How *not* to lie: handling missing data and uncertainty through visualization

Ayush Noori and Aneesh Muppidi

CS 271: Topics in Data Visualization

Harvard School of Engineering and Applied Sciences

@ayushnoori @aneesh-muppidi0319

OUTLINE

Line graphs and bias?

The visual weight of a line depends on its arc length, drawing attention.

Examining variability overweighting

Is there bias in perceptual line graph average estimation?

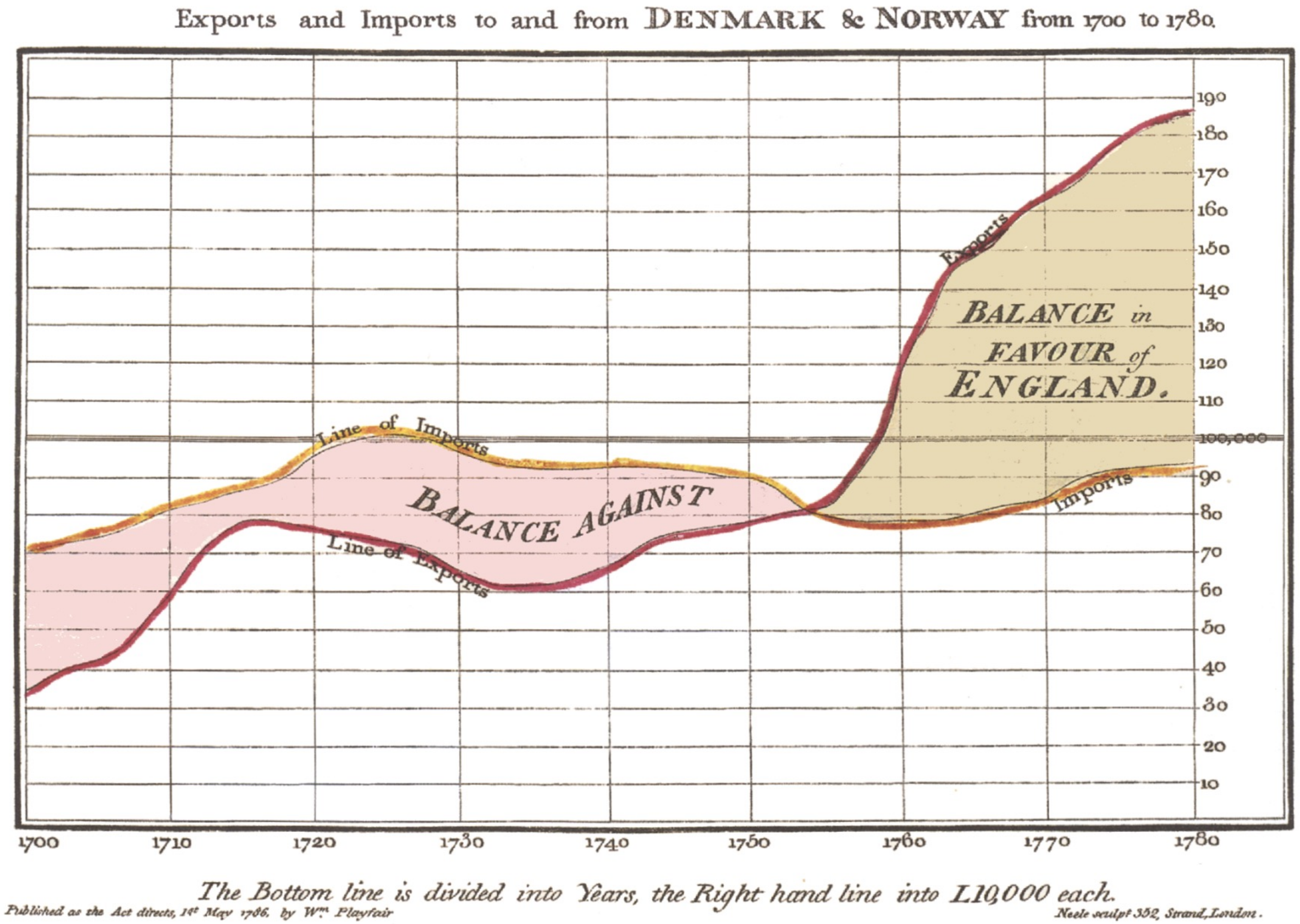
Influence of mark type

Can the degree of variability overweighting be influenced by mark type?

