = e^{\beta x_i + e_i} \quad (3.4)$$

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

The parameter,  $\beta$ , was selected to reflect the rate of exponential growth with  $e_i$  generated by rejection sampling in order to guarantee the points shown align with that of the fitted line displayed in the initial plot frame. A nonlinear least squares regression is then fit to the simulated points in order to obtain the best fit line and fitted values in 0.25 increments across the domain,  $(x_m, \hat{y}_{m,NLS}), k = 1, \dots, 4x_{max} + 1$ . The data simulation function then outputs a list of point data and line data both indicating the parameter identification,  $x$  value, and corresponding simulated or fitted  $y$  value. The data simulation procedure is described in Algorithm 4.

---

**Algorithm 4** Prediction of Exponential Trends Data Simulation

---

- **Input Parameters:**  $\beta$  growth rate; standard deviation from exponential curve  $\sigma$ ; sample size of points  $N = 30$ ; domain  $x_{min}$  and  $x_{max}$ ; fitted value increment  $x_{by} = 0.25$ .

- **Output Parameters:** List of point data and line data each indicating the parameter identification,  $x$  value, and corresponding simulated or fitted  $y$  value.
- 1: Randomly select and jitter  $N = 30$   $x$ -values along the domain,  $x_{i=1:N} \in [0, 20]$ .
  - 2: Generate “good” errors,  $e_{i=1:N}$  based on  $N(0, \sigma)$  by setting a constraint requiring the mean of the first  $\frac{1}{3}N$  errors  $< |2\sigma|$ .
  - 3: Simulate point data based on  $y_i = e^{\beta x_i + e_i}$ .
  - 4: Fit the equation  $\log(y_i) = \beta x_i$  to obtain an estimated starting value  $\beta_0$ .
  - 5: Obtain nonlinear least squares regression coefficient,  $\hat{\beta}_{NLS}$ , for the simulated point data fitting using the `nls` function in the base `stats` R package.
  - 6: Obtain fitted values every 0.25 increment across the domain from the nonlinear least squares regression  $\hat{y}_{m,NLS} = e^{\hat{\beta}_{NLS} x_m}$ .
  - 7: Output data list of point data and line data each indicating the parameter identification,  $x$  value, and corresponding simulated or fitted  $y$  value.
-

Model equation parameter,  $\beta$ , was selected to reflect two exponential growth rates (low:  $\beta = 0.10, \sigma = 0.09$  and high:  $\beta = 0.23, \sigma = 0.25$ ) as determined by visual inspection with growth rate parameter selection from the lineup study in [Chapter 2](#) used as a starting point. Each growth rate parameter was used to simulate data across a domain of 0 to 20. The two simulated data sets (low and high exponential growth rates) were then shown four times each by truncating the points shown at both 50% and 75% of the domain as well as on both the log and linear scales for a total of eight interactive plots reflecting a factorial treatment design. [Appendix B](#) displays visual examples of all eight interactive plots. Aesthetic design choices were made consistent across each of the interactive ‘You Draw It’ plots; the  $y$ -axis extended 50% below the lower limit of the simulated data range and 200% beyond the upper limit of the simulated data range to allow for users to draw outside the data set range, and participants were asked to start drawing at 50% of the domain (for example, at  $x = 10$ ). Reflecting the treatment design for each plot, the  $y$ -axis was assigned to be displayed on either the linear scale or log scale.

### 3.4.2 Results

A LOESS smoother (local regression) was fit to each user line to allow for visual inspection. For each participant  $l = 1, \dots, N_{\text{participant}}$ , the final data set used for analysis contained  $x_{ijklm}, y_{ijklm, \text{drawn}}, \hat{y}_{ijklm, \text{loess}}$ , and  $\hat{y}_{ijklm, \text{NLS}}$  for growth rate  $i = 1, 2$ , points truncated  $j = 1, 2$ , scale  $k = 1, 2$  and  $x_{ijklm}$  value for increment  $m = 1, \dots, 81$ . [Fig. 3.11](#) displays spaghetti plots for each of the eight treatment combinations. The spaghetti plot with a high growth rate suggests participants underestimated the exponential trend when asked to draw a trend line on the linear scale compared to when asked to draw a trend line on the log scale. In particular, this suggestion is most noticeable when points are truncated at 50% with the

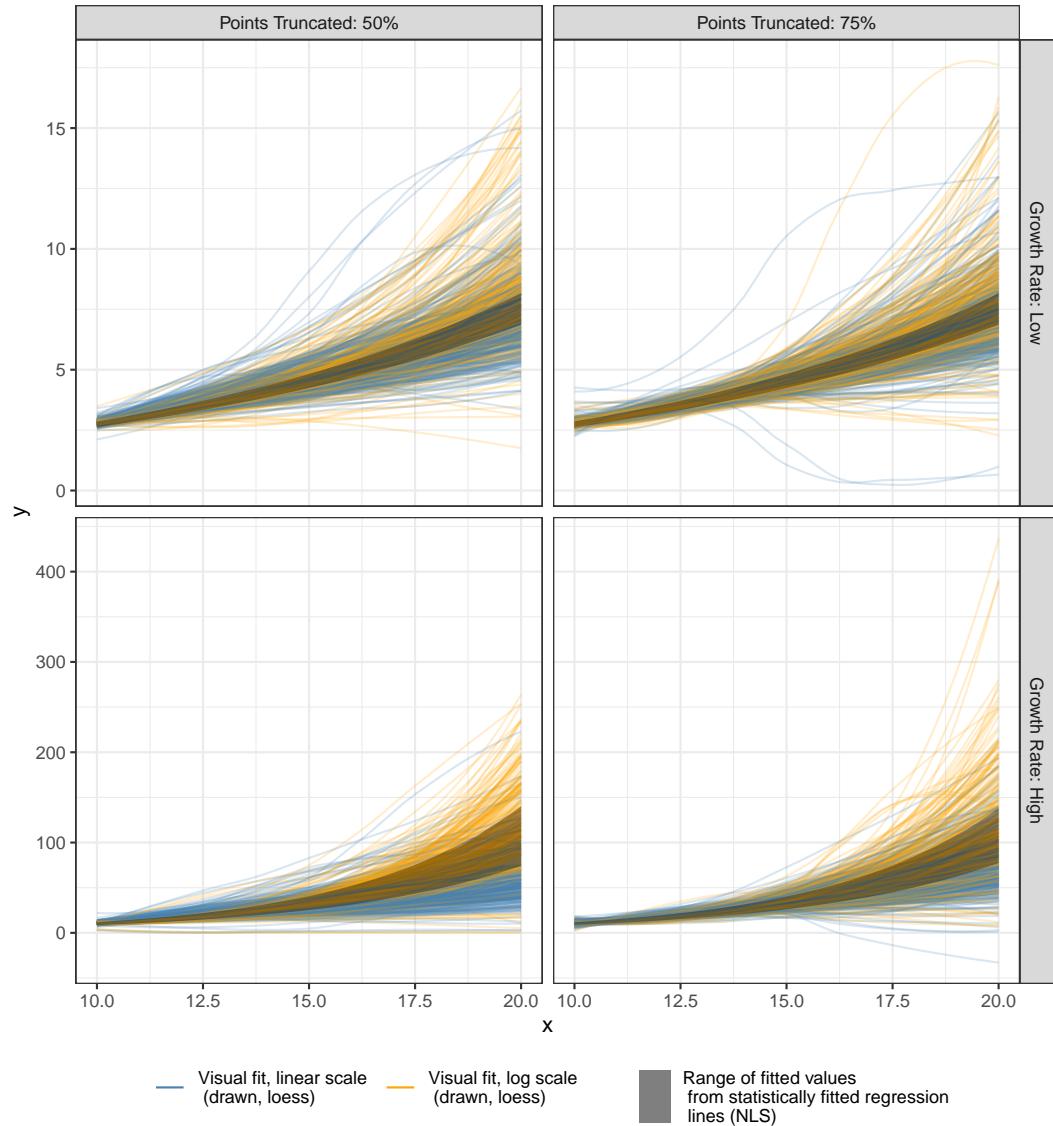


Figure 3.11: Spaghetti plot of results from the exponential prediction sub-study. Participants drawn lines on the linear scale are shown in blue and the log scale are shown in orange. Variability in the statistically fitted regression lines occurred due to a unique data set being simulated for each individual; the gray band shows the range fitted values from the statistically fitted regression lines.

underestimation beginning at a later  $x$  value when points are truncated at 75%.

Allowing for flexibility, the `bam` function in the `mgcv` package (S. Wood, 2003, 2004, 2011, 2017; S. Wood et al., 2016) was used to fit a GAMM to estimate trends

of vertical residuals from the participant drawn line in relation to the NLS fitted values ( $e_{ijklm,NLS} = y_{ijklm,drawn} - \hat{y}_{ijklm,NLS}$ ) across the domain. The combination between growth rate, point truncation, and scale was treated as a fixed effect with no estimated intercept and a separate smoothing spline for  $x$  was estimated for each treatment combination. A random participant effect accounting for variation due to participant and a random spline for each participant accounted for variation in spline for each participant. The GAMM equation for residuals is given by:

$$y_{ijklm,drawn} - \hat{y}_{ijklm,NLS} = e_{ijklm,nls} = \tau_{ijk} + s_{ijk}(x_{ijklm}) + p_l + s_l(x_{ijklm}) \quad (3.5)$$

where

- $y_{ijklm,drawn}$  is the drawn y-value for the  $l^{th}$  participant,  $m^{th}$  increment, and  $ijk^{th}$  treatment combination
- $\hat{y}_{ijklm,NLS}$  is the NLS fitted y-value for the  $l^{th}$  participant,  $m^{th}$  increment, and  $ijk^{th}$  treatment combination
- $e_{ijklm,NLS}$  is the residual between the drawn y-value and fitted y-value for the  $l^{th}$  participant,  $m^{th}$  increment, and  $ijk^{th}$  treatment combination
- $\tau_{ijk}$  is the intercept for the  $i^{th}$  growth rate,  $j^{th}$  point truncation, and  $k^{th}$  scale treatment combination
- $s_{ijk}$  is the smoothing spline for the  $ijk^{th}$  treatment combination
- $x_{ijklm}$  is the  $x$ -value for the  $l^{th}$  participant,  $m^{th}$  increment, and  $ijk^{th}$  treatment combination
- $p_l \sim N(0, \sigma_{\text{participant}}^2)$  is the error due to the  $l^{th}$  participant's characteristics
- $s_l$  is the random smoothing spline for the  $l^{th}$  participant.

Fig. 3.12 shows the estimated trends of the residuals (vertical deviation of

participant drawn points from NLS fitted points) as modeled by the GAMM. Examining the plots, the estimated trends of residuals for predictions made on the linear scale (blue) appear to deviate from the  $y = 0$  horizontal (dashed) line indicating underestimation of exponential growth. In comparisons, the estimated trends of residuals for predictions made on the log scale (orange) follow closely to the  $y = 0$  horizontal (dashed) line, implying exponential trends predicted on the log scale are more accurate than those predicted on the linear scale. In particular, this trend is more prominent in high exponential growth rates where underestimation becomes prominent after the aid of points is removed.

### 3.5 Discussion and Conclusion

The intent of this chapter was to establish ‘You Draw It’ as a method and tool for testing graphics then use this tool to determine the cognitive implications of displaying data on the log scale. Eye Fitting Straight Lines in the Modern Era replicated the results found in Mosteller et al. (1981). When shown points following a linear trend, participants tended to fit the slope of the first principal component over the slope of the least squares regression line. This trend was most prominent when shown data simulated with larger variances. The reproducibility of these results serve as evidence of the reliability of the ‘You Draw It’ method.

In Prediction of Exponential Trends, the ‘You Draw It’ method was used to test an individual’s ability to make predictions for exponentially increasing data. Results indicate that underestimation of exponential growth occurs when participants were asked to draw trend lines on the linear scale and that there was an improvement in accuracy when trends were drawn on the log scale. This phenomena is strongly supported for high exponential growth rates. Improvement in predictions are made

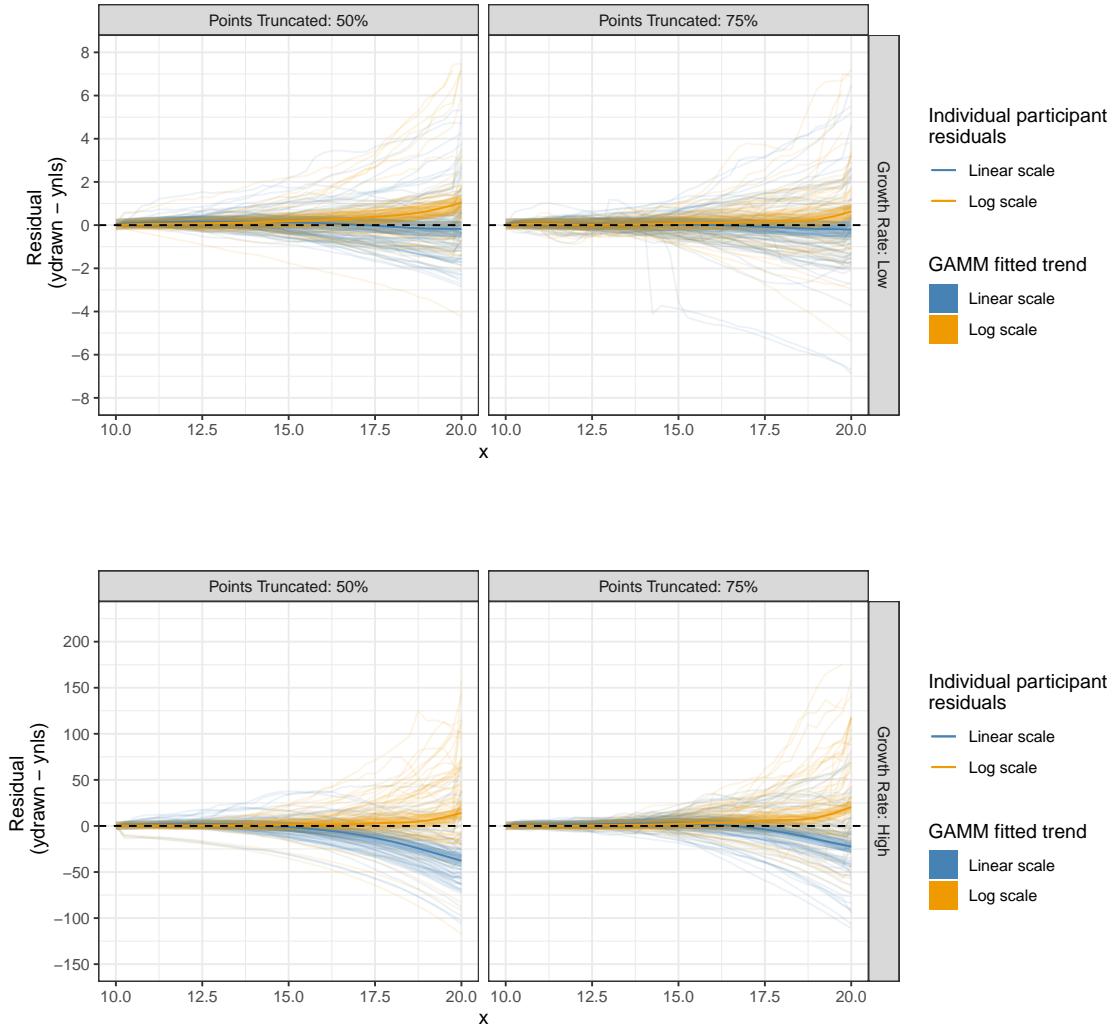


Figure 3.12: Estimated trends of residuals (vertical deviation of participant drawn points from NLS fitted points) as fit by the generalized additive mixed model. Deviation for visual trends predicted on the linear scale are shown in blue and deviation for visual trends predicted on the log scale are shown in orange. A random sample of 75 participants was selected to display the individual participant residuals behind the overall trend.

when points along the exponential trend are shown as indicated by the discrepancy in results for treatments with points truncated at 50% compared to 75% of the domain.

The results of this study suggest that there are cognitive advantages to log scales when making predictions of exponential trends. Participants' predictions were more accurate at high growth rates when participants drew trend lines on the log scale compared to the linear scale. Further investigation is necessary to determine the implications of using log scales when translating exponential graphs to numerical values; we address this problem in the next chapter.

## Chapter 4

### Numerical Translation and Estimation

#### 4.1 Introduction

The previous two chapters explored the use of log scales through differentiation and visual prediction of trends. These graphical tasks were conducted independent of context - no information about the data itself or even numerical scale values were provided to participants; instead, participants focused how our visual system perceives and identifies patterns in exponential growth. In order to understand the cognitive implications of displaying exponentially increasing data on a log scale, it is essential to evaluate graph comprehension as it relates to the contextual scenario of the data shown. This is a complex inferential process which requires participants to engage with the data by quantitatively transforming information in the chart (W. Cleveland & McGill, 1984, 1985). In this study, we asked participants to translate a graph of exponentially increasing data into real value quantities and extend their estimations by comparing two data points.

##### 4.1.1 Graph Comprehension

Graph comprehension is heavily dependent on the questions being asked of the viewer; therefore, how these questions are phrased is an important aspect of

comprehension and must be given deliberate consideration (Graesser et al., 2014). Evaluation of how viewers explore a new and complex graphic requires long-term interaction with the chart displaying the data (Becker, Moore, & Lawrence, 2019). While it is difficult to obtain an accurate representation of a viewers understanding of the graphic with a fixed set of numerical estimates, three levels of graph comprehension have emerged from literature (Curcio, 1987; Friel et al., 2001; Glazer, 2011; Jolliffe, 1991; R. Wood, 1968). The three behaviors related to graph comprehension involve (1) literal reading of the data (elementary level), (2) reading between the data (intermediate level), and (3) reading beyond the data (advanced level). We present examples of questions associated with the three levels of questioning in [section 1.6.1](#). For instance, if shown a line graph of the value of a certain stock over time, an elementary level question might prompt the viewer to answer, “what was the value of stock X on June 15th?” and an intermediate level question would extend these estimates to ask the viewer, “over the first five days, how did the value of stock X change” (Friel et al., 2001). In addition to the graph’s visual features and questioning, it is important for researchers to give careful consideration to the context of the graphic on the viewers comprehension.

#### 4.1.2 Estimation Biases

Certain well-known biases such as the tendency to round to multiples of five or ten or to anchor estimates to visual cues arise from open-ended estimation tasks (Tan & Benbasat, 1990). Viewers may anchor their estimates to grid lines or round their approximations to rounded figures due to natural preference (Godlonton et al., 2018; Myers, 1954; Tan, 1994). Estimation accuracy is also affected by scale and axis labels (Dunham & Osborne, 1991); when alternate grid lines are labeled, viewers often read unlabeled grid lines as halves (Beeby & Taylor, 1973). This

misrepresentation is highlighted for asymmetric scales, such as a log scale, since spatial distance does not equate to numerical or quantitative difference. Therefore, careful consideration must be given to the choice of scale for the graphic and how the viewer will interpret the data and information displayed.

## 4.2 Study Design

Participants in this study were asked to answer six questions related to each of two contextual scenarios and an associated scatter plot shown for a total of twelve questions. The text for each scenario is presented below; the context of both scenarios was selected to be similar. Each text describes a situation in which a fictional intergalactic species is exponentially increasing in population over a time chosen to reflect the popular culture media depiction of that species (*Star trek*, 1967; *Star wars*, 1977, 1983). For simplicity, we will refer to these fictional time components as a year throughout the rest of the chapter.

**Tribble scenario.** Hi, we're Tribbles! We were taken from our native planet, Iota Germinorum IV, and brought abroad Starfleet in stardate 4500. A Starfleet scientist, Edward Larkin, genetically engineered us to increase our reproductive rate in an attempt to solve a planetary food shortage. The Tribble population on Starfleet over the next 50 Stardates (equivalent to 1 week universe time) is illustrated in the graph. We need your help answering a few questions regarding the population of Tribbles.