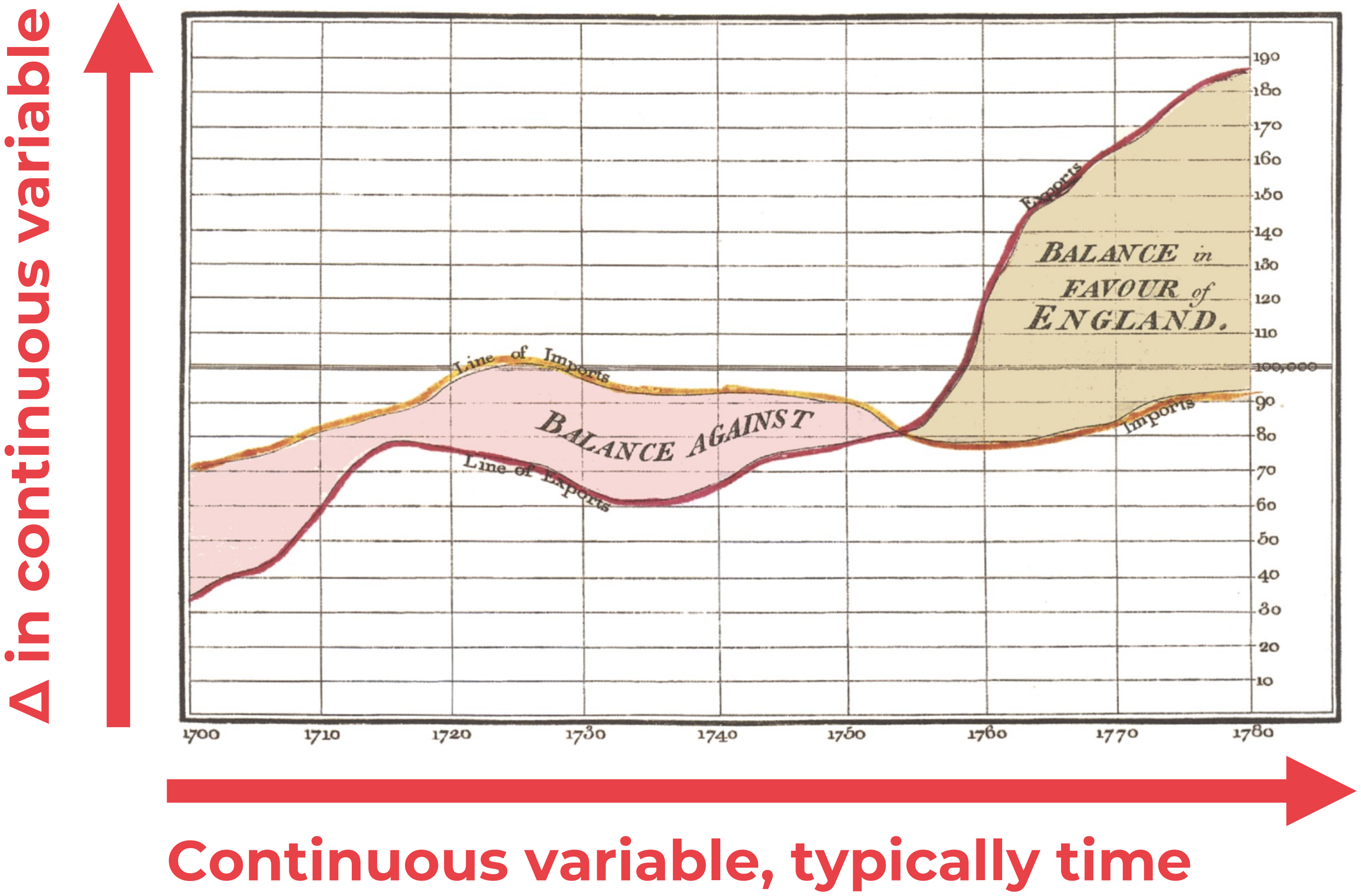
How *not* to lie with visualization

Evaluation in large patient cohorts and patient-derived brain organoids.

Line graphs show changes in two continuous variables and are a common visualization type.



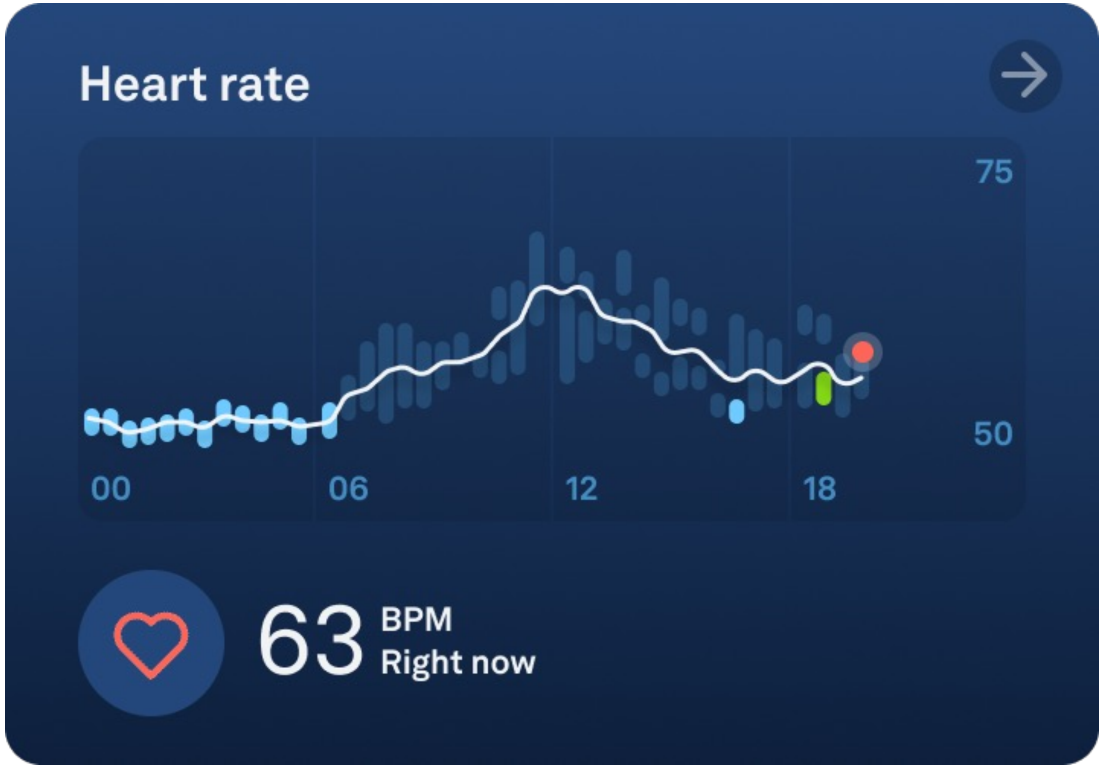
Line graphs show changes in two continuous variables and are a common visualization type.



William Playfair, Commercial and Political Atlas, 1786

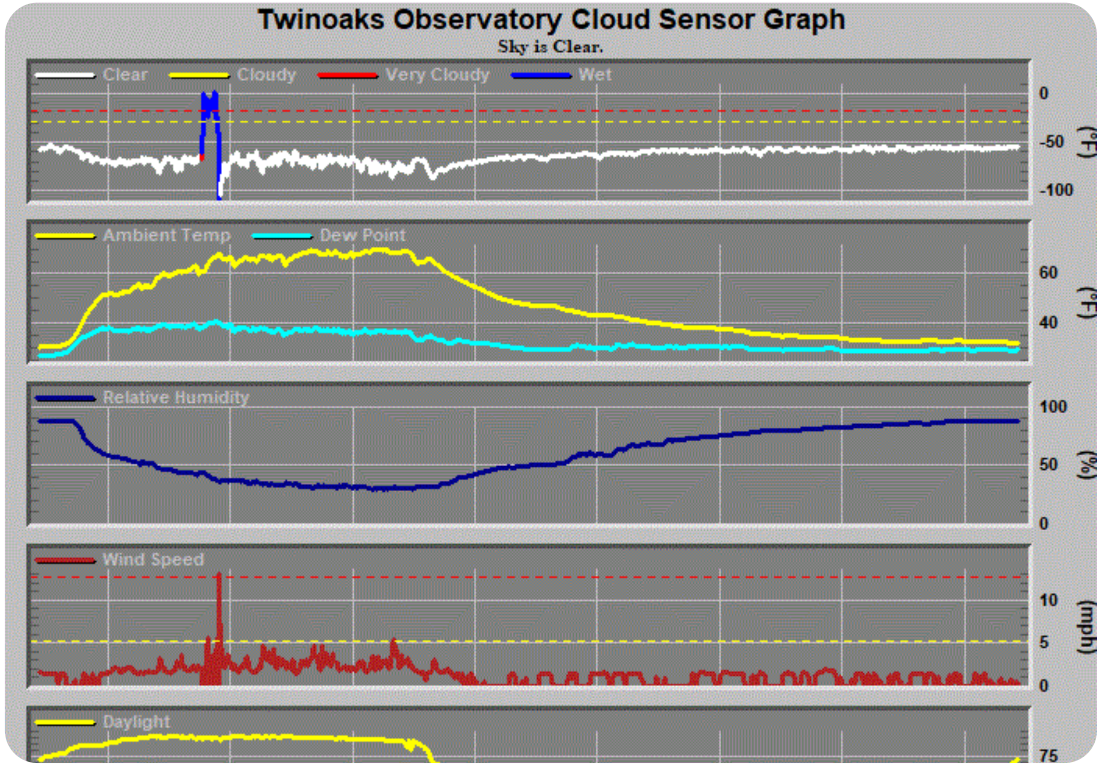
Line graphs are used to visualize time series data, e.g., stocks, sensor data, vitals, etc.

STOCKS

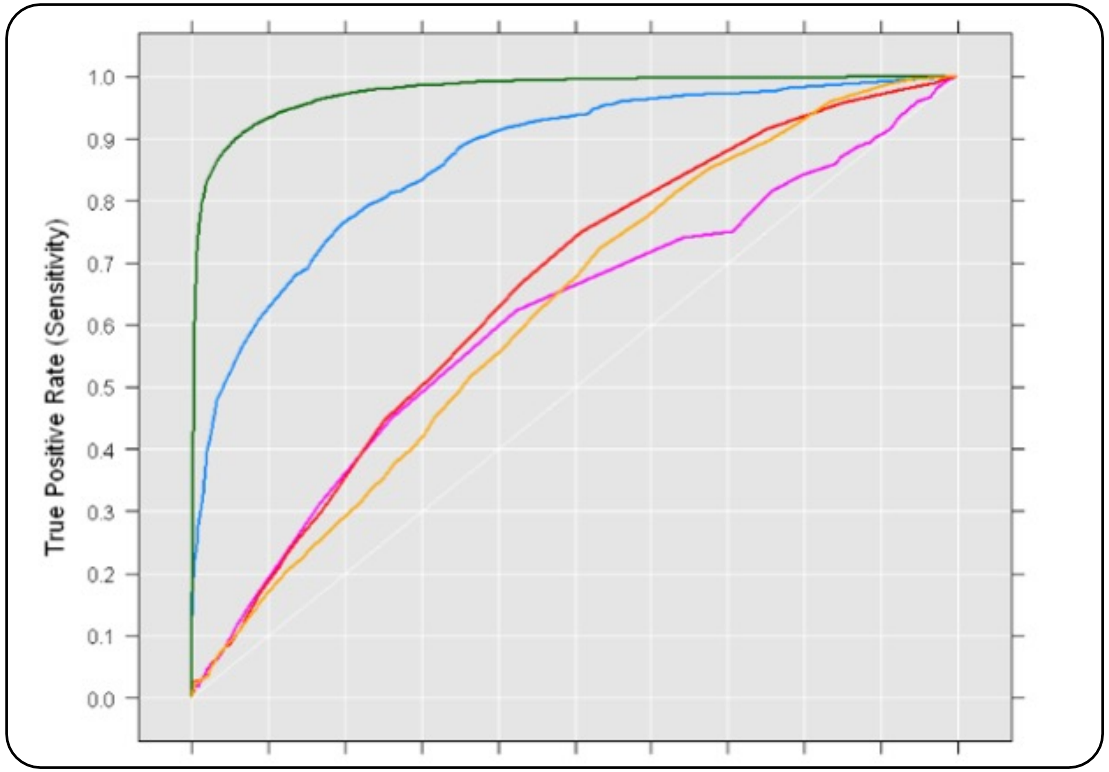


HUMAN VITALS

Line graphs are used to visualize time series data, e.g., stocks, sensor data, vitals, etc.

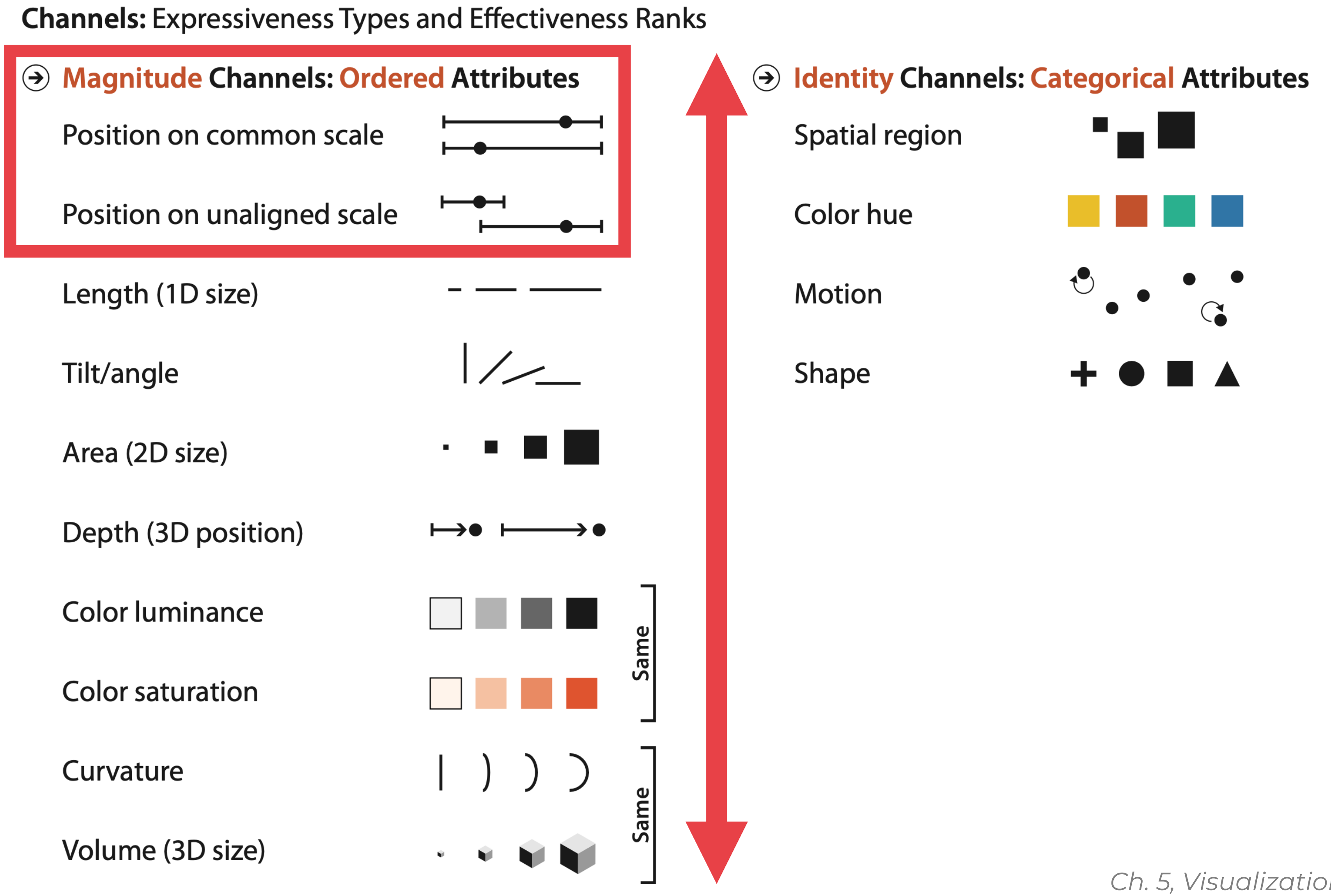


SENSOR DATA



MACHINE LEARNING

Visual encoding of time series as position is considered effective vs. hue, shape, etc.



Ch. 5, Visualization Analysis and Design

However, perceptual biases can lead to misinterpretation of visualizations.

COVER STORY

The Good, the Bad, and the Biased:

Five Ways Visualizations Can Mislead (and How to Fix Them)

Danielle Albers Szafir, University of Colorado Boulder

Insights

→ Visualizations allow people to readily analyze and communicate data. However, many common visualization designs lead to engaging imagery but false conclusions.

→ By understanding what people see when they look at a visualization, we can design visualizations that support more accurate data analysis and avoid unnecessary biases.