**Ewok scenario.** Hi, we're Ewoks! We are native to the forest moon of Endor. After the Galactic Civil War, some Ewoks traveled offworld to help Rebel veterans as 'therapy Ewoks' and began to repopulate. The

Ewok population After the Battle of Yavin (ABY) is illustrated in the graph. We need your help answering a few questions regarding the population of Ewoks offworld.

Fictional illustrations of the figures used in context were modified from artwork by Horst and included on the main page for each scenario. The scale of the graphic and data set displayed was randomly assigned to scenarios for each individual. For instance, a participant may have seen a scatter plot of data set two displayed on the linear scale paired with the Ewok scenario text and a scatter plot of data set one displayed on the log scale paired with the Tribble scenario text. The order of the two scenarios and their assigned data set and scale was randomly assigned to each individual.

We selected the six questions (Table 4.1) for graph comprehension based on the three defined levels of questioning. In each scenario, participants were first asked an open ended question, which required them to spend time exploring the data displayed in the graphic, followed by a random order of two elementary level questions and three intermediate level questions. We did not focus on advanced level questioning since extrapolation and interpolation was addressed in Chapter 2.

The estimation study in this chapter was completed last in the series of the three graphical studies and took about fifteen minutes for participants to answer all twelve estimation questions. Participants completed the series of graphical tests using a R Shiny application found [here](#). For each of the quantitative translation questions, participants were provided a basic calculator and scratchpad to aid in their estimation of values. We recorded the inputted and evaluated calculations and scratch work of each participant in order to better understand participant strategies for estimation.

Table 4.1: Estimation questions

Question type	Tribble scenario	Ewok scenario
Open Ended	Between stardates 4530 and 4540, how does the population of Tribbles change?	Between 30 and 40 ABY, how does the population of Ewoks change?
Elementary Q1	What is the population of Tribbles in stardate 4510?	What is the population of Ewoks in 10 ABY?
Elementary Q2	In what stardate does the population of Tribbles reach 4,000?	In what ABY does the population of Ewoks reach 4,000?
Intermediate Q1	From 4520 to 4540, the population increases by _____ Tribbles.	From 20 ABY to 40 ABY, the population increases by _____ Ewoks.
Intermediate Q2	How many times more Tribbles are there in 4540 than in 4520?	How many times more Ewoks are there in 40 ABY than in 20 ABY?
Intermediate Q3	How long does it take for the population of Tribbles in stardate 4510 to double?	How long does it take for the population of Ewoks in 10 ABY to double?

### 4.3 Data Generation

We generated two unique data sets with the same underlying parameter coefficients, but different errors randomly generated from the same error distribution. For each data set, a total of  $N = 50$  points  $(x_i, y_i), i = 1, \dots, N$  were generated for single increments of  $x_i \in [0, 50]$  where  $x$  and  $y$  have an exponential relationship. Data were simulated based on a three parameter exponential model

with multiplicative errors:

$$y_i = \alpha e^{\beta x_i + e_i} + \theta \quad (4.1)$$

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

The underlying parameter coefficients were selected to follow a similar growth rate and shape as the previous two studies by visual inspection while ensuring in a maximum magnitude of around 50,000. The resulting parameters selected for data generation were  $\alpha = 130$ ,  $\beta = 0.12$ ,  $\theta = 50$ , and  $\sigma = 1.5$ .

Fig. 4.1 display scatter plots of the two unique data sets on both the linear and log base two scales; a log of base two was selected in order to aid in participants estimation of time until the population doubled in ‘Intermediate Q3’ (Table 4.1). Participants were shown the graphic of both data sets on either the linear or log scale with labels adjusted to reflect the associated scenario context and scale. Grid lines for the  $y$ -axis were set to be consistent for the same scale across both data sets with the linear scale increasing by 5,000 and the log base two scale doubling, thus demonstrating the additive and multiplicative contextual appearance and interpretation of each scale respectively. Minor  $y$ -axis grid lines were removed to avoid participants anchoring to the midway point between grid lines; this is particularly important on the log scale since a half-way grid line spatially does not correspond to a half-way point numerically. Grid lines for the  $x$ -axis spanned a range of 50 years with major grid lines every ten years apart and minor grid lines indicating every five years. The time unit labels on the  $x$ -axis reflected 0 to 50 ABY (After Battle of Yavin) for the Ewok scenario and were adjusted to 4500 to 4550 Stardates for the Tribble Scenario to align with the associated popular media depiction of each figure as well as disguise the use of the same underlying data

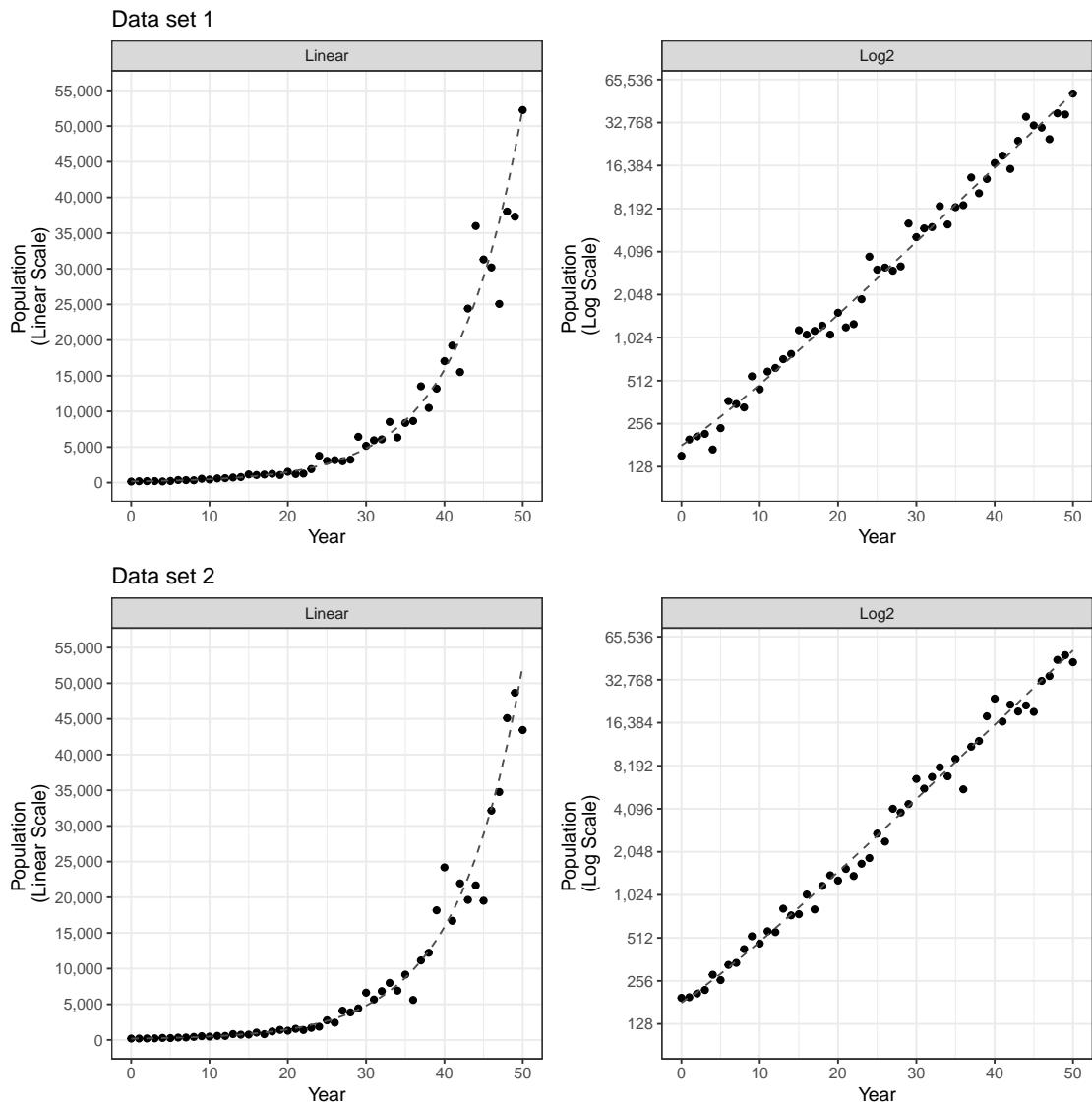


Figure 4.1: Scatter plots of the two unique data sets displayed on both the linear and log base two scales.

simulation model and estimation questions across both scenarios.

## 4.4 Results

Participant recruitment and study deployment was conducted via Prolific, a crowd sourcing website, on Wednesday, March 23, 2022 during which 302 individuals each completed all six estimation questions for each scenario (total of twelve questions per individual). The data set used for analysis contained the unique participant identification and indicated the scenario, scale, data set, and estimation question along with the participant text response or quantitative estimate, calculation input and evaluation, and associated scratch work. A total of 145 participants answered questions related to data set one on the linear scale and data set two on the log scale with 157 participants answering questions related to data set one on the log scale and data set two on the linear scale. Sketches for each question are used to demonstrate the estimation tasks participants were asked to conduct. An array of graphical displays allow for visual inspection of participant responses and provide suggestions about the cognitive implications of displaying exponentially increasing data on the log scale.

### 4.4.1 Open Ended

Before participants were asked to estimate numeric quantities, they were asked to provide an open ended response and describe how the population changed over time. This required participants to spend time exploring the graphic and reflect upon how the data displayed related to the contextual application. The `tidytext` and `corpus` packages in R (Perry, 2021; Silge & Robinson, 2016) were used to extract and stem words from participant text responses; stop words such as ‘the’ and ‘is’ as well as

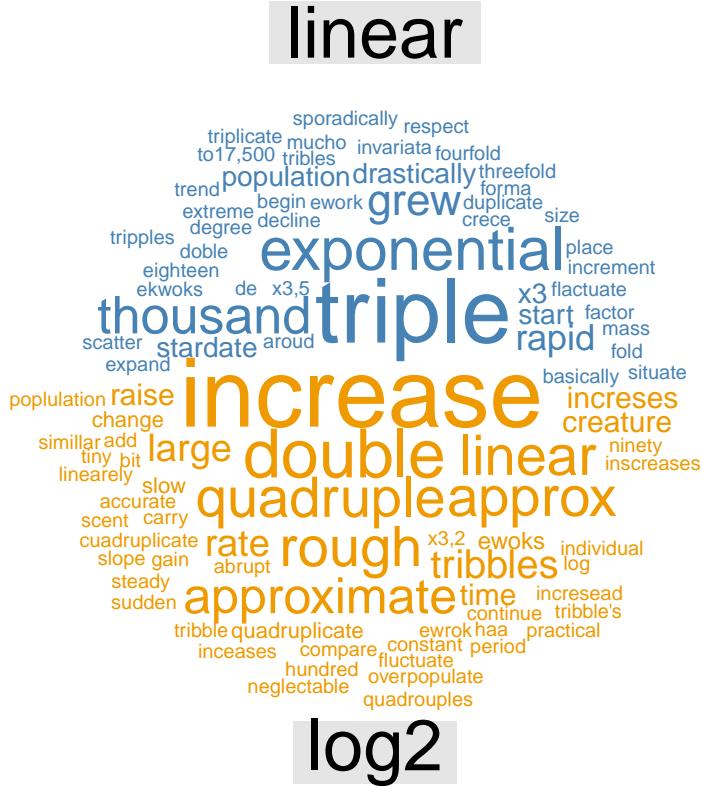


Figure 4.2: The open ended question results are displayed in a comparison word cloud which compares frequencies of words across the two scales. The maximum deviation in frequency is mapped to the size of the word and the position and color of the word is determined by the scale - linear (blue), log (orange).

numbers were removed from the cleaned word responses. The `wordcloud` package (Fellows, 2018) was used to create a cloud comparing frequencies of words across the two scales (Fig. 4.2). The comparison word cloud is generated by defining  $p_{i,j}$  as the rate in which word  $i$  occurs when describing the data on scale  $j$  where  $p_j$  is the average rate across the scales  $\sum_i \frac{p_{i,j}}{N \text{ scales}}$ . The maximum deviation for each word is calculated by  $\max_i(p_{i,j} - p_j)$  and mapped to the size of the word with the position of the word determined by the scale in which the maximum occurs.

The comparison word cloud illustrates the general terminology participants used

when describing the scatter plots shown on each scale. Participants more frequently referred to terms such as ‘exponential’ and ‘rapid’ when shown the scatter plot on the linear scale while ‘double’ and ‘quadruple’ were often used to describe the graphic when shown on the log scale; indicating participants read the  $y$ -axis labels and noticed the doubling grid lines. Participants often used “triple” to describe the data when displayed on the linear scale; one explanation might be that participants were roughly estimating the multiplicative change between grid lines. For example, in year 40, the trend lands roughly around 15,000 and ends near 45,000 (three times as large) in year 50. The use of the term ‘linear’ when participants are describing the appearance of the data displayed on the log scale suggests that a portion of participants described the visual appearance of the data independent of the axis labels; without further context, we do not have enough information to determine whether this implies participants were not recognizing the data was exponentially increasing rather than linearly increasing due to the change in contextual appearance caused by the choice of scale.

#### **4.4.2 Elementary Q1: Estimation of population**

In order to examine the effect of scale on literal reading of the data, participants were asked, *“What is the population in year 10?”* (Fig. 4.3). The true estimated population in year 10 based on the underlying parameter estimates was 481.61 with simulated points of 445.48 and 466.9 for data sets one and two respectively. The median participant estimate across both scales and data sets was 500, with innerquartile ranges of 500 and 400 for data set one and data set two respectively when displayed on the linear scale and 48 and 12 for data set one and data set two respectively when displayed on the log scale.

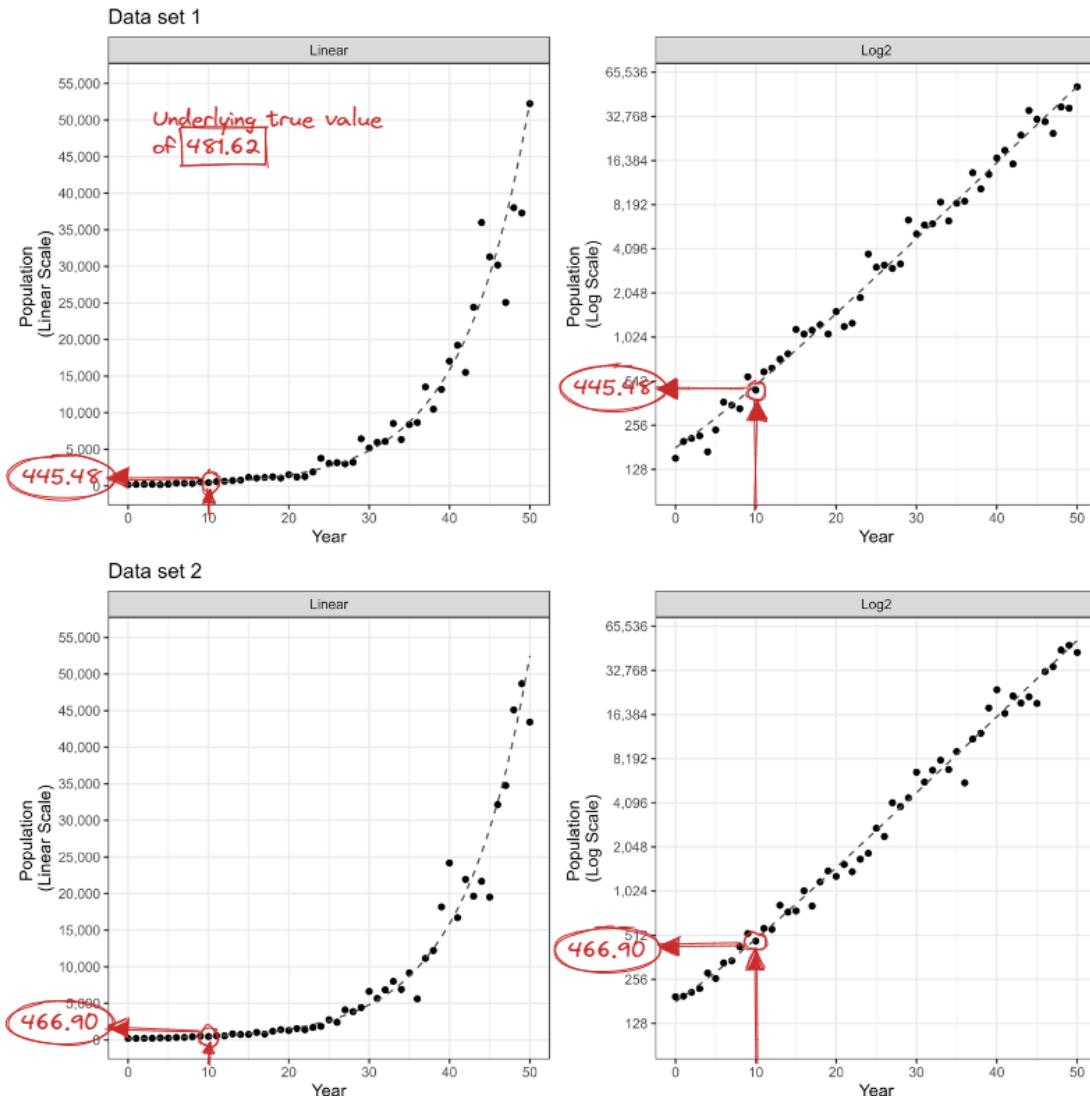


Figure 4.3: Sketch of the estimation procedure asked in Elementary Q1. Participants first locate 10 along the  $x$ -axis and move upward until they believe they have found the correct location on the curve; then participants look to the  $y$ -axis for their estimated population.

Density plots were used to illustrate the distribution of the quantitative estimates provided by participants. (Fig. 4.4) reveals a larger variance in quantitative population estimates made on the linear scale compared to the log scale. As expected, it is clear that participants were anchoring to grid lines and base ten values as highlighted by the high density of estimates at 512 and 500 on the log scale as well as local maximums near multiples of ten such as 500 and 1000.

During the study, participants were explicitly asked to estimate the population during year 10; this value corresponds to a low magnitude where the population is condensed in a small region on the linear scale as opposed to later in time when larger magnitudes in population can be seen. While the results provided support for less variability in the estimated population in year 10 on the log scale, it is important to evaluate the accuracy of estimates along the full domain. In two estimation questions related to intermediate level reading between the data, participants are asked to provide an increase and change in population between years 20 and 40, thus requiring participants to make first level estimates at these locations (Fig. 4.10 and Fig. 4.14). In order to understand the effect of the location along the domain and in turn the magnitude of the population being estimated, we extracted first level estimates for years 20 and 40 from participant calculations and scratch work.