Data visualizations allow people to readily explore and communicate knowledge drawn from data. Visualization methods range from standard scatterplots and line graphs to intricate interactive systems for analyzing large data volumes at a glance. But how can we craft visualizations that effectively communicate the right information from our data? What aspects of data and design need to come together to develop accurate insights? The answer lies in the way we see the world: People use their visual and cognitive systems (i.e., our eyes and brain) to extract meaning from visualized data. However, flashy visualizations are not always optimized to help people see what matters. This article reviews common visualization practices that may inhibit effective analysis, why these designs are problematic, and how to avoid them. The discussion illustrates a need to better understand how visualizations can support flexible and accurate data analysis while mitigating potential sources of bias.

Glancing at the bar chart in Figure 1 will likely convince you that one method performs twice as well as the other. However, this visualization is misleading: The true difference between methods is only 5 percent. Talks and articles frequently feature flashy visualizations like this—visualizations

Somewhere Over the Rainbow: An Empirical Assessment of Quantitative Colormaps

Yang Liu
University of Washington
Seattle, WA, USA
yliu0@cs.washington.edu

Jeffrey Heer
University of Washington
Seattle, WA, USA
jheer@uw.edu

ABSTRACT

An essential goal of quantitative color encoding is the accurate mapping of perceptual dimensions of color to the logical structure of data. Prior research identifies weaknesses of "rainbow" colormaps and advocates for ramping in luminance, while recent work contributes multi-hue colormaps generated using perceptually-uniform color models. We contribute a comparative analysis of different colormap types, with a focus on comparing single- and multi-hue schemes. We present a suite of experiments in which subjects perform relative distance judgments among color triplets drawn systematically from each of four single-hue and five multi-hue colormaps. We characterize speed and accuracy across each colormap, and identify conditions that degrade performance. We also find that a combination of perceptual color space and color naming measures more accurately predict user performance than either alone, though the overall accuracy is poor. Based on these results, we distill recommendations on how to design more effective color encodings for scalar data.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

Colormaps; Color Models; Graphical Perception; Visualization; Quantitative Methods; Lab Study.

INTRODUCTION

The rainbow colormap – a scheme spanning the most saturated colors in the spectrum – has been a staple (or depending on one's perspective, eyesore) of visualization practice for many years. Despite its popularity, researchers have documented a number of deficiencies that may hinder accurate reading of the visualized data [4, 26, 36, 42]. They raise the following criticisms: the rainbow colormap is unfriendly to color-blind users [26], it lacks perceptual ordering [4], it fails to capture

CHI 2018, April 21–26, 2018, Montréal, QC, Canada.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-5620-6/18/04 ...$15.00.
http://dx.doi.org/10.1145/3173574.3174172

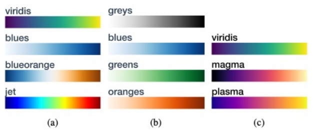


Figure 1: **Colormaps under study.** We evaluate four single-hue, three perceptually-uniform multi-hue, a diverging, and a rainbow colormap(s). We divide them into (a) assorted, (b) single-hue and (c) multi-hue groups, with two colormaps repeated across groups for replication.

the nuances of variations for data with high spatial frequencies [36], and it is ineffective at conveying gradients due to banding effects at hue boundaries [4, 42].

Each of these problems may be traced to a naïve ramping through the space of color hues. In response, a common color design guideline for scalar quantitative data is to use a single-hue colormap that ramps primarily in luminance [6] (from dark to light, or vice versa). Changes in luminance provide a strong perceptual cue for ordering, consistent across individuals and cultures. Moreover, the human visual system has higher-resolution processing pathways for achromatic vision than for chromatic vision [23], supporting discrimination of higher spatial frequencies in the luminance channel.

These considerations raise a natural question: are the above criticisms specific to the rainbow colormap, or do they apply to multi-hue colormaps more generally? Defenders of rainbow and other multi-hue colormaps may cite not only aesthetic concerns, but also a potential for increased visual discrimination. By ramping through hue in addition to luminance, might viewers benefit from greater color separation across a colormap and thereby discern both small and large value differences more reliably? New multi-hue colormaps – the *viridis* colormap and its variants [38] – have recently been adopted by many visualization tools as a replacement for rainbow colormaps. These colormaps were formed by tracing curves through a perceptually-uniform color model, simultaneously ramping in both hue and luminance, while avoiding red-green contrast to respect the most common form of color vision deficiency.

Though existing guidelines and designs for quantitative color derive from both first principles and experience, they have not been comprehensively evaluated. In this work, we investigate

Eurographics Conference on Visualization (EuroVis) 2019
M. Gleicher, H. Leitte, and I. Viola
(Guest Editors)

Volume 38 (2019), Number 3

Examining Implicit Discretization in Spectral Schemes

P. S. Quinan[1], L. M. Padilla[2,3], S. H. Creem-Regehr[3] and M. Meyer[1]

[1]School of Computing, University of Utah
[2]Department of Psychology, Northwestern University
[3]Department of Psychology, University of Utah

Abstract

Two of the primary reasons rainbow color maps are considered ineffective trace back to the idea that they implicitly discretize encoded data into hue-based bands, yet no research addresses what this discretization looks like or how consistent it is across individuals. This paper presents an exploratory study designed to empirically investigate the implicit discretization of common spectral schemes and explore whether the phenomenon can be modeled by variations in lightness, chroma, and hue. Our results suggest that three commonly used rainbow color maps are implicitly discretized with consistency across individuals. The results also indicate, however, that this implicit discretization varies across different datasets, in a way that suggests the visualization community's understanding of both rainbow color maps, and more generally effective color usage, remains incomplete.

CCS Concepts

• *Human-centered computing* → *Empirical studies in visualization;*

1. Introduction

Two of the primary reasons rainbow color maps are considered harmful stem from an argument that they implicitly discretize encoded data into hue-based bands [BT07, BRT95, Mor09]. The literature argues that this perceived banding both highlights non-existent relationships in the data through the creation of false boundaries and masks real relationships within a given band [BT07]. Our current understanding of the implicit discretization in rainbow color maps, however, is based on a combination of generalized knowledge about how humans perceive the visible spectrum [RLK92] and anecdotal evidence that has yet to be empirically tested [BT07].

To our knowledge, no work has empirically evaluated the perceived banding in rainbow color maps or characterized the potential differences across color scales, datasets, or individuals. Understanding and characterizing any perceived banding in rainbow color maps is important, as precise knowledge of how people perceive these bands is essential for leveraging implicit discretization in color-map design to improve the performance of some tasks [PQMC17, DPR*18]. An improved understanding of implicit discretization also ensures that the visualization guidance regarding rainbow color maps has proper scientific foundations [Kos16].

To better understand the perceived banding in rainbow color maps in color displays, we conducted an exploratory study aimed at assessing both whether rainbow color maps implicitly discretize data and how implicit discretization varies across different individuals, datasets, and spectral schemes. Participants were shown sets of color mapped visualizations and asked to first count the color categories/boundaries that they perceived and then interactively delineate those categories/boundaries. Participants' delineations were then compared against potential boundary locations derived from variation in the perceptual dimensions of color for each spectral scheme. Although previous work has attributed some perceived bands in rainbow color maps to variation in luminance [BRT95, BT07, Mor09], luminance alone cannot explain the banding perceived in all spectral schemes [KRC02]. This study expands the investigation of banding effects to variations across all three perceptual dimensions of color: lightness, chroma, and hue.