In order to examine whether participants who used the provided resources for estimation differed in their numerical estimations from those who did not, we first compared population estimates from the explicitly asked year 10 location; these comparisons are provided in Appendix 3b. About half of the participants fell into the category which provided scratch work and half did not. It was determined there was no substantial difference or bias in estimates between the two groups, therefore,

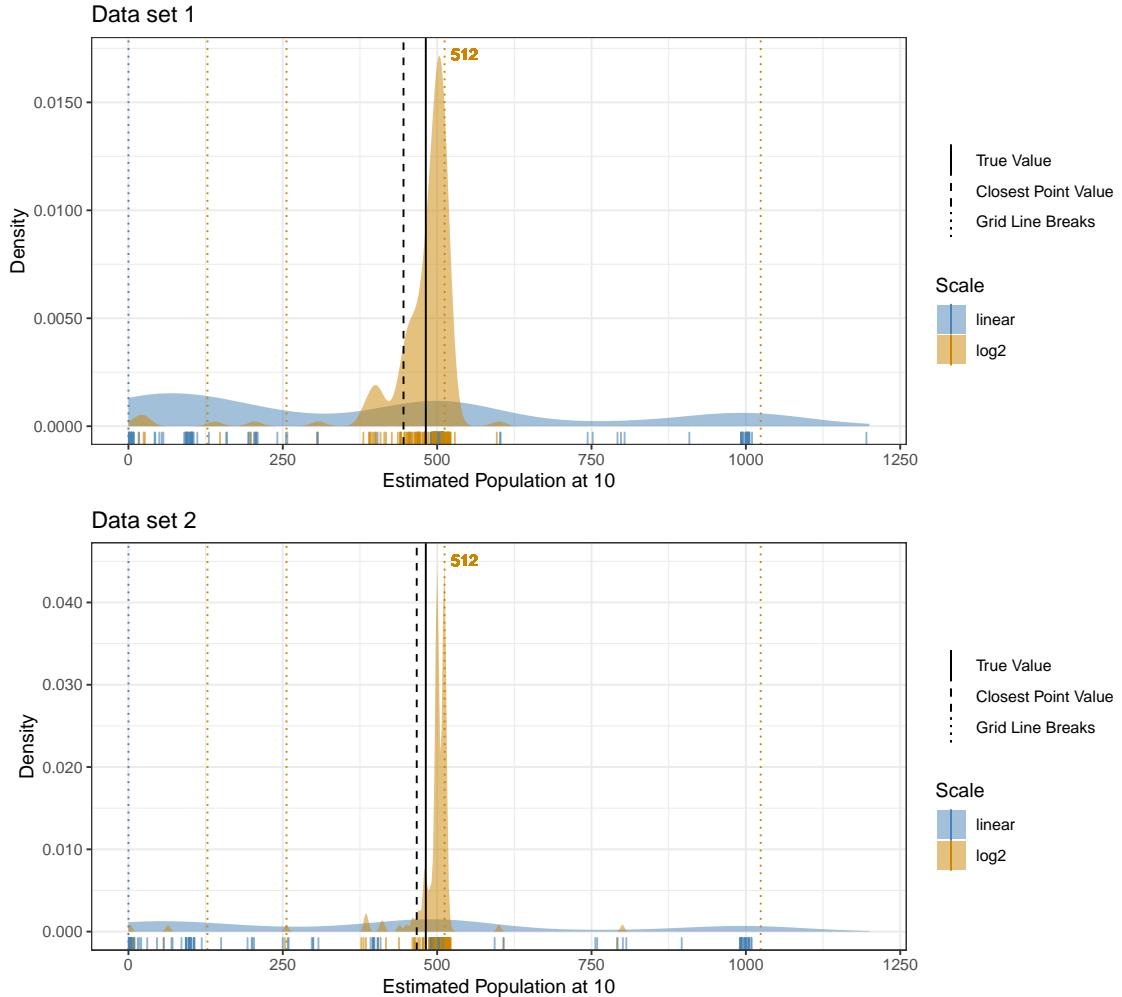


Figure 4.4: Density of the participant estimates for the population in year 10. Colors are associated to scale - linear (blue) and log (orange) - and vertical lines indicate the true value based on the underlying model equation (black solid), closest point value based on the simulated data set (black dashed), and grid lines shown on the graphs (blue/orange dotted). A jittered rug plot along the  $x$ -axis shows where participant estimates were made. The two unique data sets are shown separately.

we proceeded to examine the estimated populations across scales from the first level estimates.

The true population from the underlying parameters in year 20 was 1483.01 with closest simulated point values of 1529.19 and 1288.9 for data sets one and two respectively; this location still results in a relatively low magnitude of population, but is closer to the crux of the exponential curve. In year 40, the true population from the underlying parameters in year 40 is 15846.35 with closest simulated point values of 17046.94 and 24186.34 for data sets one and two respectively. It is important to note that there is a difference in simulated point values in year 40 between the two data sets; as a result, the multiplicative error causes larger variance in simulated points for later years and larger population magnitudes.

Population estimates for year 10 from participants who used the scratchpad and first level estimates for years 20 and 40 are shown with spaghetti plots in Fig. 4.5 and Fig. 4.6 displayed on both the linear and log scale to aid in visual evaluation. The scale in which the estimate was made is indicated blue for linear and orange for log with the segments mapped from the participant estimated population to the true year based on the underlying data equation. Previously noted, the simulated point corresponding to year 40 in data set two has a large deviation from the true underlying data equation; Fig. 4.6 highlights that some participants were reading the data points as opposed to first detecting the underlying trend and making estimates based on the identified trend. This provides argument that estimates are highly subjective to the particular data set. As the year increases, we observe an increased accuracy in estimates made on the linear scale while estimates made on the log scale suffer in accuracy due to strong anchoring to grid lines and the larger quantitative difference between grid lines as population magnitudes increase. For

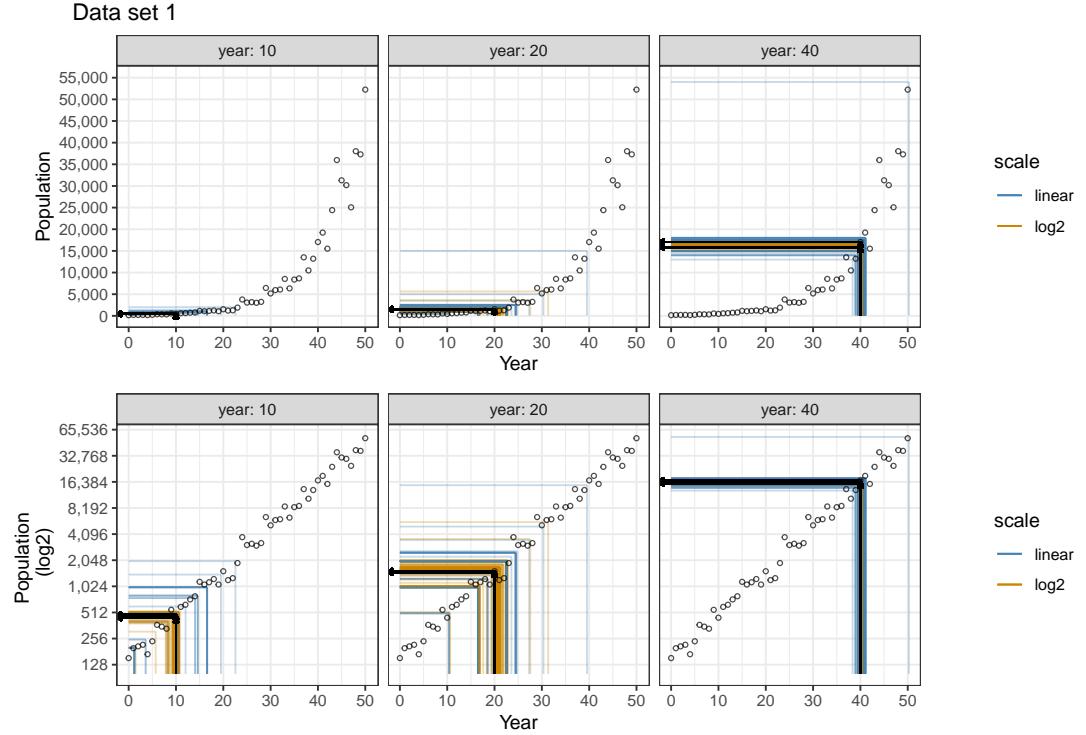


Figure 4.5: Visual evaluation of participants' estimates of the population at years 10, 20, and 40 on data set 1. When work was shown, first level estimates were extracted from participant calculations and scratch pad notes for years 20 and 40. Spaghetti plots are displayed on the linear scale (top) and log scale (bottom) with both scale estimates shown on each - linear (blue), log (orange). The year was calculated from the underlying model equation based on the population estimate provided by the participant. Gray arrows indicate the true value and closest point value as demonstrated in Fig. 4.3.

instance, on the log scale, there was a tendency to overestimate the population for year 20 from data set one, underestimate the population for year 20 from data set two, and overestimate the population for year 40 from data set two. Inaccurate first level estimations can lead to consequences in estimations which require participants to make comparisons between two points (e.g. Intermediate Q1 and Q2).

In extracting participant first level estimates from their calculation and scratch work, we observed participants were resistant to estimating between grid lines and

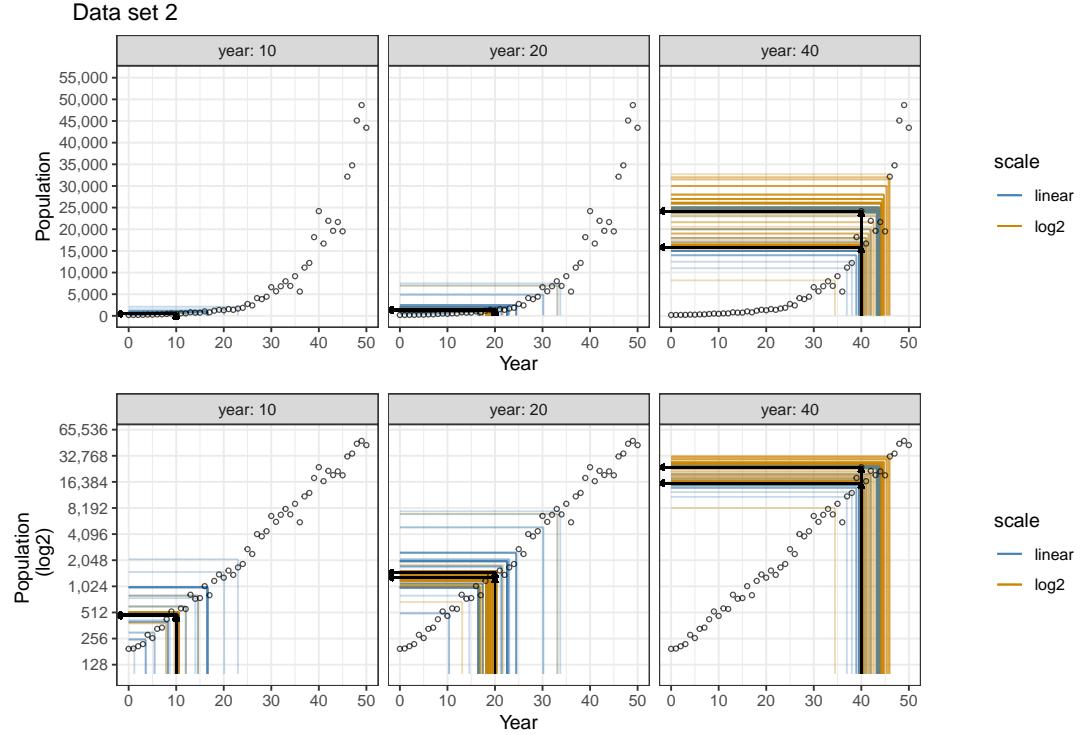


Figure 4.6: Visual evaluation of participants' estimates of the population at years 10, 20, and 40 on data set 2. When work was shown, first level estimates were extracted from participant calculations and scratch pad notes for years 20 and 40. Spaghetti plots are displayed on the linear scale (top) and log scale (bottom) with both scale estimates shown on each - linear (blue), log (orange). The year was calculated from the underlying model equation based on the population estimate provided by the participant. Gray arrows indicate the true value and closest point value as demonstrated in Fig. 4.3.

had a greater tendency to anchor their estimates to the grid line estimates on the log scale. Fig. 4.7 illustrates the number of participants who provided that estimate on either the linear or log base two scale. True values are based on the underlying model equation, closest simulated point values, and grid line breaks are indicated by the horizontal line types. In particular, for year 40 in data set one, the closest point (17046.94) falls close to the log grid line (16384); participants greatly anchored to the grid line of 16384 with some participants adjusting to 16500 or 17000, anchoring again to a base ten value. In a similar situation, for year 40 in data set two, the

closest point (24186.34) falls close to the linear grid line (25000); more participants adjusted their estimates to 24500 or 24000 rather than anchoring to the grid line. This suggests that participants were more likely to provide estimates which deviated from grid lines when making estimates on the linear scale, indicating they are more comfortable with interpreting values on a linear scale as opposed to the log scale. When participants made estimates between grid lines on the log scale as indicated by their scratch work, they tended to estimate “halfway” between the two values indicated by the grid line breaks. For example, 1536 was a common population estimate for year 20 in data set one because visually the location estimation lands about halfway between grid lines 1024 and 2048 (Fig. 4.10 and Fig. 4.14). Another common halfway point on the log scale occurred at 24576 which visually lands between grid lines 16384 and 32768 for year 40 in data set two. Participant calculations and scratch work provides support that participants equated these as halfway numerically as indicated by the selected work provided below:

*Sample work 1*

$$2048 - 1024 = 1024$$

$$1024/2 = 512$$

$$512 + 1024 = 1536$$

*Sample work 2*

$$2048 + 1024 = 3072$$

$$3072/2 = 1536$$

*Sample work 3*

$$32768 - 16384 = 16384$$

$$32768 - 16384 = 16384$$

$$16384 * 2 = 32768$$

$$16384/2 = 8192$$

$$8192 + 16384 = 24576.$$

In particular, sample work 3 demonstrates the participant processing the log base two mapping as they repeatedly calculate the distance between two grid lines by subtraction and multiplication; they however then go on to estimate halfway between the two grid lines by equating spatial distance and quantitative difference. This indicates a lack of understanding of log mapping where the spatial equivalence does not correspond to numeric equivalence; in other words, spatially halfway

between two grid lines does not result in a numeric value halfway between the quantitative grid line labels.

In conclusion, Elementary Q1 and the first level population estimates extracted from participant calculations and scratch work indicate that accuracy for low magnitudes are more accurate with lower variance in those estimates on the log scale than on the linear scale. Accuracy of population estimates made on the linear scale improve as the magnitude of the population increases. The results also provided support for the idea}that participants have a strong tendency to anchor their estimates to both grid lines and a base ten framework with resistance to estimating between grid lines on the log scale in particular, leading to a sacrifice in accuracy for larger magnitudes. Participant calculations and scratch work revealed a lack of understanding of logarithmic mapping due to considering spatial distance as indicative of numerical distance.

#### **4.4.3 Elementary Q2: Estimation of time**

In addition to estimating the population from a given year, participants were asked, "*In what year does the population reach 4000?*" (Fig. 4.8). This required literal reading of the data by mapping a value given on the  $y$ -axis to its corresponding value on the  $x$ -axis. The true estimated year based on the underlying equation in which the population reached exactly 4000 was 28.45. Unlike the previous question, there was no exact simulated point that aligned with the quantity to be estimated; the closest points for data set one occurred at years 24 (population 3774.9) and 30 (population 5174.12) and for data set two at years 27 (population 3859.22) and 28 (population 4099.69). The median year estimated by participants for data set one was 24 on both scales with innerquartile ranges of 1

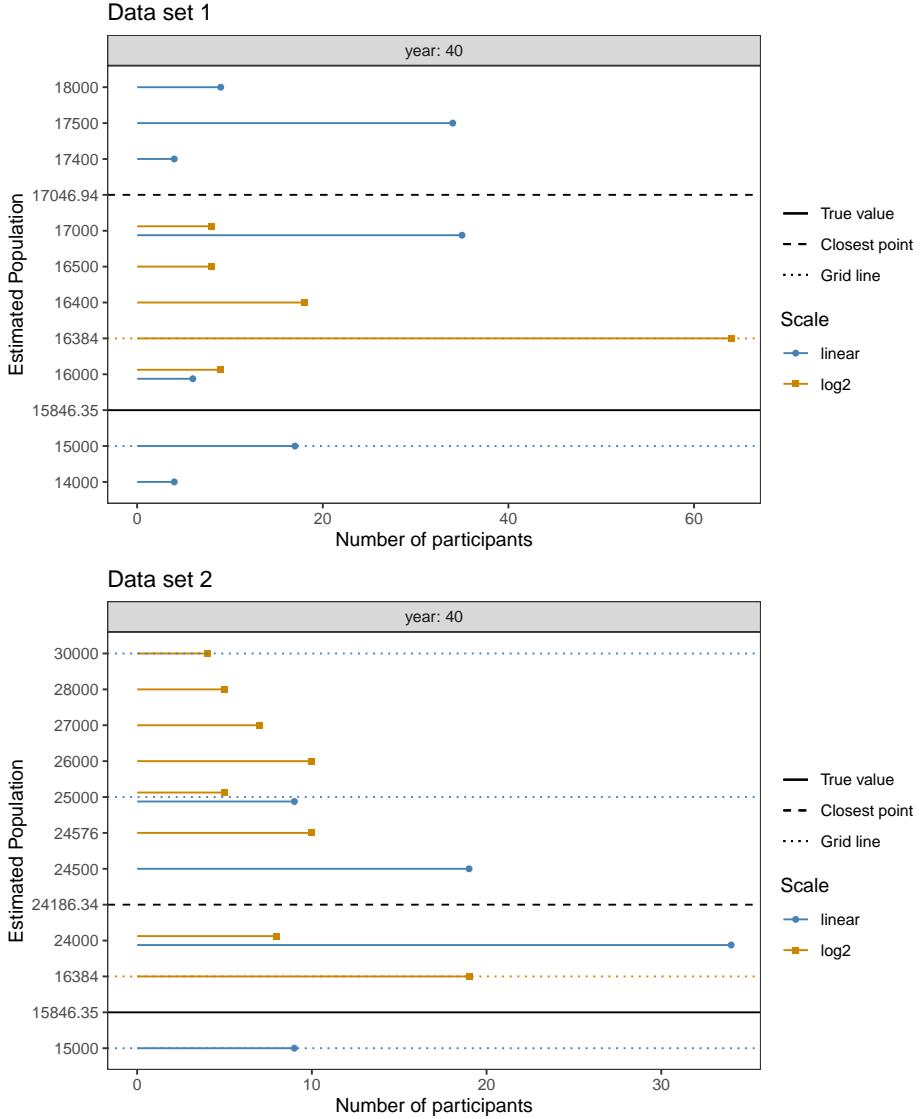


Figure 4.7: Estimated populations in year 40 provided by more than three participants are shown in the dot chart. The  $x$ -axis indicates the number of participants who provided the estimate marked on the  $y$ -axis in ascending numerical order. Colors are associated to scale - linear (blue) and log (orange) - and horizontal lines indicate the true value based on the underlying model equation (black solid), closest point value based on the simulated data set (black dashed), and grid lines shown on the graphs (blue/orange dotted). The two unique data sets are shown separately.

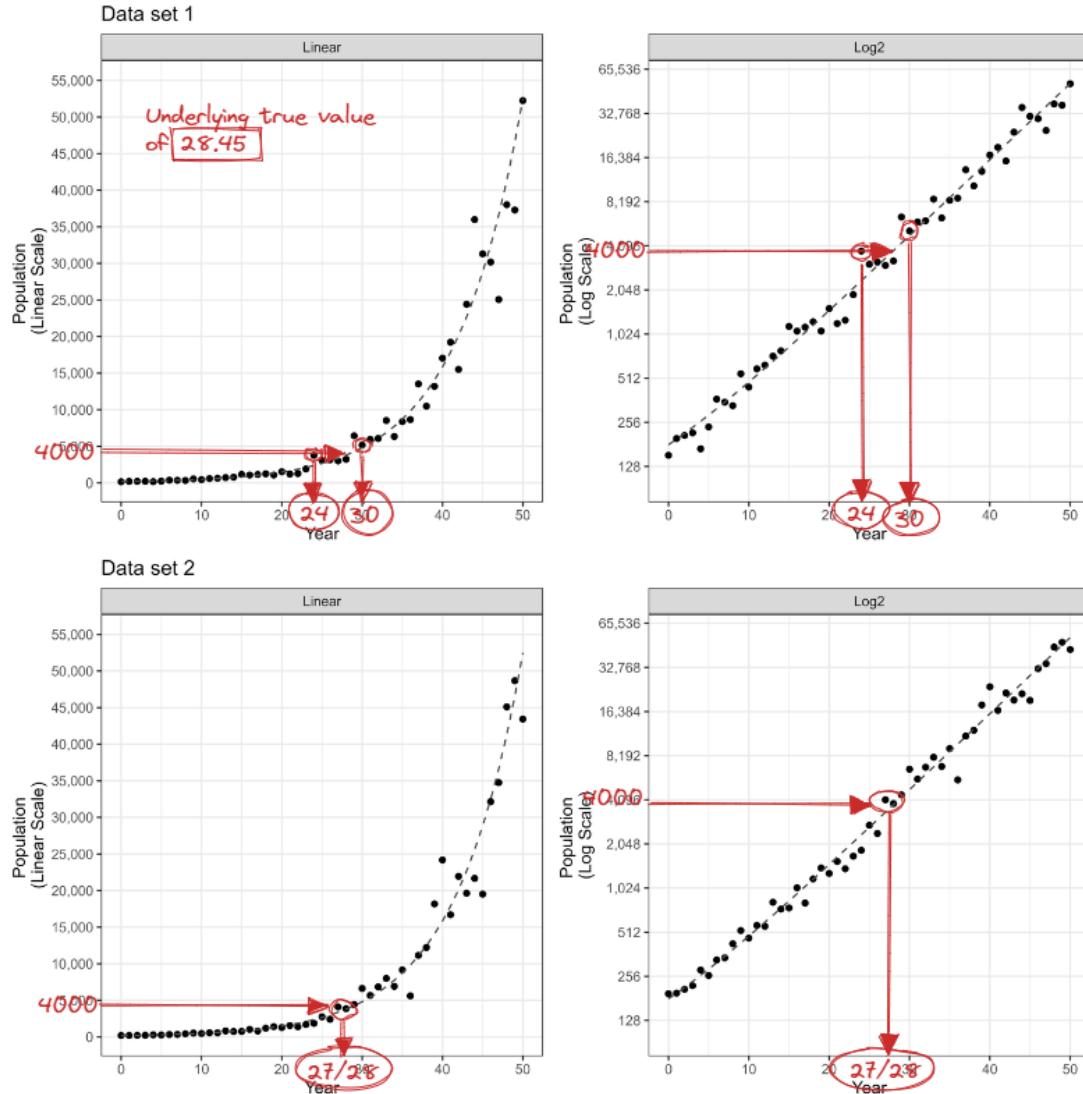


Figure 4.8: Sketch of the estimation procedure asked in Elementary Q2. Participants first locate 4000 along the  $y$ -axis and move to the right until they believe they have found the correct location on the curve; then participants look down to the  $x$ -axis for their estimated year.

and 3 for the linear and log scale respectively; the median for data set two occurred at 27 for both data sets with innerquartile ranges of 2 and 1 for the linear and log scale respectively.

While a small portion of participants provided estimates of years 5, 10 and 15,

the density plots in Fig. 4.9 focus on reasonable participant estimates between years 20 and 35. A population of 4000 occurs around a medium magnitude and is thus distinguishable on the linear scale, making the estimated location more visible. Participants were consistently accurate across both the linear and log scales with a larger variance for data set one when estimates were made on the log scale. One possible explanation for the difference in variation between data sets is that some participants were first visually fitting a trend on the log scale (results in a visually linear trend) while some participants were basing their estimates off the closest point (year 24). These competing strategies are clearly visible on the plot: some participants overestimated the closest point, while others made estimates more consistent with the true value based on the underlying (mean) equation. On the log scale, participants were able to strongly anchor their estimates to the grid line break of 4096 and provide accurate year estimates by counting between grid lines on the  $x$ -axis with few participants making estimates between years (for example, 27.5). However, participants still had a tendency to anchor to a base ten framework as indicated by an increase in the density of estimates occurring at year 30.

Results from Elementary Q2 provide support that participants accurately estimated the year in which the population reaches 4000 on both scales. The accuracy on the linear scale can be explained by the visibility of a medium magnitude along with participant ability to make accurate estimates between grid lines on a linear scale. The population given aligned closely with grid line 4096 on the log scale, allowing participants to strongly anchor to the grid line for their estimation. In particular, for data set one, participants were slightly more likely to base their estimates off the underlying trend line on the log scale than on the linear scale. Estimated years were often provided in whole numbers and few participants

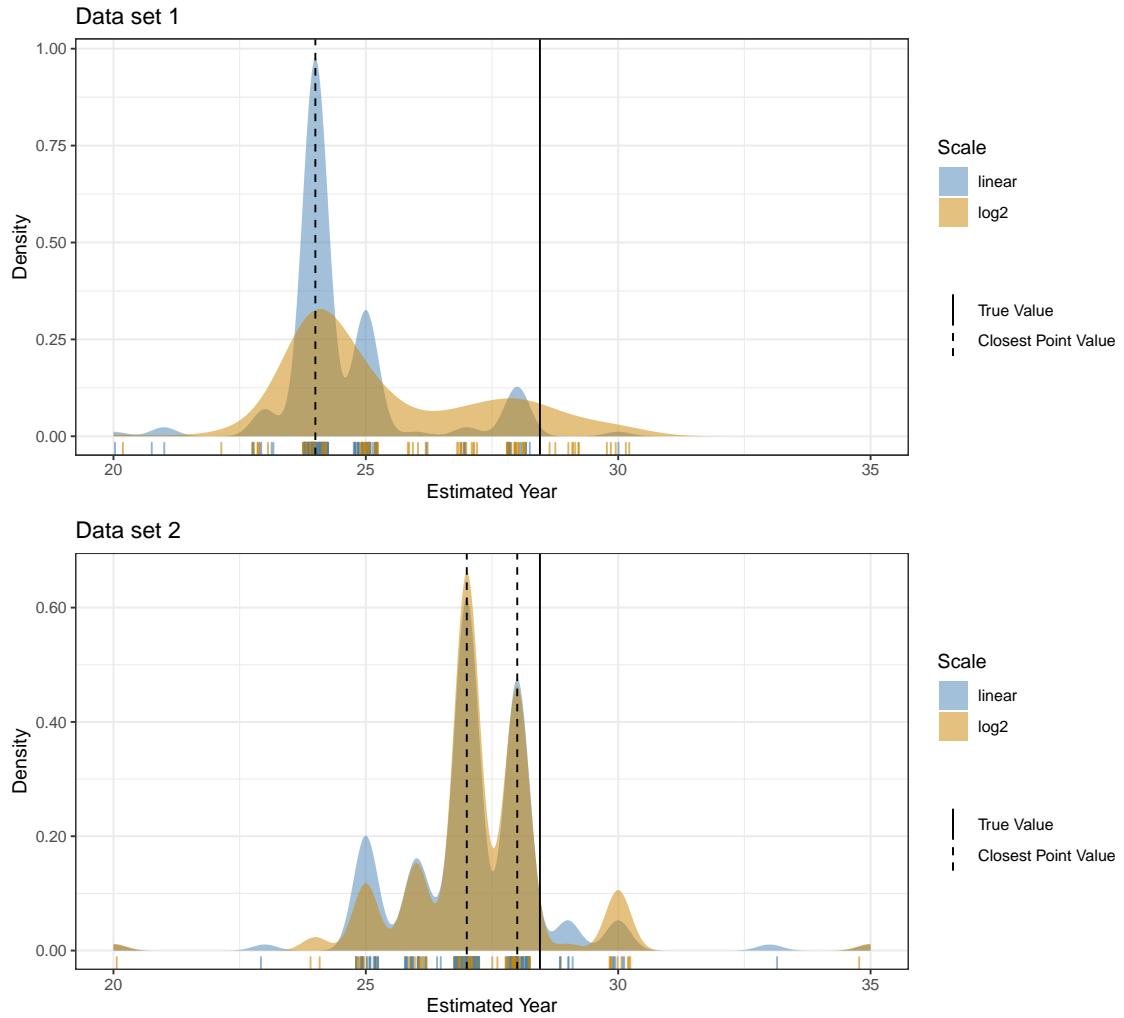


Figure 4.9: Density of the participant estimates for the year in which the population reaches 4000. Colors are associated to scale - linear (blue) and log (orange) - and vertical lines indicate the true value based on the underlying model equation (black solid) and closest point value based on the simulated data set (black dashed). A jittered rug plot along the  $x$ -axis shows where participant estimates were made. The two unique data sets are shown separately. The plot shows anchoring occurred to the closest point as shown by an increase in density around the dashed line. Density peaks occurred at whole values indicating rounding errors.

showed an understanding that the population of interest could occur between years.

#### 4.4.4 Intermediate Q1: Additive increase in population

Intermediate level questions required participants to read between the data and make comparisons between points. Participants were asked, "*From 20 to 40, the population increases by \_\_\_\_\_ [creatures].*" (Fig. 4.10). The questioning was selected carefully to prompt participants to make an additive comparison of populations between two years. In order to make this comparison, participants must have first made an accurate first level estimate in both years and then subtract the two estimates. Sample participant work below shows correct logic on both the linear and log scales:

*Sample work 4: correct logic (linear)*

$$15000 - 2500 = 12500$$

Scratchpad: In 20 ABY the population of Ewoks was 2500,

in 40 ABY the population was 15 000,

i would make a subtraction

*Sample work 5: correct logic (log)*

$$2048 - 1024 = 1024$$

$$1024 - 512 = 512$$

$$1024 + 512 = 1536$$

$$16384 - 1536 = 14848$$

Scratchpad: 20 aby 1536

40 16384.

The true estimated increase in population from year 20 to 40 based on the underlying equation is 14363.34 (15846.35 - 1483.01) with increases based on the closest points of 15517.75 (17046.94 - 1529.18) and 22897.45 (24186.34 - 1288.91) for data sets one and two respectively. The median estimated increase for data set one was 15000 (IQR = 3000) for the linear scale and 14784 (IQR = 2000) for the log scale while data set two resulted in larger estimates and variability with a median increase of 17500 (IQR = 10625) and 16500 (IQR = 8952) for the linear and log scale respectively. The discrepancy in the summary between the two data sets provides further support that participants were inspecting the simulated data points

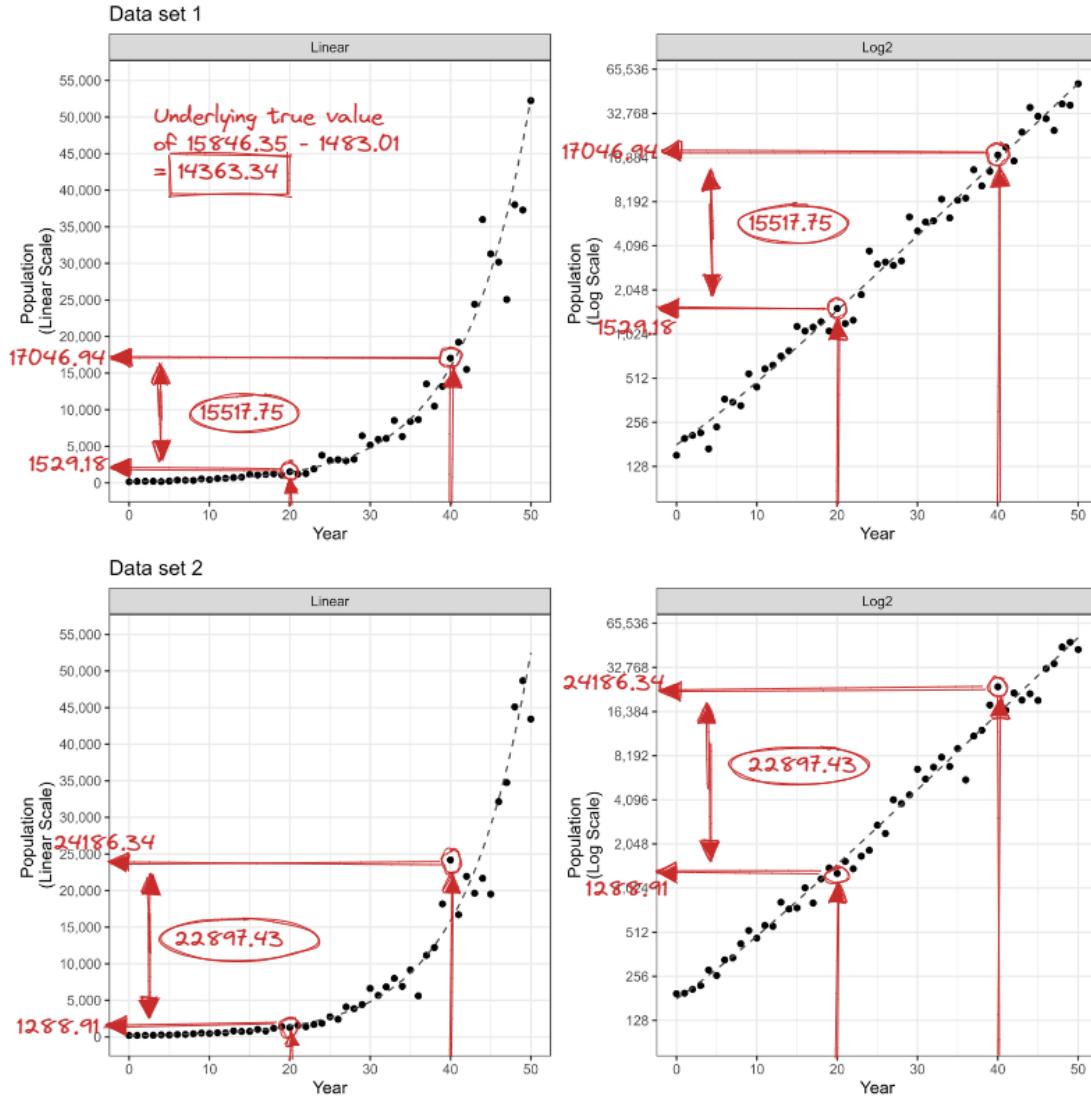


Figure 4.10: Sketch of the estimation procedure asked in Intermediate Q1. Participants make first level population estimates at years 20 and 40, then calculate the difference between the two values.

in order to make their estimates.

Fig. 4.11 and Fig. 4.12 display the density for estimated increases in population as made by participants for data set one and two respectively. There were a considerable amount of estimated increases near zero indicating that some

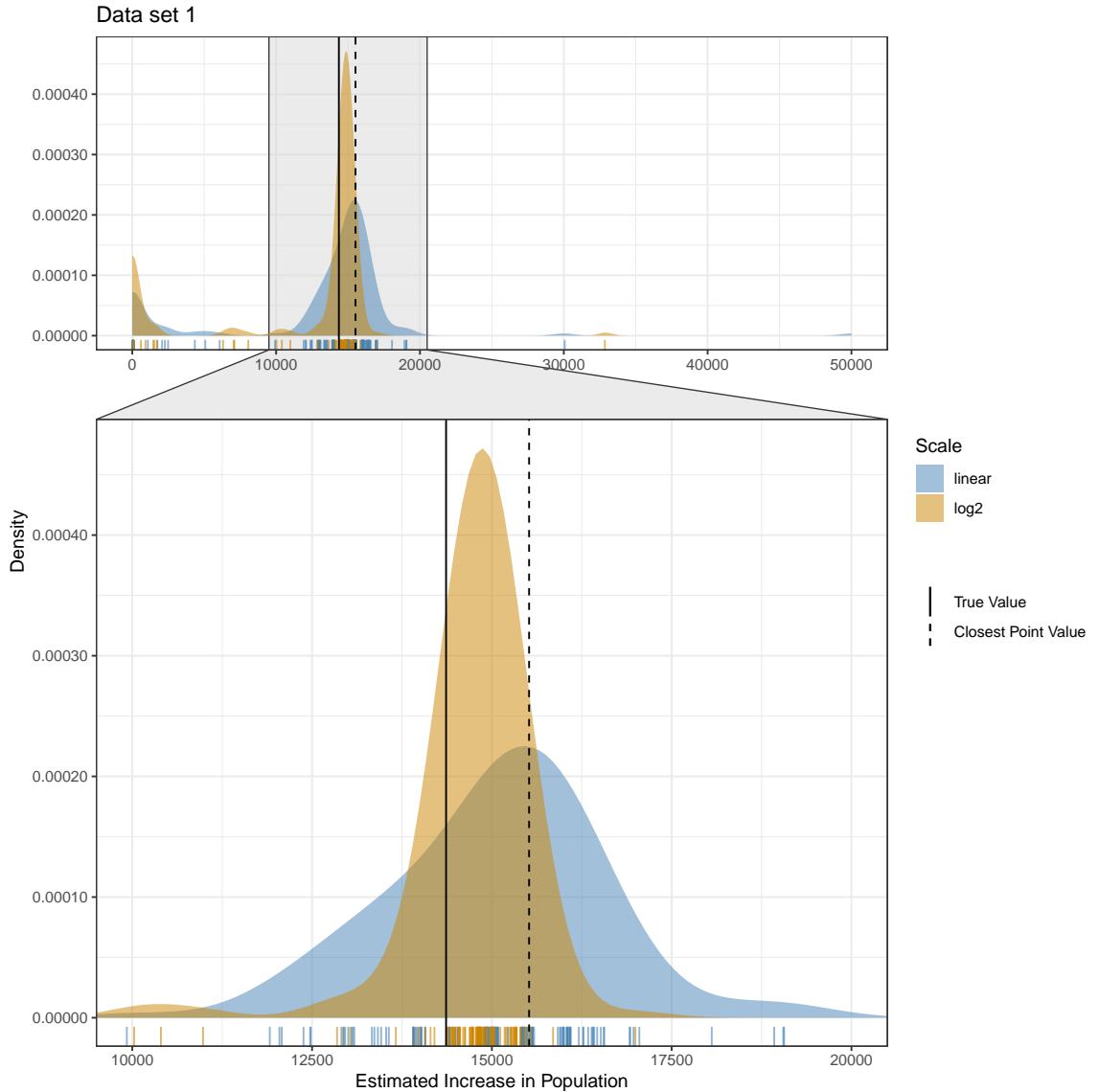


Figure 4.11: Density of the participant estimates for the difference in population between years 20 and 40 for data set 1. Colors are associated to scale - linear (blue) and log (orange) - and vertical lines indicate the true difference based on the underlying model equation (black solid) and closest point difference based on the simulated data set (black dashed). A jittered rug plot along the  $x$ -axis shows where participant estimates were made. The plot shows an improvement in accuracy when estimates are made on the linear scale as opposed to the log scale as indicated by the linear peak at the closest point.