The study results suggest that rainbow color maps *are* implicitly discretized with consistency across individuals. Additionally, the results show correspondences between participants' responses and variation in each perceptual dimension of color. The results also indicate that the discretization produced by a given color map varies in unexpected and unpredictable ways across different datasets, revealing practical challenges for common tasks like drawing comparisons across datasets. Further, the findings suggest that the visualization community's current understanding of both rainbow color maps, and more generally effective color usage, remain incomplete.

The remainder of this paper is outlined as follows. Section 2 summarizes related work conducted in the visualization, vision science, and cognitive science communities. Section 3 discusses the wide range of definitions for the term *rainbow color map*. Section 4 details both the study's aims and methods. Section 5 then outlines the results of the study, which we discuss further in Section 6, before summarizing our conclusions in Section 7.

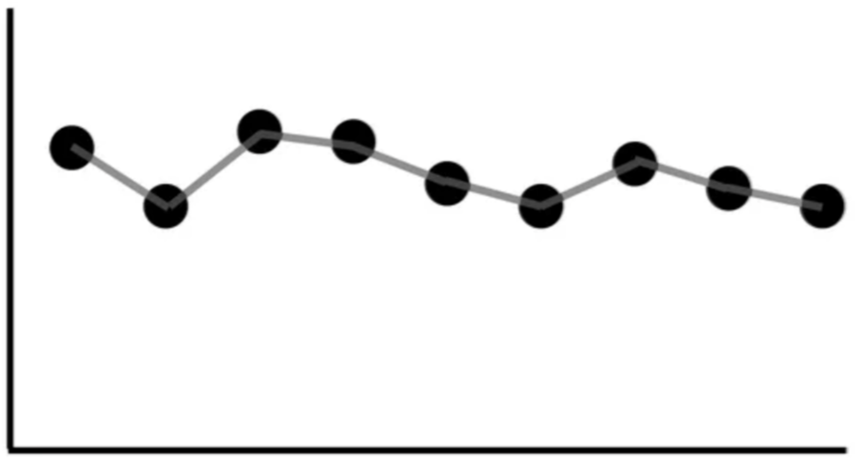
2. Related Work

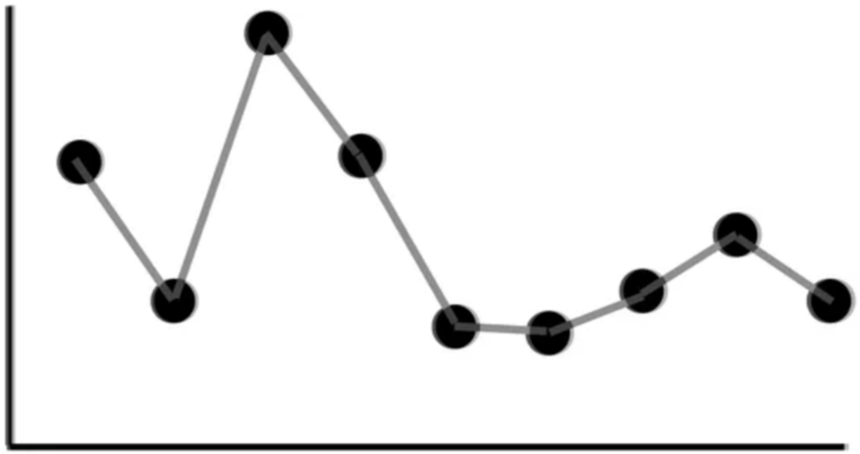
This paper builds on work from two distinct bodies of literature: the visualization community's prior work regarding rainbow color

BACKGROUND

RAINBOW COLOR MAPS

The length of a line drawn varies not only with duration but also with variability of the values.





The arc length of a line affects how much visual weight a line has.

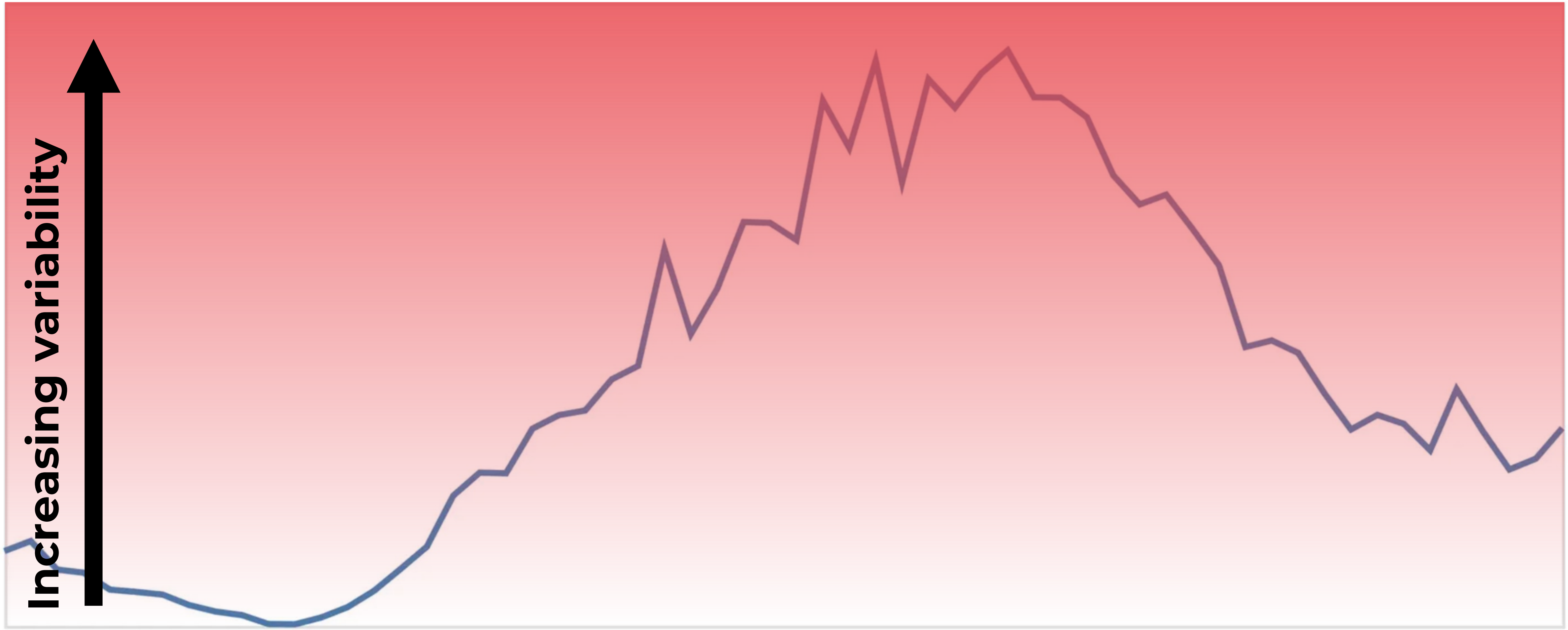
Regions with increased variability capture attention, potentially biasing average estimates.