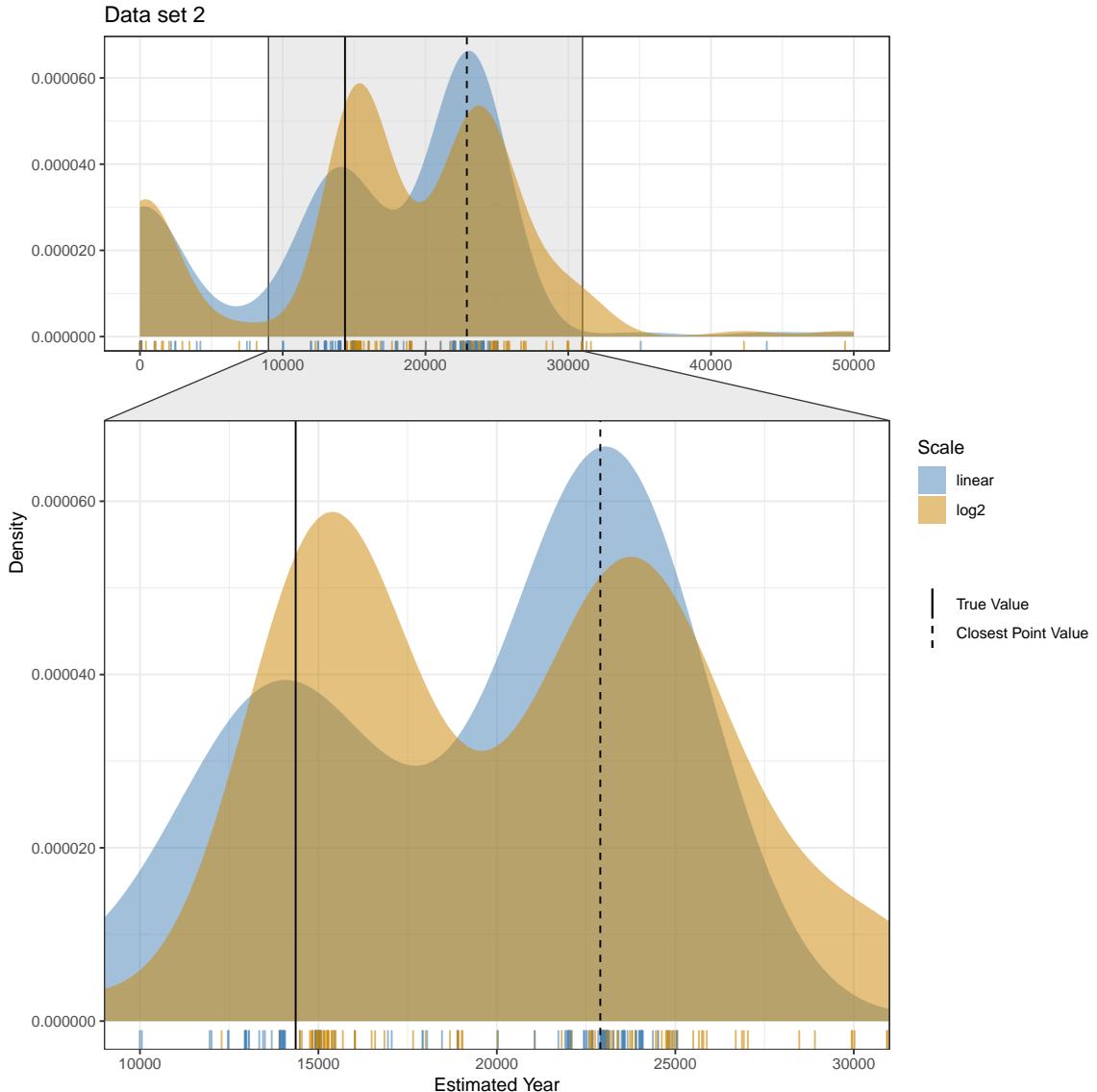


Figure 4.12: Density of the participant estimates for the difference in population between years 20 and 40 for data set 2. Colors are associated to scale - linear (blue) and log (orange) - and vertical lines indicate the true difference based on the underlying model equation (black solid) and closest point difference based on the simulated data set (black dashed). A jittered rug plot along the  $x$ -axis shows where participant estimates were made. The plot shows an improvement in accuracy when estimates are made on the linear scale as opposed to the log scale as indicated by the linear peaks at the true difference and closest point. The two peaks illustrate how participants were reading reading the data points and not only the underlying trend.

participants were misinterpreting the value they were asked to estimate. Sample participant work below shows common incorrect logic on both the linear and log scales:

*Sample work 6: incorrect logic (linear)*

$$24000/2000 = 12$$

*Sample work 7: incorrect logic (log)*

$$16380/1026 = 15.96$$

*Sample work 8: changed logic (log)*

$$2048 - 1024 = 1024$$

$$1024 + 512 = 1536$$

$$16384 - 1536 = 14848$$

$$14848/1536 = 9.67.$$

In particular, sample work 8 shows how the participant first estimated halfway between the log grid lines and correctly subtracted the populations for the two given years before incorrectly changing their logic to divide the two populations. One potential source of misinterpretation of this questions might be the particular order in which participants were asked the questions. For example, if participants were asked to provide an estimated increase in population after having been asked Intermediate Q2 which prompts participants to provide a multiplicative change in

population, they may be more likely to misinterpret Intermediate Q1. However, participants answering questions on the second scenario would have seen both questioning frameworks in the previous scenario context.

Estimates for the increase in population between year 20 and year 40 was distinctly more accurate for estimates made on the linear scale as indicated by the peak density occurring near the closest point and true value vertical lines. The slight shifts in the density on the log scale suggest participants are making inaccurate first level estimates. One explanation might be that participants were anchoring to the grid lines much stronger on the log scale as opposed to being more likely to adjust their estimates between grid lines on the linear scale. Common responses (Fig. 4.13) on the log scale come from anchoring to grid lines ( $16384 - 1024 = 15360$ ), halfway numerically between grid lines ( $16384 - 1536 = 14848$ ;  $24576 - 1536 = 2340$ ), and base ten ( $16384 - 2000 = 14784$ ) while participants on the linear scale anchored to multiples of 500 and 1000. This was dependent on the location of simulated points in relation to the grid lines and lead to an underestimation in difference for data set one and an overestimation in difference for data set two. Variance in estimates appeared to be consistent across both scales for data set two with a smaller variance on the log scale for data set one.

Responses from Intermediate Q1 required participants to use their first level estimates in order to make an additive comparison of populations between two years. Some participants misinterpreted the question, making a multiplicative comparison, thus providing estimates closer to zero. This was supported by examining select participant calculation and scratchpad work. The estimated increase in population was more accurate on the linear scale with the lack of accuracy on the log scale affected by participant resistance to and misunderstanding

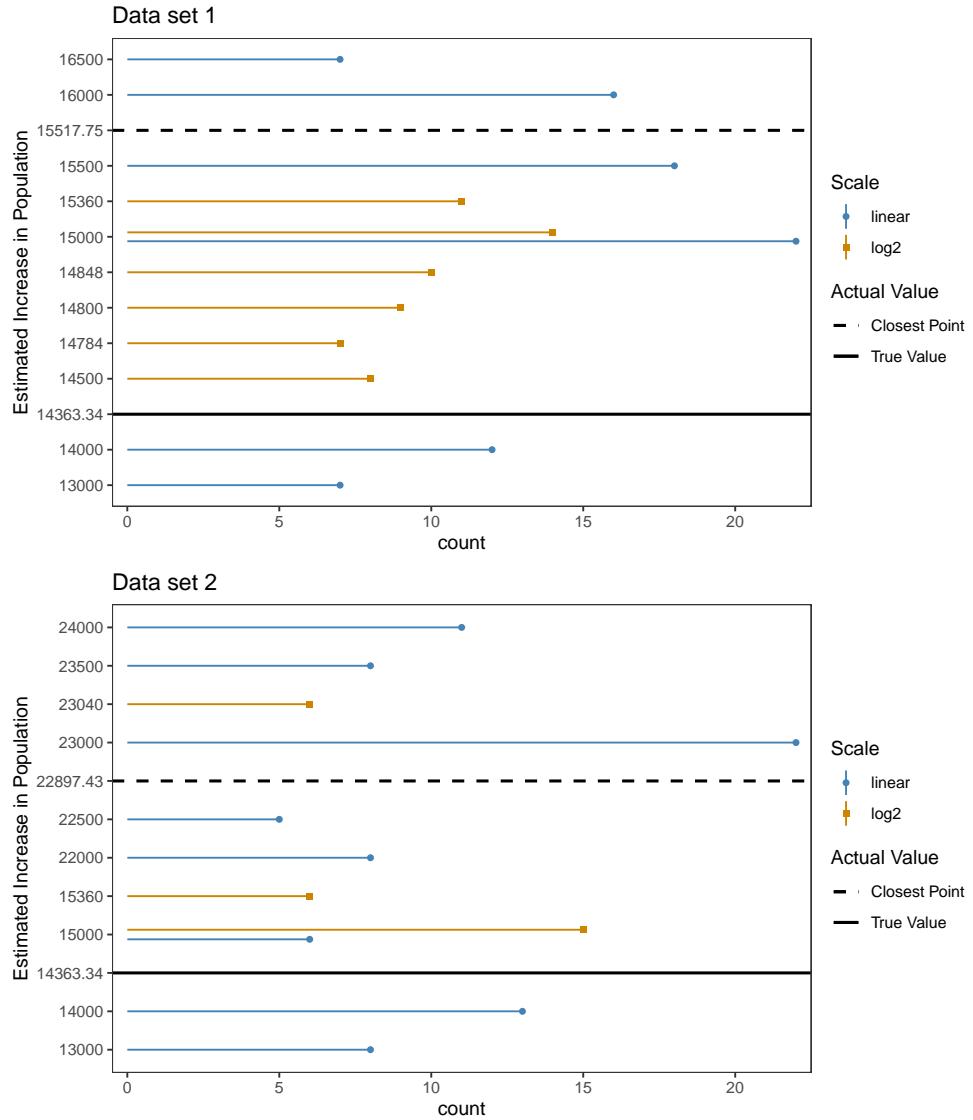


Figure 4.13: The dot charts show estimates for the difference in population between years 20 and 40 provided by more than three participants. The  $x$ -axis indicates the number of participants who provided the estimate marked on the  $y$ -axis in ascending numerical order. Colors are associated to scale - linear (blue) and log (orange) - and horizontal lines indicate the true difference based on the underlying model equation (black solid) and closest point difference based on the simulated data set (black dashed). The two unique data sets are shown separately.

of making estimates between log grid lines.

#### 4.4.5 Intermediate Q2: Multiplicative change in population

Previously, we explored how participants made an additive comparison of populations between two years. In addition, participants were asked, "*How many times more [creatures] are there in 40 than in 20?*" (Fig. 4.14). The questioning was selected carefully to prompt participants to make a multiplicative comparison between two years. Similar to Intermediate Q1, in order to make this comparison, participants must have made accurate first level estimates in both years and then divide the two estimates. Participants may also have made this comparison on the log scale by understanding the multiplicative nature of the grid lines. Sample participant work below shows correct logic on both the linear and log scales:

*Sample work 9: correct logic (linear)*

$$17500/1400 = 12.5$$

Scratchpad: same as before, but a division

*Sample work 10: correct logic (linear)*

$$17000 - 1000 = 16000$$

$$17000/1000 = 17$$

*Sample work 11: correct logic (log)*

$$24/1.4 = 17.14$$

Scratchpad: around 24k tribbles were at 4540, and

1.4k at 4520, make a division and thats

how many times (without the k)

*Sample work 12: correct logic (log)*

$$2048 * 5 = 10240$$

$$2048 * 6 = 12288$$

$$2048 * 7 = 14336$$

$$2048 * 8 = 16384$$

$$2048 * 8 = 16384.$$

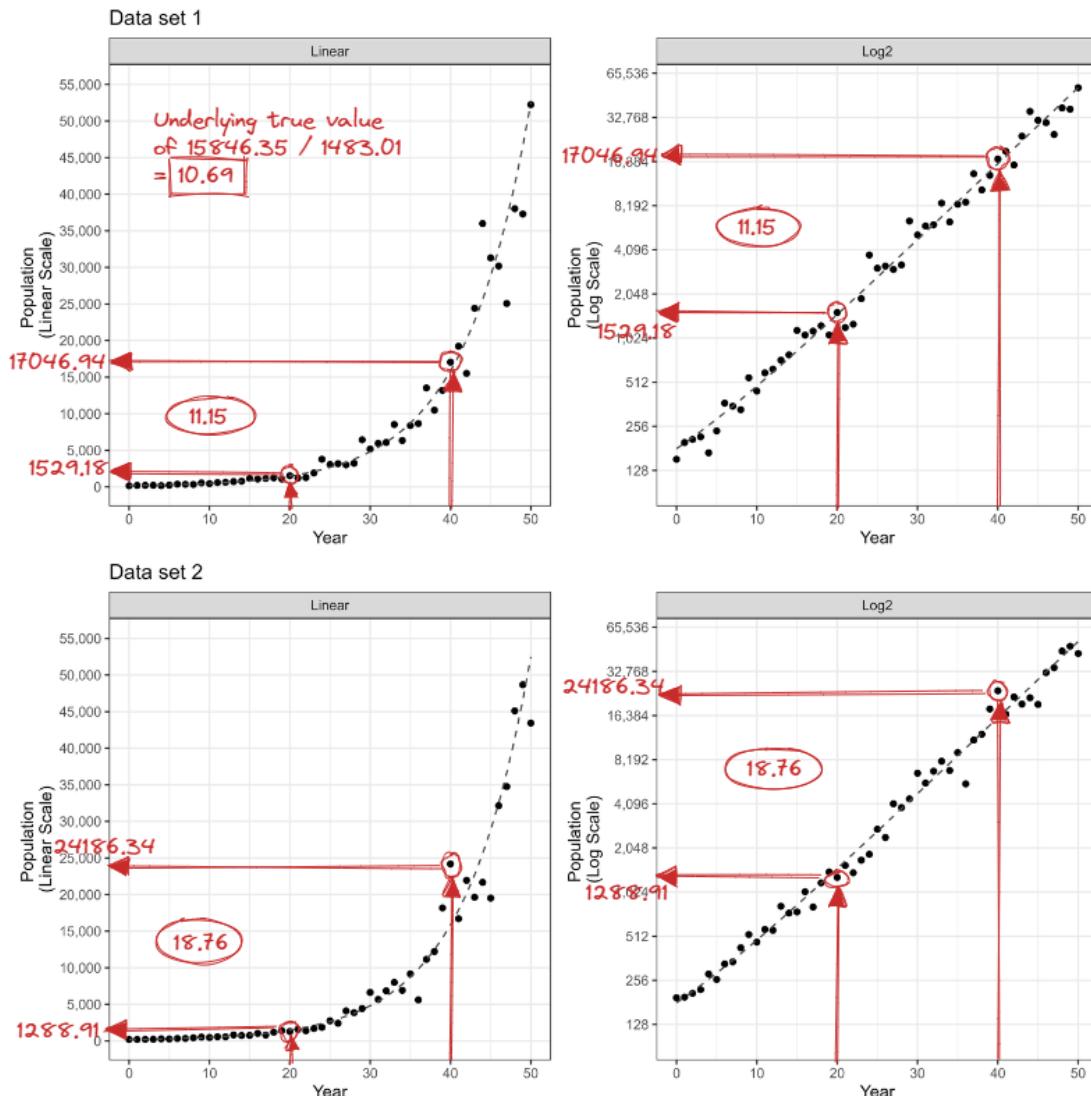


Figure 4.14: Sketch of the estimation procedure asked in Intermediate Q2. Participants make first level population estimates at years 20 and 40, then calculate the ratio between the two values.

The scratch work from participants gave insight about the estimation strategy participants followed when determining the estimated change in population. For instance, sample work 10 shows the participant first incorrectly calculated the additive increase in population before correcting their calculation through division while sample work 12 shows how the participant used a trial and error method. The true change in population based on the underlying equation was 10.69 times as many ( $15846.35/1483.01$ ) with changes based on the closest points of 11.1 ( $17046.94/1529.18$ ) and 18.8 ( $24186.34/1288.91$ ) for data sets one and two respectively. The median estimated change for data set one was 11.7 (IQR = 8.5) for the linear scale and 10.7 (IQR = 6) for the log scale while data set two resulted in larger estimates and variability with a median change of 15.3 (IQR = 14) and 16 (IQR = 8.5) for the linear and log scale respectively. The inconsistency between the two data sets aligns with previous evidence that participants were making estimates by reading the simulated data rather than based on the underlying trend.

As seen in the results for Intermediate Q1, some participants struggled to understand the value they were being asked to estimate. Similarly, Fig. 4.15 illustrates a substantial number of participants provided estimates that more closely reflected that of the additive increase in population rather than the multiplicative change. Fig. 4.17 highlights that 15000 was still a common participant response. Sample work below demonstrate common incorrect logic and calculations conducted by participants:

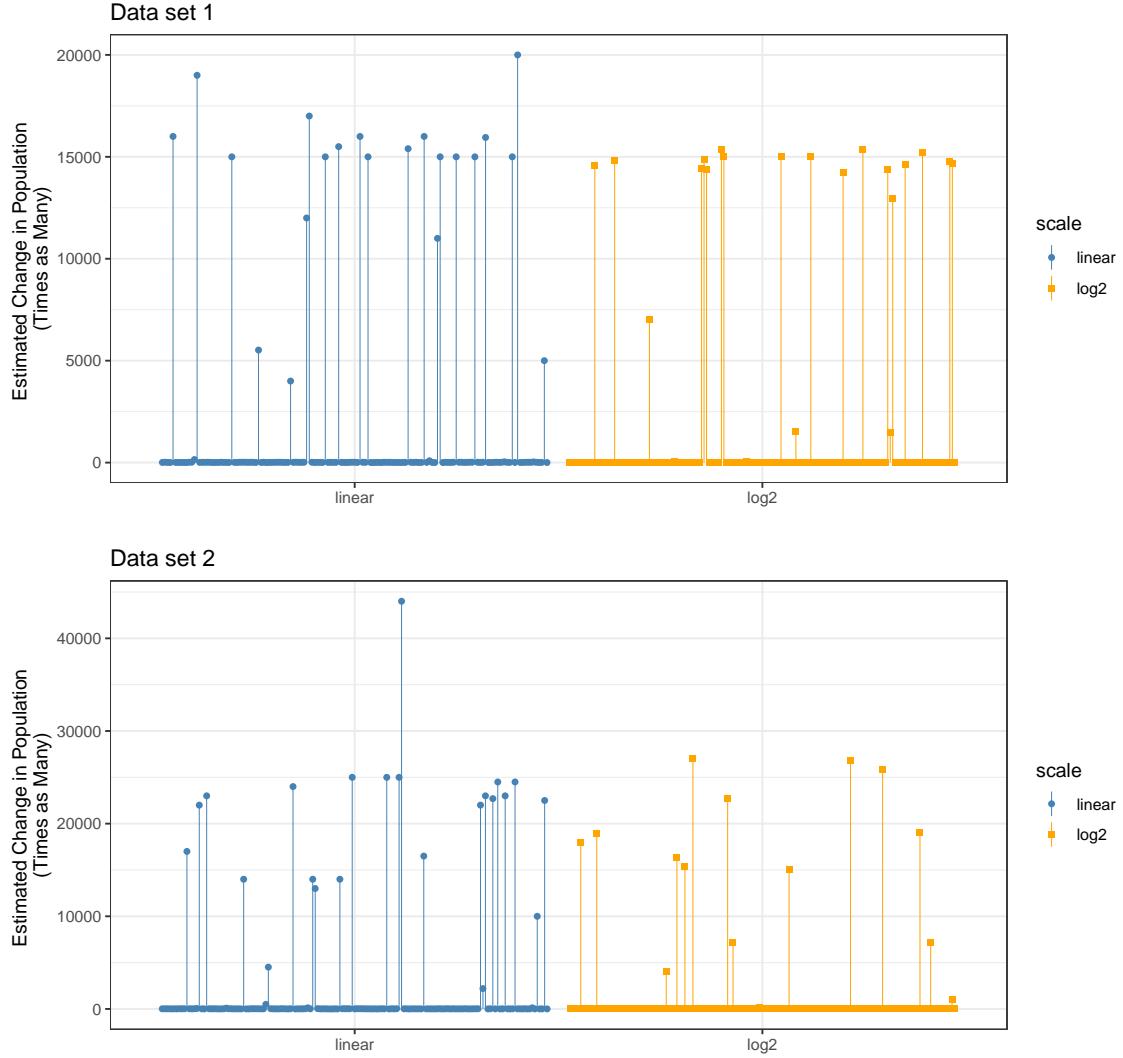


Figure 4.15: Displays the observed estimated change in population for Intermediate Q2. The colors indicate scale - linear (blue) and log (orange) with participants dodged. The plot shows a substantial number of participants provided estimates that more closely reflected that of the additive increase in population rather than the multiplicative change.

*Sample work 13: incorrect logic (linear)*

$$23800 - 1100 = 22700$$

*Sample work 14: incorrect logic (log)*

$$16384 - 1536 = 14848.$$

Evaluating reasonable participant responses for the change in population between 0 times as many and 35 times as many, Fig. 4.16 indicates participants tended to make more accurate and less variable estimates on the log scale than on the linear scale. Fig. 4.15 shows common responses provided by participants.

Overall, responses for Intermediate Q2 provided further support that participants tended to misinterpret the quantity they were being asked to estimate. The density plots of responses suggest that the log scale has a slight advantage over the linear scale for estimating the multiplicative change in population. While anchoring to grid lines and base ten values for first level estimates still occurred, many participants further anchored their responses to whole values.