We proved variability bias at the start of class!
See demo in Colab Notebook.

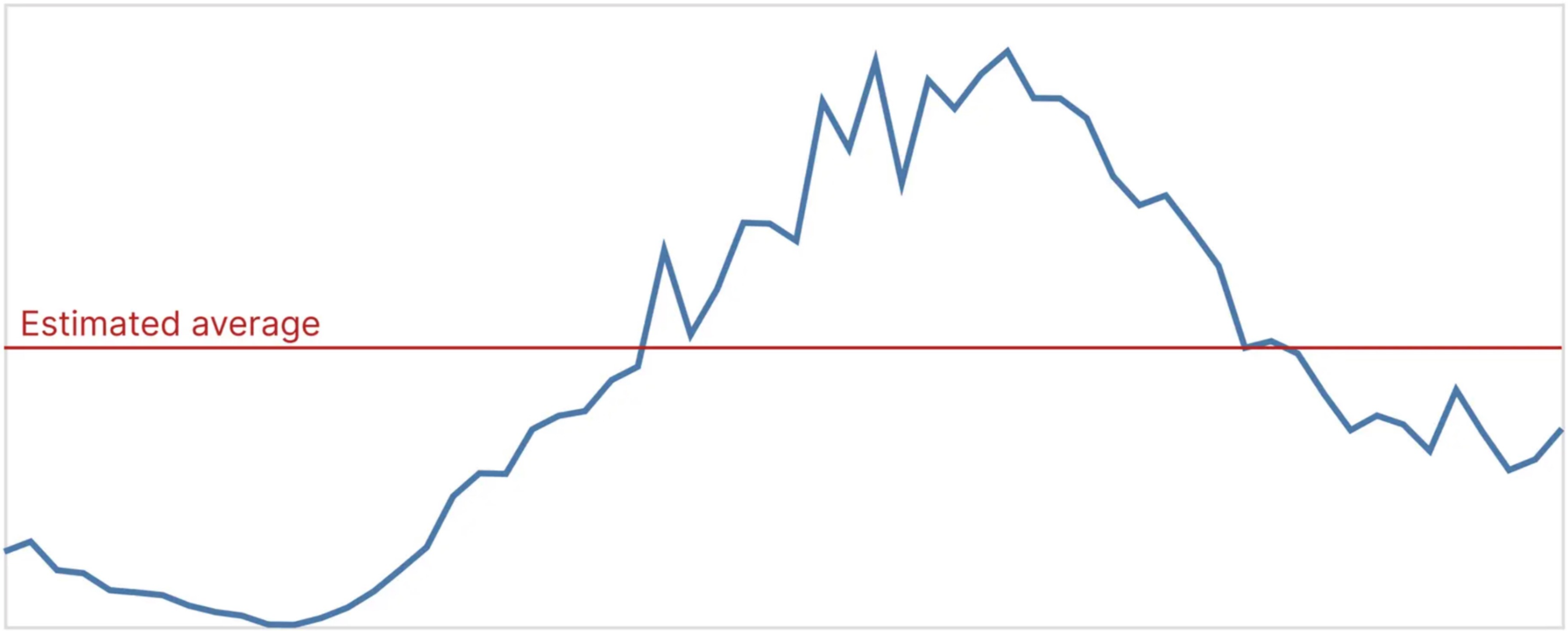
Aneeshers/
CS271VariabilityBiasExp

bit.ly/
CS271VariabilityBiasExp

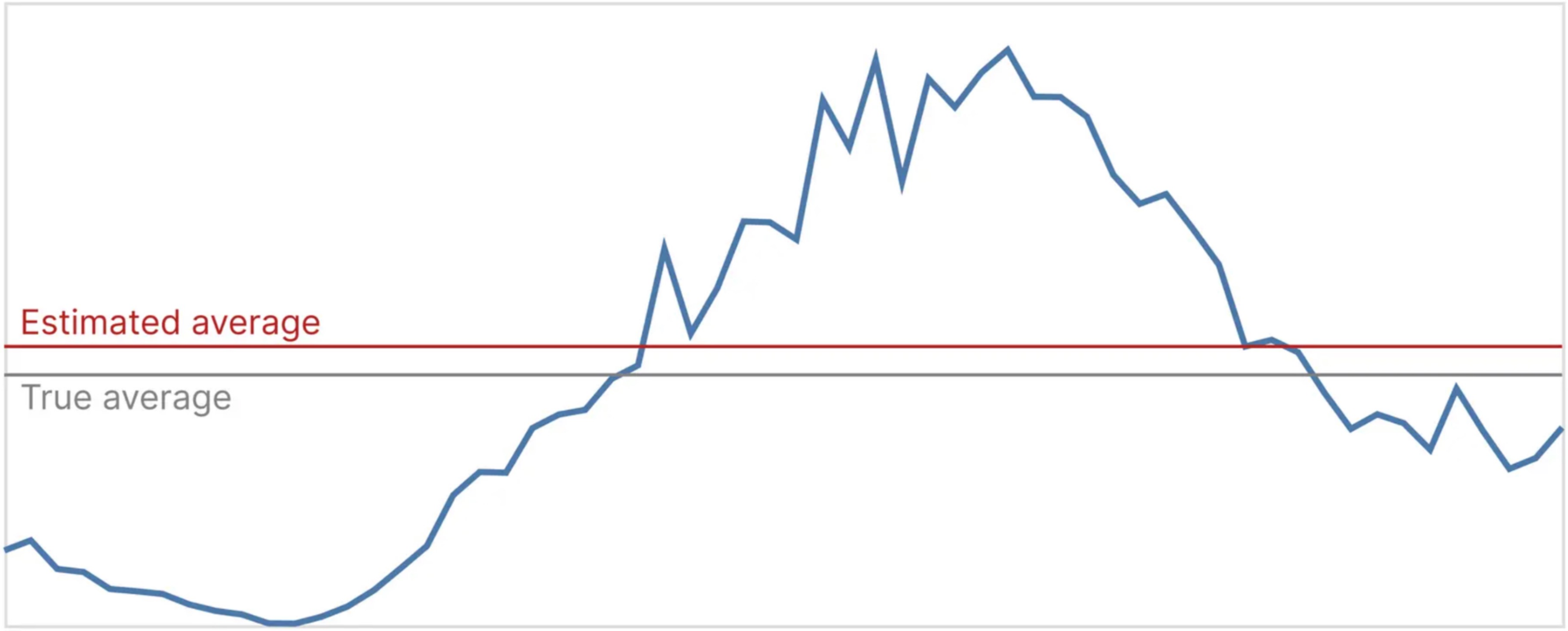
Regions of increased variability may capture visual attention, biasing average estimates.



Regions of increased variability may capture visual attention, biasing average estimates.



Regions of increased variability may capture visual attention, biasing average estimates.



Moritz *et al.* explore how average estimates in line graphs are biased by higher variability.

Average Estimates in Line Graphs Are Biased Toward Areas of Higher Variability

Dominik Moritz, Lace M. Padilla, Francis Nguyen, and Steven L. Franconeri

Fig. 1: Demonstration of the bias toward variability for three mark types showing the same data. The red line shows the mean estimated averages across all participants in our second experiment. The line chart (center) shows a bias of the estimated average toward higher variability in the higher y-values. The bias is smallest when the data is shown as points equally spaced along the x-axis (left). The bias in line charts is in the same direction as the bias of estimates of points sampled at equal intervals along the arc of the line (right).