#### 4.4.6 Intermediate Q3: Time until population doubles

An alternative multiplicative comparison between two points was to determine the amount of time it took for a value to double. Participants were asked, “*How long does it take for the population in year 10 to double?*” (Fig. 4.18). In order to accurately evaluate this comparison, participants must have made a first level estimate for the population in year 10, asked in Elementary Q1. Participants then

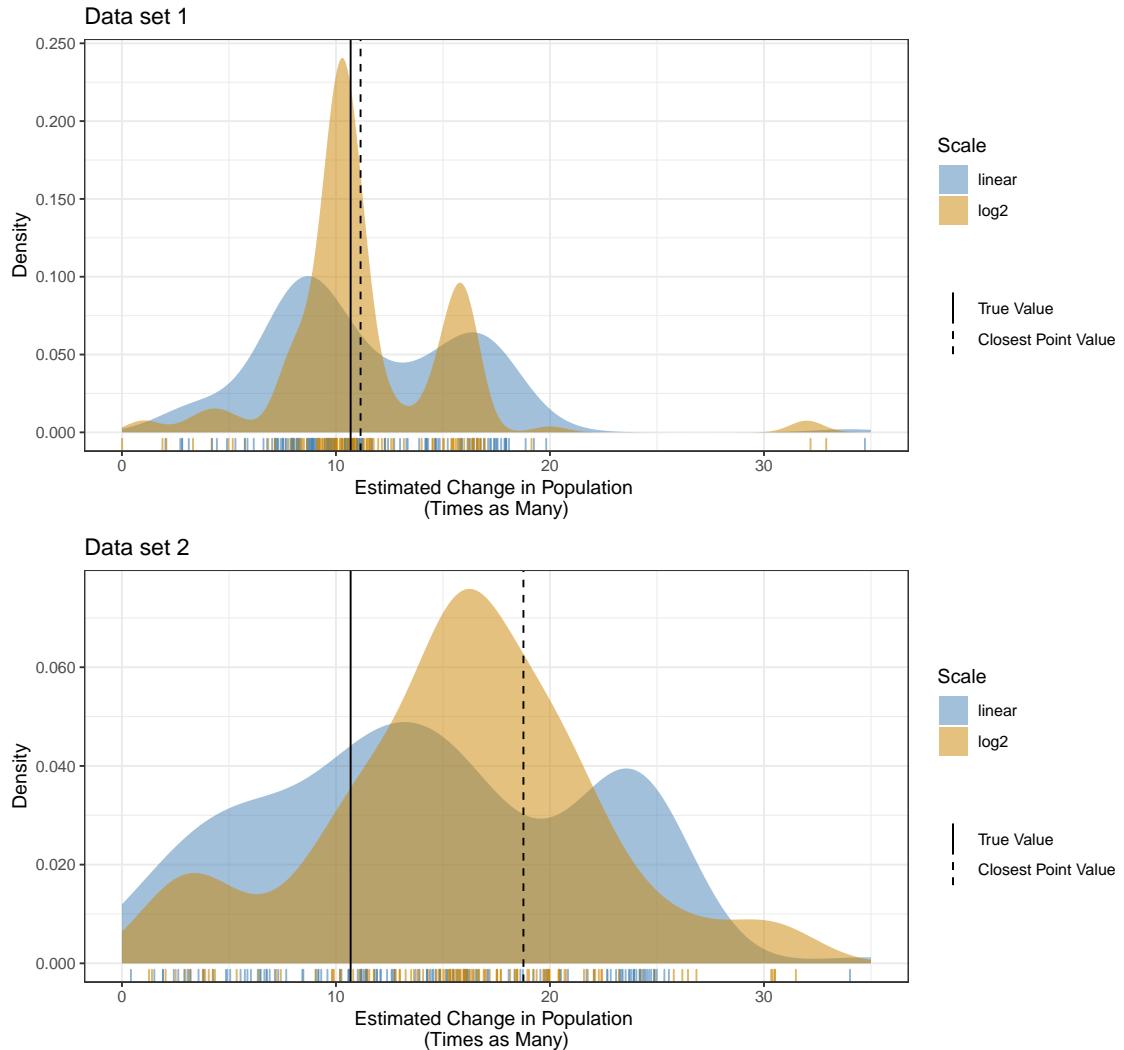


Figure 4.16: Density of the participant estimates for the multiplicative change in population between years 20 and 40. Colors are associated to scale - linear (blue) and log (orange) - and vertical lines indicate the true change based on the underlying model equation (black solid) and closest point change based on the simulated data set (black dashed). A jittered rug plot along the  $x$ -axis shows where participant estimates were made. Data sets are plotted separately. The density plots show participants tended to be make more accurate and less variable estimates on the log scale than on the linear scale.

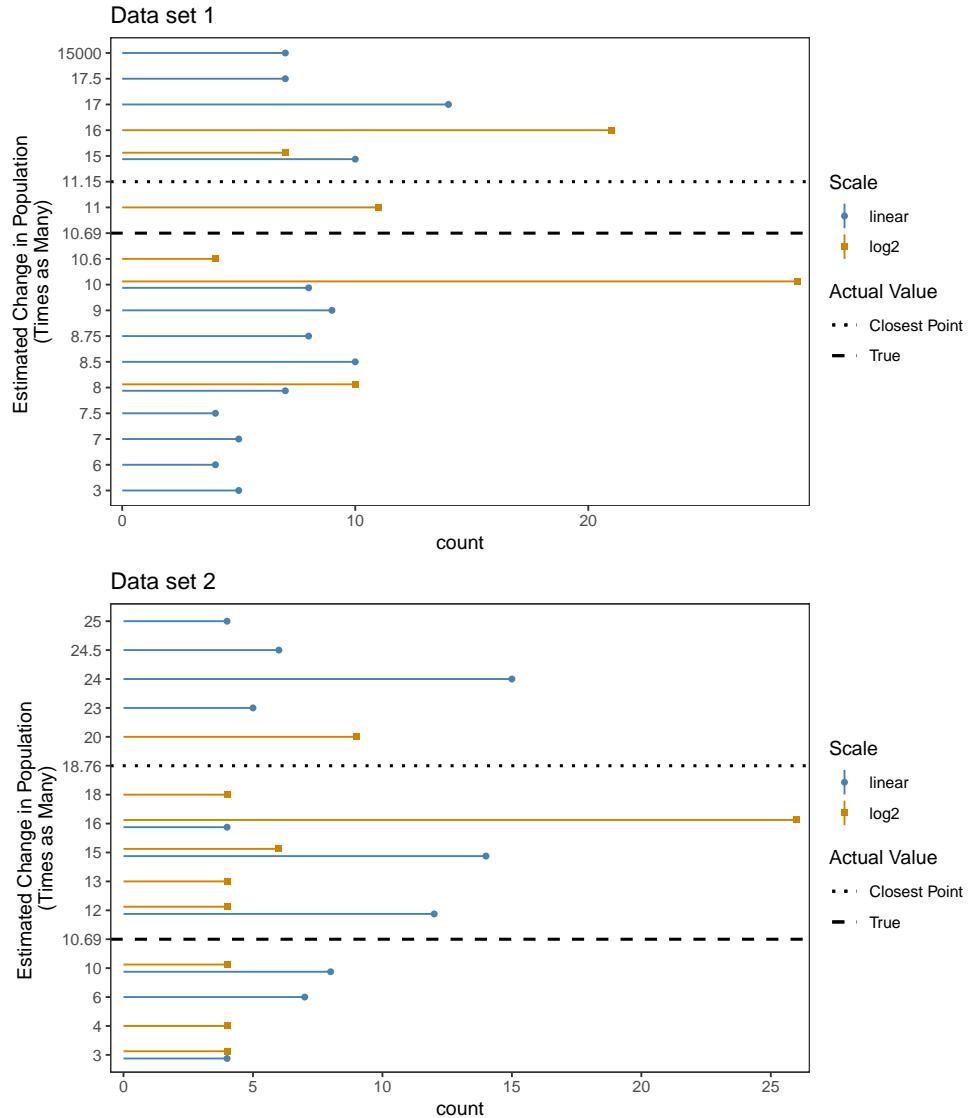


Figure 4.17: The dot charts show estimates for the multiplicative change in population between years 20 and 40 provided by more than three participants. The  $x$ -axis indicates the number of participants who provided the estimate marked on the  $y$ -axis in ascending numerical order. Colors are associated to scale - linear (blue) and log (orange) - and horizontal lines indicate the true change based on the underlying model equation (black solid) and closest point change based on the simulated data set (black dashed). The two unique data sets are shown separately. We see that 15000 is still a common response, demonstrating a common misunderstanding of the value being asked.

needed to double their first level estimate in order to extract the year in which this value occurred and finally subtract year 10. Alternatively, on the log scale, participants could have made this comparison without actually extracting the numeric values and instead relied on their spatial distance equating one increase in grid lines to a double in population. This estimation strategy would have required keen understanding of the log base two scale. Making a judgement based on participant calculations and scratch work, most participants selected the former approach when they estimated the number of years it took for the population to double. One participant stated, “4510 has 512 Tribble Population and to double it it needs to have 1,024 so it would take approximately 5 years. In 4515 they would have double the population.” while another participant indicated “10Aby near to 460.8 so double is 921.6; aprox. 5 years.”

Based on the true underlying equation, the population in year 10 (481.62) doubled to 963.24 in year 16.25, thus it took 6.25 years for the population to double. Closest simulated points resulted in the population doubling in 4 years ( $445 \times 2 = 890.97$  in 14) for data set one and 6 years ( $466.90 \times 2 = 933.79$  in 16) for data set two. The median participant response for data set one was 5 (IQR = 5) and 5 (IQR = 2) for the linear and log scale respectively with a median for data set two of 8 (IQR = 6) and 6 (IQR = 2) for the linear and log scale respectively. Fig. 4.19 illustrates the estimated number of years until the population in year 10 doubled. While there appears to be similar accuracy across both scales, the variance in estimates was considerably smaller for the log scale. The large variance in the linear scale may be explained by the location of reference year 10 which resulted in a population of low magnitude and is visually difficult to estimate. As indicated by peaks in density, there was strong anchoring which occurs at multiples of five.

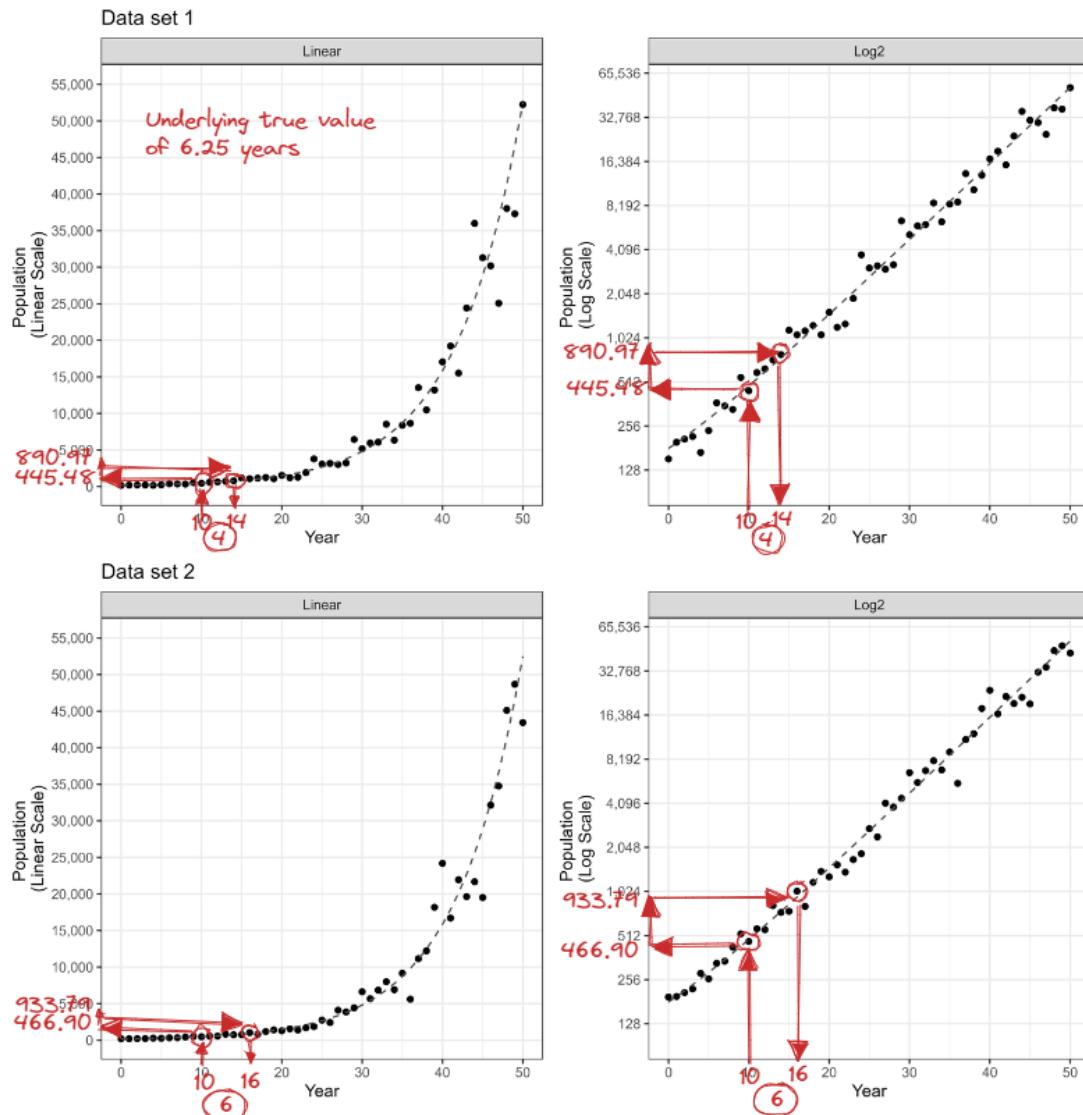


Figure 4.18: Sketch of the estimation procedure asked in Intermediate Q3. Participants make a first level population estimate at year 10, then double this population and estimate the year in which that occurs. Lastly, participants subtracted year 10 to determine how long it took for the population to double. Another estimation strategy could involve participants spatially judging the double of population on the log base 2 scale.

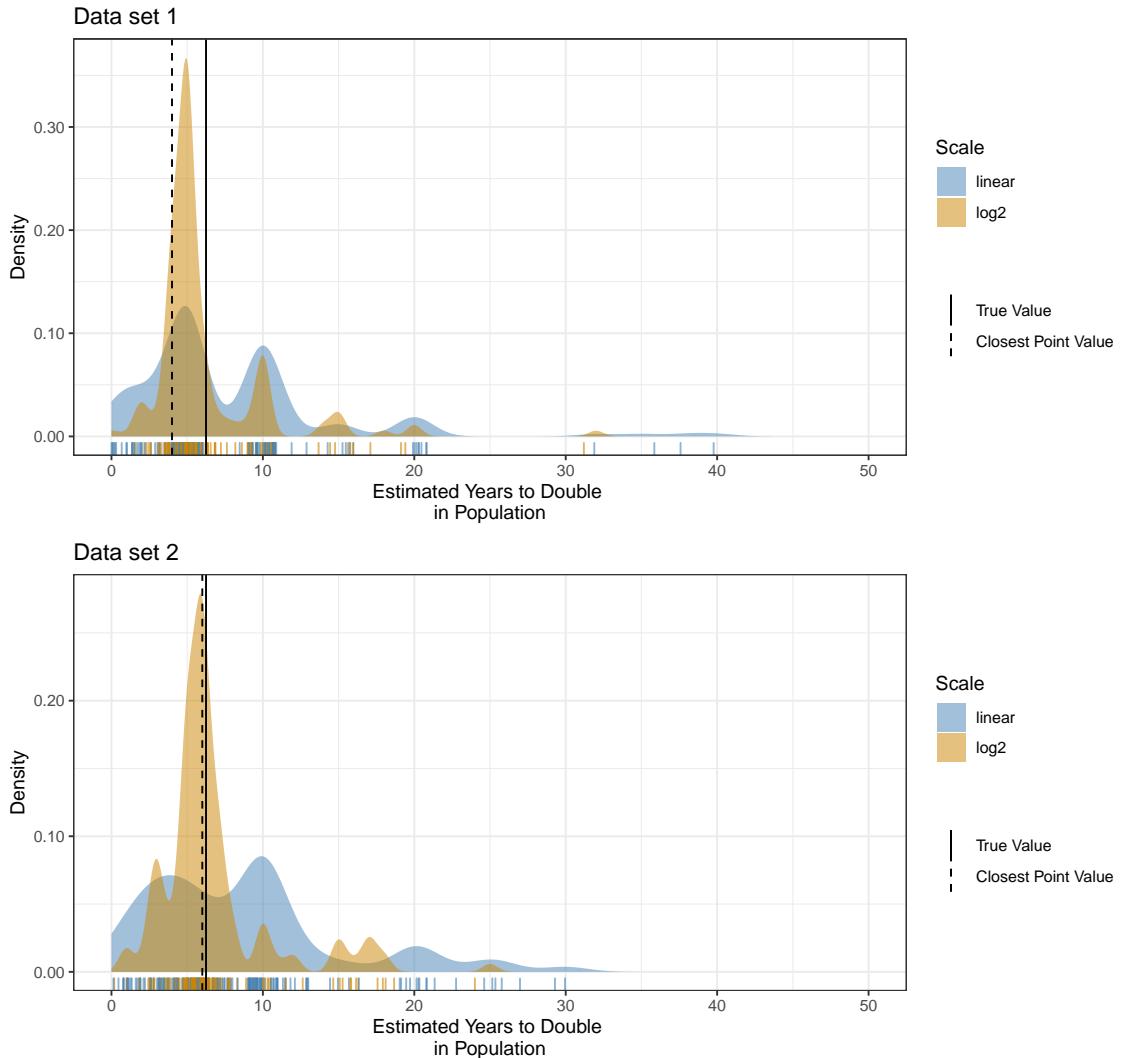


Figure 4.19: Density of the participant estimates for how long it took for the population in year 10 to double. Colors are associated to scale - linear (blue) and log (orange) - and vertical lines indicate the true number of years based on the underlying model equation (black solid) and closest point number of years based on the simulated data set (black dashed). A jittered rug plot along the  $x$ -axis shows where participant estimates were made. The two unique data sets are shown separately. The plot shows larger variability for estimates made on the linear scale.

In summary, as indicated by participant scratch work, they tended to make a first level estimates for the reference year 10 rather than visually judging the distance between grid lines on the log scale. The estimated number of years until the population doubled from a reference year of 10 resulted in lower variability on the log scale as opposed to a larger variability on the linear scale. One explanation might be the low magnitude of population in year 10 and further exploration would be needed to justify for other reference years. Common responses strongly suggest participants were anchoring to multiples of 5 years.

## 4.5 Discussion and Conclusion

This study was intended to help inform and aid in understanding the cognitive implications of displaying exponentially increasing data on a log scale. We evaluated graph comprehension as it relates to two contextual scenarios by asking participants a series of six questions (one open ended, two elementary level, and three intermediate) which required them to quantitatively transform information in the chart. Results provided an understanding of the advantages and disadvantages of the log scale.

In general, results suggest that understanding log logic is difficult as indicated by the misunderstanding in Intermediate Q1 and Q2. This is also supported in following participants estimation strategies for making first level estimates in Intermediate Q3 rather than relying on their spatial awareness between grid lines on the log scale. The accuracy of estimates greatly depends on the location of the value being estimated in relation to the magnitude. For example, accuracy and variability of population estimates made on the linear scale improved as the year of interest increased and thus the magnitude of population increased, making the point more

visible. Alternatively, there was a slight sacrifice in the accuracy of population estimates on the log scale as the year of interest increased. This was due to participant resistance to estimate between grid lines on the log scale and inaccurate representation of equating spatial distance to quantitative difference. As the magnitude of population increases, there was a more noticeable effect of the resistance and lack of understanding. Inaccurate first level estimations can lead to consequences in estimations which require participants to make comparisons between two points. Density plots showed an advantage of the linear scale when estimating an additive increase in population and a slight advantage of the log scale when estimating a multiplicative change in population.

It was also found that estimates were subjective to the simulated data set as shown by the discrepancy between data set one and data set two. This implies a large portion of participants were reading the actual simulated data points as opposed to basing estimates on the underlying visual trend of the data. Common responses revealed participants bias to anchoring their estimates to the grid lines, particularly on the log scale, as well as to base ten values. The strong tendency to anchor to grid lines on the log scale resulted in a sacrifice in accuracy as quantitative differences between grid lines increased. This also implies that accuracy strongly depended on the location of the simulated point in relation to the grid lines.