Abstract— We investigate *variability overweighting*, a previously undocumented bias in line graphs, where estimates of average value are biased toward areas of higher variability in that line. We found this effect across two preregistered experiments with 140 and 420 participants. These experiments also show that the bias is reduced when using a dot encoding of the same series. We can model the bias with the average of the data series and the average of the points drawn along the line. This bias might arise because higher variability leads to stronger weighting in the average calculation, either due to the longer line segments (even though those segments contain the same number of data values) or line segments with higher variability being otherwise more visually salient. Understanding and predicting this bias is important for visualization design guidelines, recommendation systems, and tool builders, as the bias can adversely affect estimates of averages and trends.

Index Terms—bias, lines graph, ensemble perception, average

1 Introduction

Since William Playfair invented line graphs in 1786 [23], they have become one of the most common data visualization types. Designers use line graphs to visualize stocks, sensor data, machine learning metrics, and human vitals (e.g., heart rate). Line graphs show a continuous variable's change over another continuous variable, typically time, as the changing position of a line mark.

We generally assume that visualizations, especially of effective visual encoding channels such as position, are perceived not perfectly but without bias [5]. The popularity of line graphs may be because the visual encoding of time series as the position of a line is considered effective relative to other visual encoding channels, such as hue, depending on the task. However, designers should be cognizant of perceptual biases that can lead to misinterpretation of visualizations [14, 29]. For example, prior work demonstrates that the background color can bias the perception of the color of marks [28], and continuous rainbow color maps are perceived as discrete categories [17, 25].

There may be unexplored biases in line charts as well. When drawing a line, the length of the line drawn varies not only with the duration of the visualized time series but also with the variability of the values (and the resulting variability of the line graph). For example, take two time series of regularly sampled values over the same duration. The first value may be constant while the second value oscillates. Both time series have the same number of values (the same duration), but in the visualization as a line graph, the second line has a longer overall length—we call this the *arc length* of the line. The arc length is the sum of the length of all line segments. Steeper line segments are longer than other line segments of the same length along x. The arc length of a line affects how much visual weight a line has (how much "ink" is needed to draw it) and how much it draws viewers' attention [30]. Within a single line, periods of the same length may have a longer or shorter arc length depending on how much the line goes up and down, which depends on the amount of variability in the visualized time series.

Estimates of average values may be biased by design features of the marks that draw viewers' attention, as found in prior work [13], and increased variability in visualized times series may capture attention. Our bottom-up attention is generally attracted to visual information that contrasts with its surroundings [30]. Marks can vary in contrast to the background and other elements, which dictates how capturing they are to our attention, referred to as *salience*. For example, areas of a line graph with high variability also have more ink (often in color) and more edges, creating high contrast with the background. Therefore, we hypothesize that average estimates in lines are biased toward areas of line graphs that have a longer relative arc length (i.e., that have a longer arc length for the same duration or that use more ink). Put differently, we hypothesize that increased variability in higher values increases the average estimate of a time series (and vice versa) in line graphs and that the bias is consistent with the salience of the line.

We tested this hypothesis in two experiments. Our first experiment showed that average estimates are biased toward the area of the line that visualizes more variable data. In the second experiment, we sought to understand the reasons for the observed bias. We hypothesized that average estimates in line graphs are consistent with the salience of a line. We, therefore, hypothesized that average estimates of points drawn along the arc of a line are more consistent with average estimates

• *Dominik Moritz is with Carnegie Mellon University. E-mail: domoritz@cmu.edu.*
• *Lace M. Padilla is with Northeastern University. E-mail: l.padilla@northeastern.edu.*
• *Steven L. Franconeri is with Northwestern University. E-Mail: franconeri@northwestern.edu.*
• *Francis Nguyen is with UBC. E-mail: frnguyen@cs.ubc.ca.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Average estimates in line graphs are biased toward areas of higher variability

Dominik Moritz, Lace M. Padilla, Francis Ngyuen, Steven L. Franocneri

OUTLINE

Line graphs and bias?

The visual weight of a line depends on its arc length, drawing attention.

Examining variability overweighting

Is there bias in perceptual line graph average estimation?

Influence of mark type

Can the degree of variability overweighting be influenced by mark type?

How *not* to lie with visualization

Evaluation in large patient cohorts and patient-derived brain organoids.

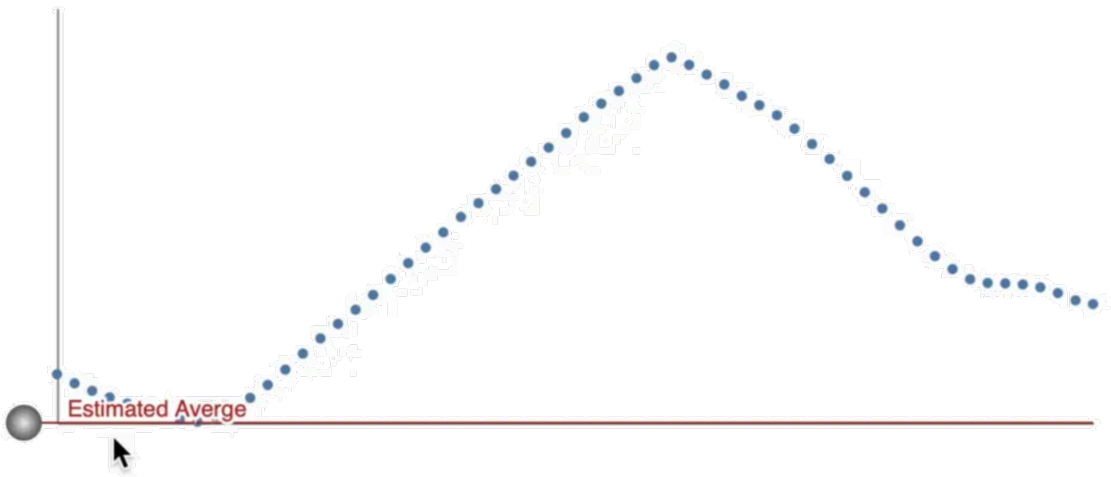
First, authors investigated the existence of bias in perceptual line graph average estimation.

Task:

What is the average closing stock price for the time range shown below?