Understanding graph comprehension is a complex process and requires long-term interaction with the chart and information being presented. With a fixed number of estimates and assessment of participant scratch work, we outlined a variety of situations in which displaying exponentially increasing data on the log scale resulted in both advantages and disadvantages, thus providing a better understanding of the cognitive implications of the use of log scales.

## Chapter 5

### Conclusion

This research evaluated the use of log scales to display exponentially increasing data from three different angles and levels of complexity: perception, prediction, and estimation. Each study provided us insight into the advantages and disadvantages of displaying exponentially increasing data on a log scale and in what context each choice of scale might be more appropriate.

The first study laid the foundation for future studies by testing participants ability to perceptually distinguish between different rates of exponential growth on a linear and log scale. This study utilized statistical lineups and did not require participants to understand exponential growth, identify log scales, or have any mathematical training; instead, any findings rely on participants ability to perceive differences in curvature and slope. An analysis on the accuracy of participants identification of the target plot provides us with insight to the perceptual implications of the choice of scale. Results from the lineup analysis indicate that the perceptual difference results from the contextual appearance of the trend. The choice of scale changes this contextual appearance leading to slight perceptual advantages for both scales depending on the curvatures of the trend lines being compared. In general, when there were large curvature differences, the choice of

scale had no impact and perceptual differences were easily identified on both scales. However, when minor differences in curvature occurred, there was a perceptual advantage for the scale which resulted in identifying the trend which contextually appeared as a curve from the trends that appeared as a line. This revealed an advantage for the log scale when differences were subtle with an exception of the linear scale leading to a perceptual advantage for identifying a trend with more curvature from one which appears to be more linear - contextually appears opposite on the log scale to identify a trend that contextually appears linear from ones that appear curved.

'You Draw It' interactive graphics were adapted from the New York Times and used to test the ability to make forecast predictions for exponentially increasing trends on both scales by drawing a visually fit line with a computer mouse.

The second study required interpretative processes to extend the pattern recognition and construct meaning. This study was supported by a sub-study which validated 'You Draw It' as a tool and method for testing statistical graphics and introduced an appropriate statistical analysis method using generalized additive mixed models for comparing visually fitted trend lines to statistical regression results. This new method was then used to test participants ability to make forecast predictions for exponentially increasing trends on both scales. The results from the analysis showed a clear underestimation of forecasting trends with high exponential growth rates when participants were asked to make predictions on the linear scale. Improvement in forecasts were made when participants were asked to make predictions on the log scale as well as when participants were provided visual aids of points along the trend line. These improvements can be explained by the change in contextual appearance of the data between the two scales; participants visually

extended a linear trend on the log scale and an exponential trend on the linear scale. The graphical tasks from the first two studies were conducted independent of scenarios or contextual applications of log scales; instead, they focused how our visual system perceives and identifies patterns in exponential growth. This did not require participants to understand or read log scales.

In order to understand the cognitive implications of displaying exponentially increasing data on a log scale, the third study evaluated graph comprehension as it relates to the contextual scenario of the data shown. This required participants to integrate information from the scenario, chart title, axes, and graphical forms. In the study, participants were asked to quantitatively transform a graph of exponentially increasing data and extend their estimates to compare two points. We evaluated graph comprehension as it relates to two contextual scenarios by asking participants a series of questions based on the elementary (literally reading the data) and intermediate (reading between the data) level questions. The results of this study inform our understanding of the cognitive implications of displaying exponentially increasing data on a log scale.

Overall, our results suggested that log logic is difficult and that we often misinterpret and miscalculate multiplicative reasoning. However, further investigation is necessary to determine if the misunderstanding occurs due to the ambiguity of language or scale. By collecting information from the calculation and scratchpad inputs, we were able to better understand the strategies used in visual estimation. Participants often made first level estimates when comparing two points rather than relying on their spatial awareness between grid lines on the log scale. They were resistant to make estimates between grid lines (anchored to grid lines) and tended to inaccurately equate spatial distance to numerical difference on the log

scale, sacrificing estimation accuracy for cognitive efficiency; this resistance and misinterpretation leads to greater consequences for larger magnitudes when distances between grid lines is large. Further testing is required to evaluate this impact on log scales of different bases, such as base ten where this spatial distance has a much larger effect than on base two. Extension studies could also provide information about the impact including minor grid lines has on anchoring; for example, ten visually unequal spaced minor grid lines (equal numerical difference) could be added to a log base ten scale to potentially aid in participant estimation. Overall, estimation accuracy for small magnitudes was improved by the use of the log scale, but sacrifices in accuracy on the log scale became apparent as magnitudes increased leading to advantages on the linear scale.

The studies conducted in this research relied on graphical tasks of varying complexity in order to help us understand the perceptual and cognitive advantages and disadvantages of displaying exponentially increasing data on the log scale. Our results showed there are perceptual advantages of the use of log scales due to the change in contextual appearance; however, our understanding of log logic is flawed when translating the information into context. We recommend consideration of both user needs and graph specific tasks when presenting data on the log scale; caution should be taken when interpretation of large magnitudes is required, but advantages may appear when it is necessary to visually identify and interpret small magnitudes on the chart. In addition to this research, further investigation is necessary to expand the use of log scales to data which does not follow an exponential trend, but where log scales are otherwise appropriate - such as polynomial curves spanning multiple orders of magnitude. It is also necessary to evaluate the implications of transforming or displaying the  $x$ -axis on a log scale, as

this work focused on the transformation of the  $y$ -axis. Follow-up studies for this research could provide further insight into the strategies of estimation by providing users the ability to interact with the graph with visual aids such as the arrows shown in sketches from [Chapter 4](#) and how we make decisions based on data displayed on a log scale in order to evaluate the effect of scale on risk adversity. This research stands as a model for conducting an extensive series of graphical tasks on the same type of data and plot in order to gain a comprehensive understanding of both the perceptual and cognitive implications of the design choices.

## Appendix A

### You Draw It Setup with Shiny

Interactive plots for the you draw it study were created using the `r2d3` package and integrating D3 source code with an R shiny application. I conducted all data simulation and processing in R and outputted two data sets - point data and line data - containing (x, y) coordinates corresponding to either a simulated point or fitted value predicted by a statistical model respectively. Then, the `r2d3` package converted the data sets in R to JSON to be interpreted by the D3.js code. I define functions in D3.js to draw the initial plot and set up drawable points for the user drawn line. Drag events in D3.js were utilized to react to observe and react to user input. Shiny Messages were used to communicate the user interaction between the D3 code and the R environment. The plot was then rendered and updated on user interaction into the R shiny application with the `RenderD3` and `d3Output` functions. Parameters for aesthetic design choices were defined in a list of options and `r2d3` passes these to the D3.js code. For instance, I specified the buffer space allowed for the *x* and *y* axes to avoid users anchoring their lines to the axes limits. For D3.js source code, visit GitHub [here](#). Fig. A.1 provides a visual aid of the process of creating the you draw it experimental study in [Chapter 3](#).

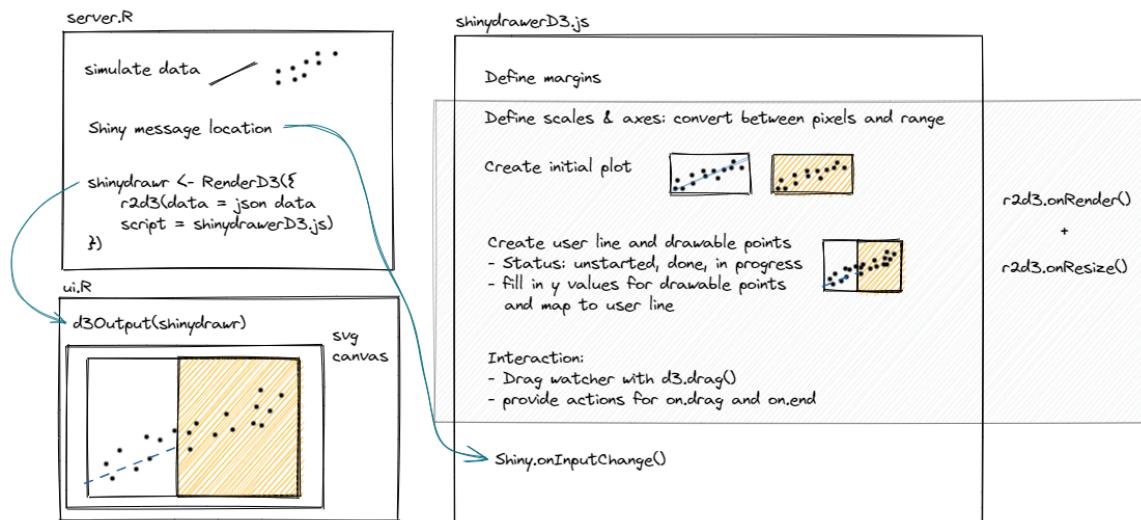


Figure A.1: Interactive plot development

## Appendix B

### Exponential Prediction Interactive Plots

The figures below illustrate the 8 interactive plots used to test exponential prediction. Two data sets were simulated with low and high exponential growth rates and shown four times each by truncating the points shown at both 50% and 75% of the domain as well as on both the log and linear scales following a 2 x 2 x 2 factorial treatment design.

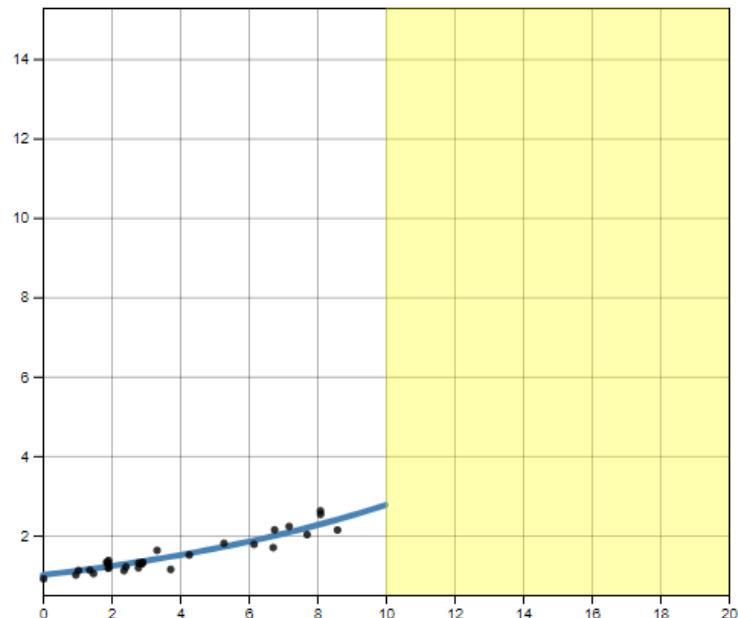


Figure B.1: Exponential Prediction: low growth rate, points truncated at 50%, linear scale

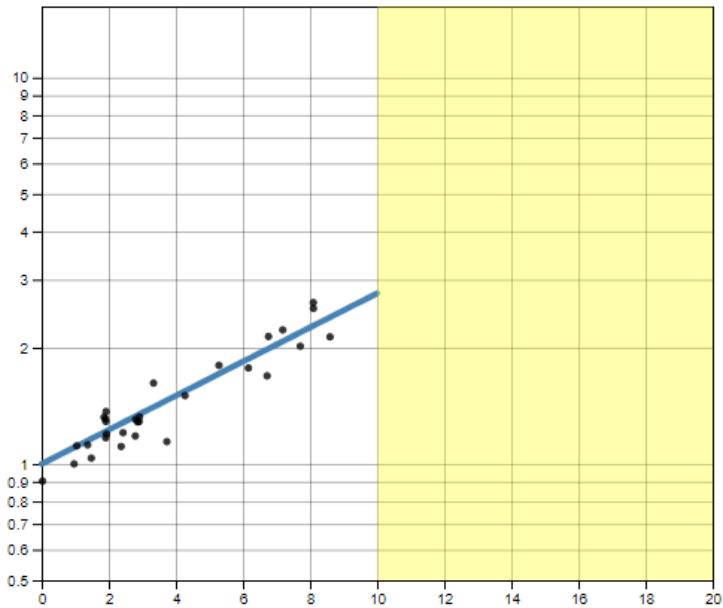


Figure B.2: Exponential Prediction: low growth rate, points truncated at 50%, log scale

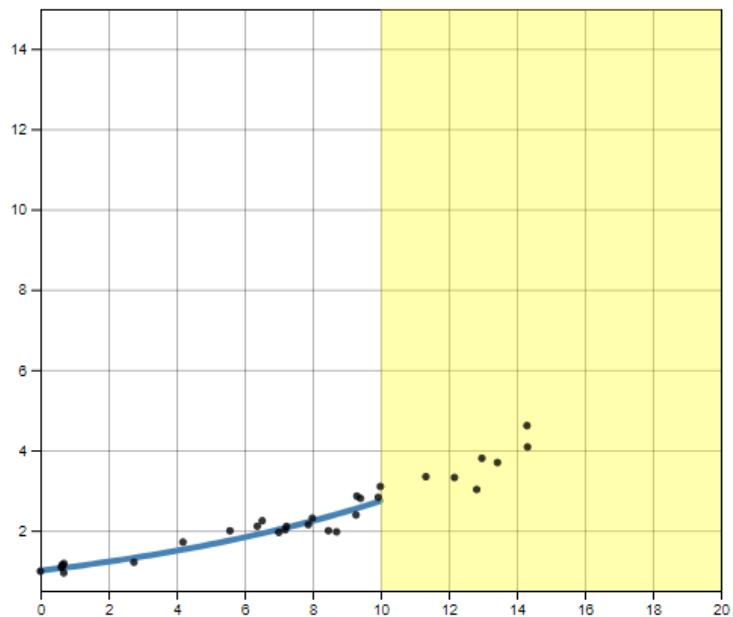


Figure B.3: Exponential Prediction: low growth rate, points truncated at 75%, linear scale

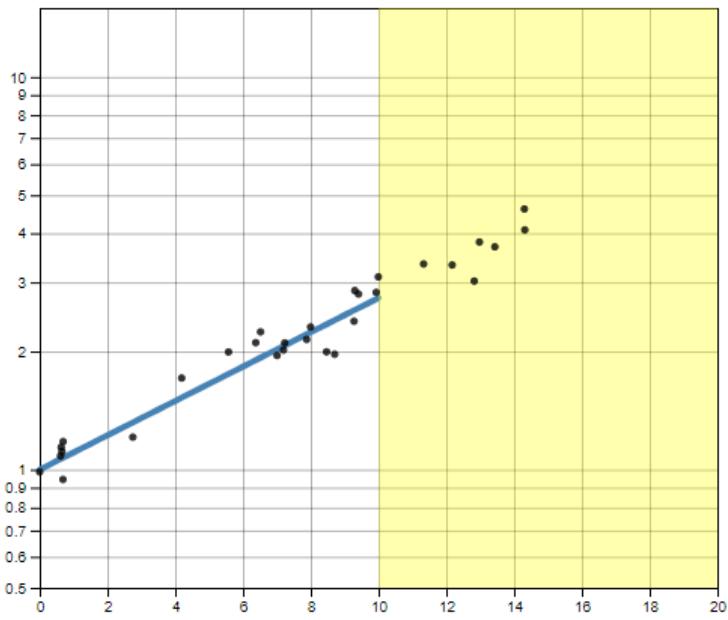


Figure B.4: Exponential Prediction: low growth rate, points truncated at 75%, log scale

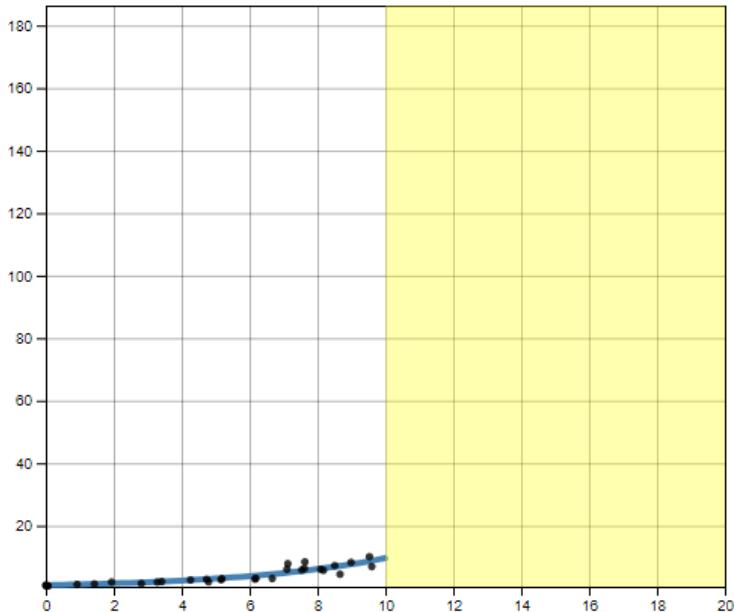


Figure B.5: Exponential Prediction: high growth rate, points truncated at 50%, linear scale

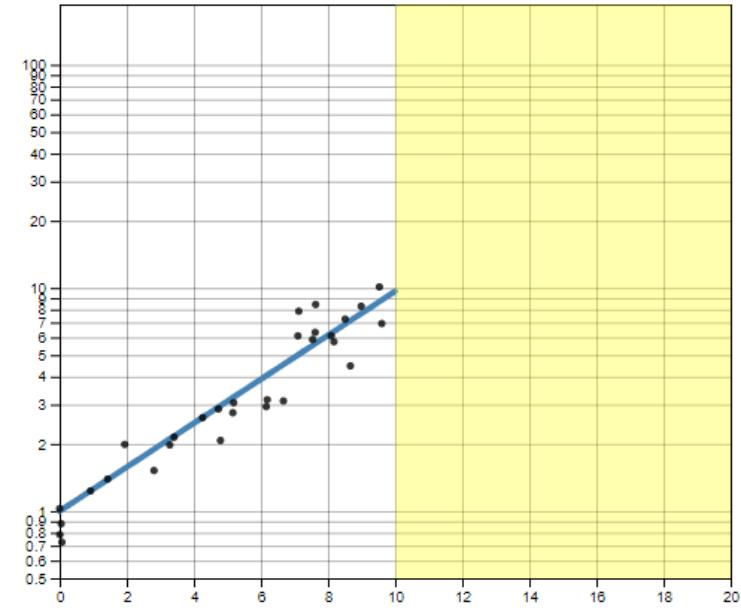


Figure B.6: Exponential Prediction: high growth rate, points truncated at 50%, log scale

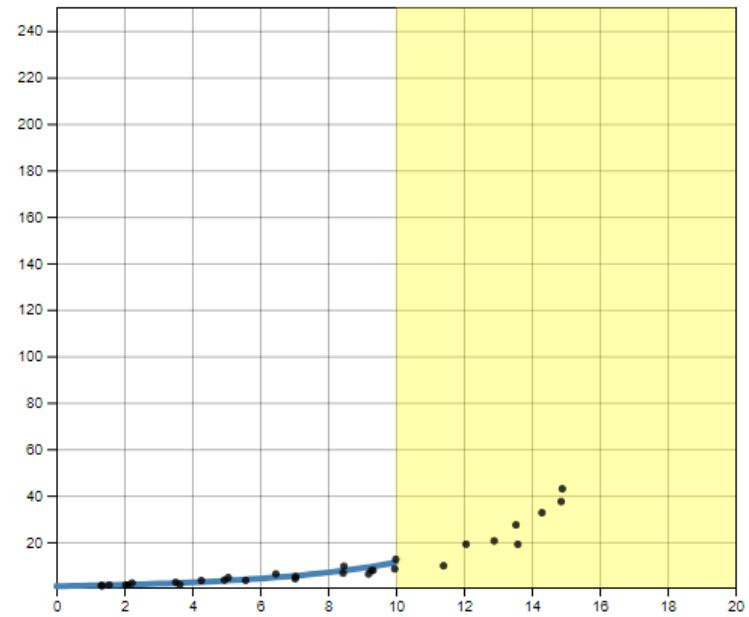


Figure B.7: Exponential Prediction: high growth rate, points truncated at 75%, linear scale

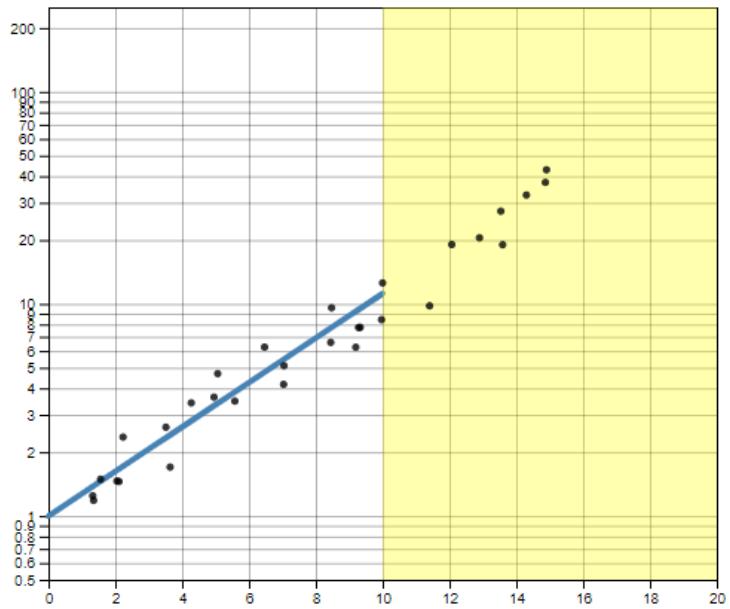


Figure B.8: Exponential Prediction: high growth rate, points truncated at 75%, log scale

## Appendix C

### Scratchwork participant comparison

In order to examine whether participants who used the provided resources for estimation differed in their numerical estimations from those who did not, we first compared population estimates from the explicitly asked year 10 location. About half of the participants fell into the category which provided scratch work and half did not (Table C.1). The was determined there was no substantial difference or bias in estimates between the two groups, therefore, we proceeded to examine the estimated populations across scales from the first level estimates (Fig. C.1). See 4.4.2 for follow-up comparisons.

Table C.1: Estimation showed work summary

Data set	Scale	Showed work	N
dataset1	linear	no	73
dataset1	linear	yes	72
dataset1	log2	no	79
dataset1	log2	yes	78
dataset2	linear	no	79
dataset2	linear	yes	78
dataset2	log2	no	73
dataset2	log2	yes	72

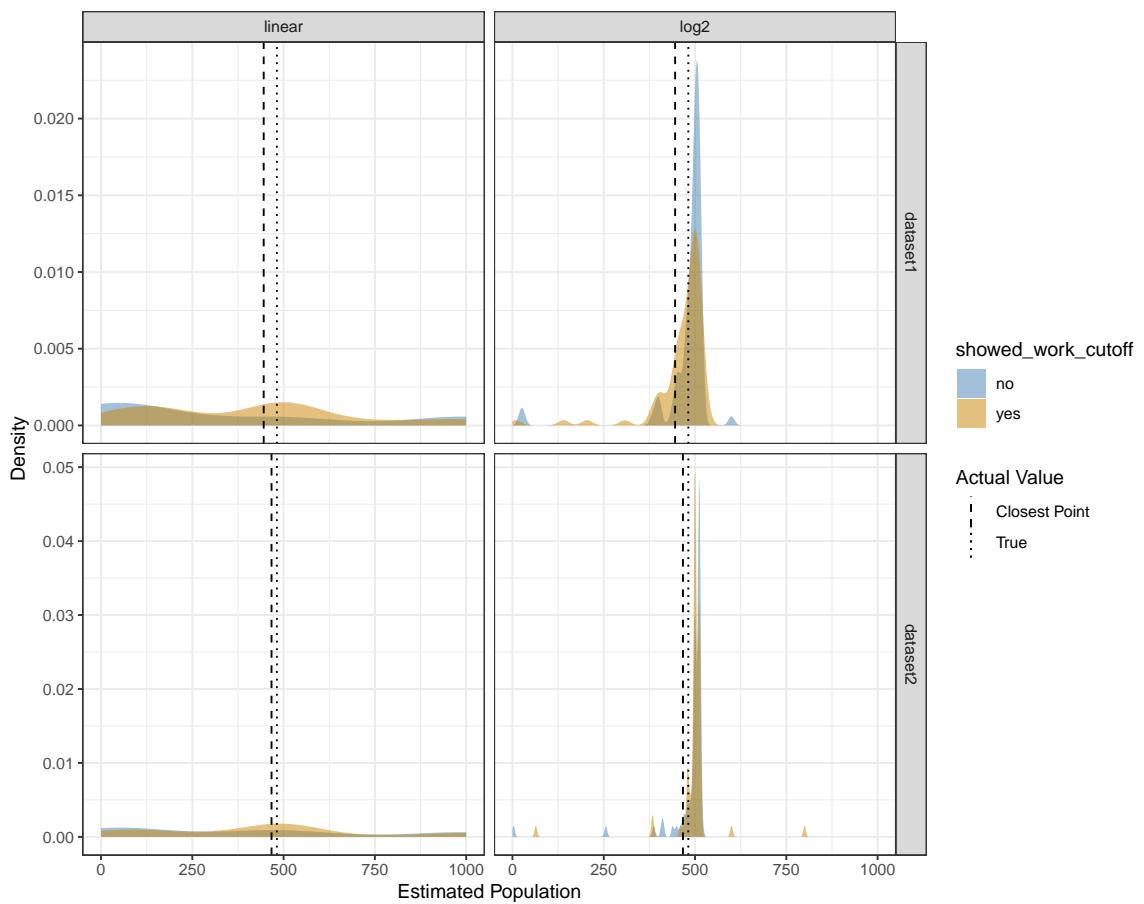


Figure C.1: Estimation showed work density plot comparison

## References

- Aisch, G., Cox, A., & Quealy, K. (2015, May). You draw it: How family income predicts children's college chances. *The New York Times*. The New York Times. Retrieved from  
<https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html>
- Amer, T. (2005). Bias due to visual illusion in the graphical presentation of accounting information. *Journal of Information Systems*, 19(1), 1–18.
- Anderson, V., & McLean, R. (1974). *Design of experiments: A realistic approach*. New York: M. Dekker.
- Andrews, R. (2022, January). Standards for graphic presentation. *Info We Trust*. Retrieved from <https://infowetrust.com/project/1915-standards>
- Barry, D., Buchanan, L., Cargill, C., Daniel, A., Delaquérière, A., Gamio, L., ... al., et. (2020, May). Remembering the 100,000 lives lost to coronavirus in america. *The New York Times*. The New York Times. Retrieved from  
<https://www.nytimes.com/interactive/2020/05/24/us/us-coronavirus-deaths-100000.html>

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*.
- Baumer, B., Kaplan, D., & Horton, N. (2021). *Texts in statistical science: Modern data science with r*. Chapman; Hall/CRC.
- Bavel, J., Baicker, K., Boggio, P., Capraro, V., Cichocka, A., Cikara, M., ... others. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460–471.
- Becker, G., Moore, S., & Lawrence, M. (2019). Trackr: A framework for enhancing discoverability and reproducibility of data visualizations and other artifacts in r. *Journal of Computational and Graphical Statistics*, 28(3), 644–658.
- Beeby, A., & Taylor, H. (1973). How well can we use graphs. *The Communicator of Scientific and Technical Information*, 17, 7–11.
- Best, L., Smith, L., & Stubbs, D. (2007). Perception of linear and nonlinear trends: Using slope and curvature information to make trend discriminations. *Perceptual and Motor Skills*, 104(3), 707–721.
- Best, R., & Boice, J. (2021, June). Build an NBA contender with our roster-shuffling machine. *FiveThirtyEight*. Retrieved from <https://projects.fivethirtyeight.com/nba-trades-2021/>
- Broersma, H., & Molenaar, I. (1985). Graphical perception of distributional aspects of data. *Computational Statistics Quarterly*, 2(1), 53–72.
- Buchanan, L., Park, H., & Pearce, A. (2017, January). You draw it: What got better or worse during obama’s presidency. *The New York Times*. The New

- York Times. Retrieved from <https://www.nytimes.com/interactive/2017/01/15/us/politics/you-draw-obama-legacy.html>
- Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D., & Wickham, H. (2009a). 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.
- Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D., & Wickham, H. (2009b). 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.
- Burn-Murdoch, J., Nevitt, C., Tilford, C., Rininsland, A., Kao, J., 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>
- Carpenter, P., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2), 75.
- Ciccione, L., & Dehaene, S. (2021). Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cognitive Psychology*, 128, 101406.
- Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402), 289–300.

- Cleveland, W. S., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society: Series A (General)*, 150(3), 192–210.
- Cleveland, W., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*. Taylor & Francis.
- Cleveland, W., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716), 828–833.
- Croxton, F., & Stein, H. (1932). Graphic comparisons by bars, squares, circles, and cubes. *Journal of the American Statistical Association*, 27(177), 54–60.
- Croxton, F., & Stryker, R. (1927). Bar charts versus circle diagrams. *Journal of the American Statistical Association*, 22(160), 473–482.
- Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382–393.
- 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.
- Deming, W. (1943). Statistical adjustment of data.
- Dongarra, J. J., Meuer, H. W., Strohmaier, E.others. (1997). TOP500 supercomputer sites. *Supercomputer*, 13, 89–111.
- Dunham, P., & Osborne, A. (1991). Learning how to see: Students graphing difficulties. *Focus on Learning Problems in Mathematics*, 13(4), 35–49.

- 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/>
- Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.
- Fellows, I. (2018). *Wordcloud: Word clouds*. Retrieved from  
<https://CRAN.R-project.org/package=wordcloud>
- Finney, D. (1951). Subjective judgment in statistical analysis: An experimental study. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2), 284–297.
- Finney, D., & Stevens, W. (1948). A table for the calculation of working probits and weights in probit analysis. *Biometrika*, 35(1/2), 191–201.
- Forrester, M., Latham, J., & Shire, B. (1990). Exploring estimation in young primary school children. *Educational Psychology*, 10(4), 283–300.
- Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158.

- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47(2), 183–210.
- Godlonton, S., Hernandez, M., & Murphy, M. (2018). Anchoring bias in recall data: Evidence from central america. *American Journal of Agricultural Economics*, 100(2), 479–501.
- Goldstein, B., & Cacciama, L. (2021). *Sensation and perception*. Cengage Learning.
- Gordon, I., & Finch, S. (2015). Statistician heal thyself: Have we lost the plot? *Journal of Computational and Graphical Statistics*, 24(4), 1210–1229.
- Graesser, A., Swamer, S., Baggett, W., & Sell, M. (2014). New models of deep comprehension. In *Models of understanding text* (pp. 9–40). Psychology Press.
- Green, T., & 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., & Kelley, T. (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.
- Heckler, A., 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). IEEE.
- Heer, J., & Agrawala, M. (2006). Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 701–708.

- Hofmann, H., Follett, L., Majumder, M., & Cook, D. (2012). Graphical tests for power comparison of competing designs. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2441–2448.
- Hogan, T., & Brezinski, K. (2003). Quantitative estimation: One, two, or three abilities? *Mathematical Thinking and Learning*, 5(4), 259–280.
- Horst, A. Artwork by allison horst. Creative Commons. Retrieved from  
<https://github.com/allisonhorst/stats-illustrations>
- Jerne, N., & Wood, E. (1949). The validity and meaning of the results of biological assays. *Biometrics*, 5(4), 273–299.
- Jolliffe, F. (1991). Assessment of the understanding of statistical concepts. In *Proceedings of the third international conference on teaching statistics* (Vol. 1, pp. 461–466).
- Jones, Gregory. (1977). Polynomial perception of exponential growth.
- Jones, Gail, Gardner, G., Taylor, A., Forrester, J., & Andre, T. (2012). Students' accuracy of measurement estimation: Context, units, and logical thinking. *School Science and Mathematics*, 112(3), 171–178.
- Joram, E., Gabriele, A., Bertheau, M., Gelman, R., & Subrahmanyam, K. (2005). Children's use of the reference point strategy for measurement estimation. *Journal for Research in Mathematics Education*, 36(1), 4–23.
- Katz, J. (2017, April). You draw it: Just how bad is the drug overdose epidemic? *The New York Times*. The New York Times. Retrieved from  
<https://www.nytimes.com/interactive/2017/04/14/upshot/drug-overdose-epidemic-you-draw-it.html>

Kruskal, W. (1975). Visions of maps and graphs. In *Proceedings of the international symposium on computer-assisted cartography, auto-carto II* (pp. 27–36). Citeseer.

Lauren. (2018, February). The worst graphs of 2017. *Dolphins*. Animalz. Retrieved from

<https://getdolphins.com/blog/the-worst-graphs-of-2017/>

Leinhardt, G., Zaslavsky, O., & Stein, M. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, 60(1), 1–64.

Lenth, R. (2021). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved from <https://CRAN.R-project.org/package=emmeans>

Lewandowsky, S., & Spence, I. (1989). The perception of statistical graphs. *Sociological Methods & Research*, 18(2-3), 200–242.

Loy, A., Follett, L., & Hofmann, H. (2016). Variations of q-q plots: The power of our eyes! *The American Statistician*, 70(2), 202–214.

Loy, A., Hofmann, H., & Cook, D. (2017). Model choice and diagnostics for linear mixed-effects models using statistics on street corners. *Journal of Computational and Graphical Statistics*, 26(3), 478–492.

Luraschi, J., & Allaire, J. (2018). *r2d3: Interface to 'D3' visualizations*. Retrieved from <https://CRAN.R-project.org/package=r2d3>

Macdonald-Ross, M. (1977). How numbers are shown. *AV Communication Review*, 25(4), 359–409.

- MacKinnon, A., & Wearing, A. (1991). Feedback and the forecasting of exponential change. *Acta Psychologica*, 76(2), 177–191.
- 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.
- Menge, D., MacPherson, A., Bytnarowicz, T., Quebbeman, A., Schwartz, N., Taylor, B., & Wolf, A. (2018). Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature Ecology & Evolution*, 2(9), 1393–1402.
- Mosteller, F., Siegel, A., Trapido, E., & Youtz, C. (1981). Eye fitting straight lines. *The American Statistician*, 35(3), 150–152.
- Myers, R. (1954). Accuracy of age reporting in the 1950 united states census. *Journal of the American Statistical Association*, 49(268), 826–831.
- Nieder, A., & Miller, E. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1), 149–157.
- Perry, P. (2021). *Corpus: Text corpus analysis*. Retrieved from <https://CRAN.R-project.org/package=corpus>
- Peterson, L., & Schramm, W. (1954). How accurately are different kinds of graphs read? *Audio Visual Communication Review*, 178–189.
- Reddit: r/CrappyDesign. (n.d.). Retrieved from [https://www.reddit.com/r/CrappyDesign/comments/ast3q6/this\\_map\\_about\\_pet\\_repartition\\_in\\_europe/](https://www.reddit.com/r/CrappyDesign/comments/ast3q6/this_map_about_pet_repartition_in_europe/)

- Reddit: r/dataisugly. (n.d.). Retrieved from  
<https://www.reddit.com/r/dataisugly/comments/tgn4vq/argh/>
- Risk levels. (2021, March). *Global Epidemics*. Brown School of Public Health.  
Retrieved from  
<https://globalepidemics.org/key-metrics-for-covid-suppression/>
- Romano, A., Sotis, C., Dominion, G., & Guidi, S. (2020). The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. *Health Economics, 29*(11), 1482–1494.
- Rost, L. C. (2020, December). You've informed the public with visualizations about the coronavirus. Thank you. *Datawrapper Blog*. Retrieved from  
<https://blog.datawrapper.de/coronavirus-data-visualization-effect-datawrapper/>
- Schneeweiss, H., Komlos, J., & Ahmad, A. (2010). Symmetric and asymmetric rounding: A review and some new results. *AStA Advances in Statistical Analysis, 94*(3), 247–271.
- Schuetzenmeister, A., & Model, F. (2021). *Mcr: Method comparison regression*. Retrieved from <https://CRAN.R-project.org/package=mcr>
- Shah, P., Mayer, R., & 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., & Braithwaite, D. (2017). Numerical development. *Annual Review of Psychology, 68*, 187–213.

- Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in r. *Journal of Open Source Software*, 1(3), 37.
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 683.
- Star trek: The original series. "The trouble with tribbles". Season 2. Episode 15.* (1967). Broadcast on NBC.
- Star wars: Episode IV – a new hope.* (1977). 20th Century Fox.
- Star wars: Episode VI - return of the jedi.* (1983). 20th Century Fox.
- Stewart-Lowndes, J., Best, B., Scarborough, C., Afflerbach, J., Frazier, M., O'Hara, C., ... Halpern, B. (2017). Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*, 1(6), 1–7.
- Szafir, D. (2018). The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them). *Interactions*, 25(4), 26–33.
- Tan, J. (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.
- Tan, J., & Benbasat, I. (1990). Processing of graphical information: A decomposition taxonomy to match data extraction tasks and graphical representations. *Information Systems Research*, 416–439.
- Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, 78(3), 392–402.

- Tol. (2021). Bitmap VS SVG. Retrieved from  
[https://commons.wikimedia.org/wiki/File:Bitmap\\_VS\\_SVG.svg](https://commons.wikimedia.org/wiki/File:Bitmap_VS_SVG.svg)
- Tory, M., & Moller, T. (2004). Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1), 72–84.
- Tufte, E. R. (1985). The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3), 15.
- Twitter: @GraphCrimes. (2022, March). The US states that are the best at solving WORDLE. Retrieved from  
<https://twitter.com/GraphCrimes/status/1500314716081692676>
- Vanderplas, S., Cook, D., & Hofmann, H. (2020). Testing statistical charts: What makes a good graph? *Annual Review of Statistics and Its Application*, 7, 61–88.
- VanderPlas, S., & Hofmann, H. (2015). Spatial reasoning and data displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 459–468.
- VanderPlas, S., & Hofmann, H. (2017). Clusters beat trend!? Testing feature hierarchy in statistical graphics. *Journal of Computational and Graphical Statistics*, 26(2), 231–242.
- Varshney, L., & Sun, J. (2013). Why do we perceive logarithmically? *Significance*, 10(1), 28–31.
- Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22(2), 219–240.
- Villa, M. (2022, February). The rouble collapse against US dollar. Twitter. Retrieved from

<https://twitter.com/emmevilla/status/1498365717720223751>

Von Bergmann, J. (2021, March). Xkcd\_exponential: Public health vs scientists.

*mountainMath*. GitHub. Retrieved from

[https://github.com/mountainMath/xkcd\\_exponential](https://github.com/mountainMath/xkcd_exponential)

Von Huhn, R. (1927). Further studies in the graphic use of circles and bars: I: A discussion of the eells' experiment. *Journal of the American Statistical Association*, 22(157), 31–36.

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

Wagenaar, W., & Sagaria, S. (1975). Misperception of exponential growth. *Perception & Psychophysics*, 18(6), 416–422.

Walker, F. (1874). Statistical atlas of the united states based on the results of the ninth census 1870: With contributions from many eminent men of science and several departments of the government comp. Under the authority of congress by francis a. walker.

Wickham, H. (2013). Graphical criticism: Some historical notes. *Journal of Computational and Graphical Statistics*, 22(1), 38–44.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Wickham, H., Cook, D., Hofmann, H., & Buja, A. (2010). Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 973–979.

Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data.* " O'Reilly Media, Inc.".

- Wilkinson, L. (2013). *The grammar of graphics*. Springer Science & Business Media.
- Wood, R. (1968). Objectives in the teaching of mathematics. *Educational Research*, 10(2), 83–98.
- Wood, S. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1), 95–114.
- Wood, S. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467), 673–686.
- Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1), 3–36.
- Wood, S. (2017). *Generalized additive models: An introduction with r* (2nd ed.). Chapman; Hall/CRC.
- Wood, S., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563.