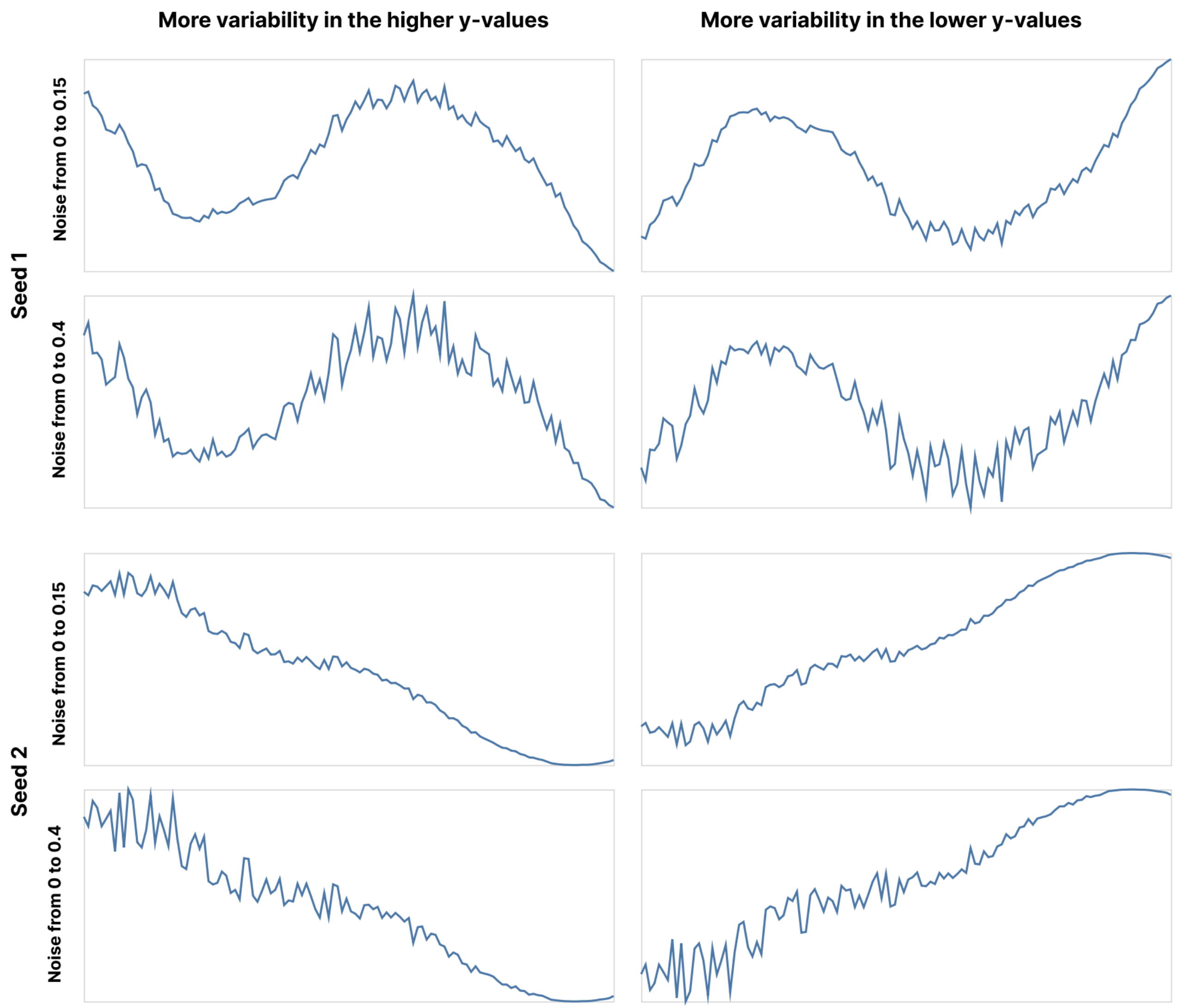
To respond, drag the red line on the chart with the grabber ( ) to indicate where you believe the **average** stock price is. Once you have indicated the average stock price, click "next" to continue.

Trial 1/48



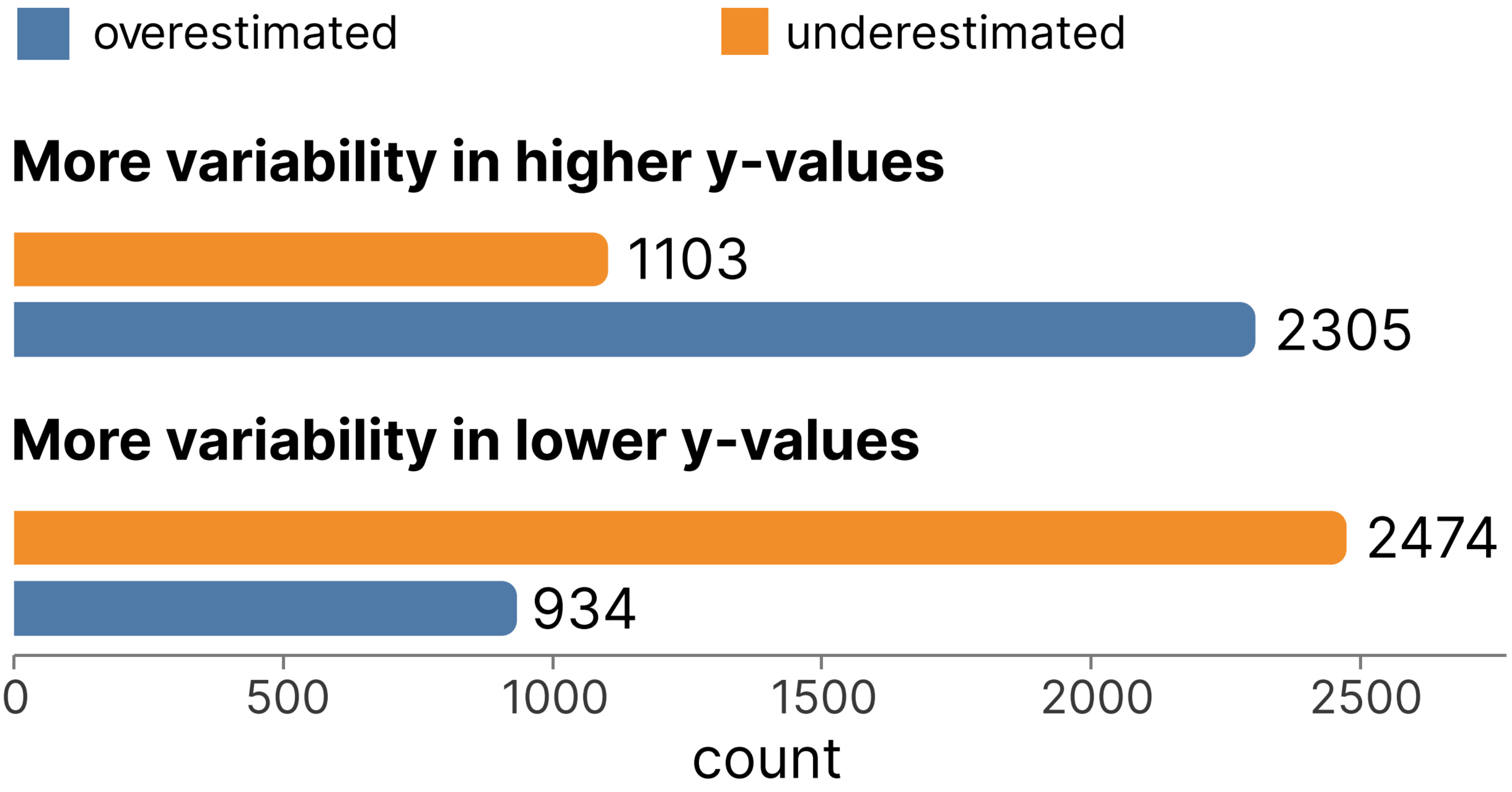
Next

2 × 2 within subjects, 140 participants shown synthetic stock data with added variability.

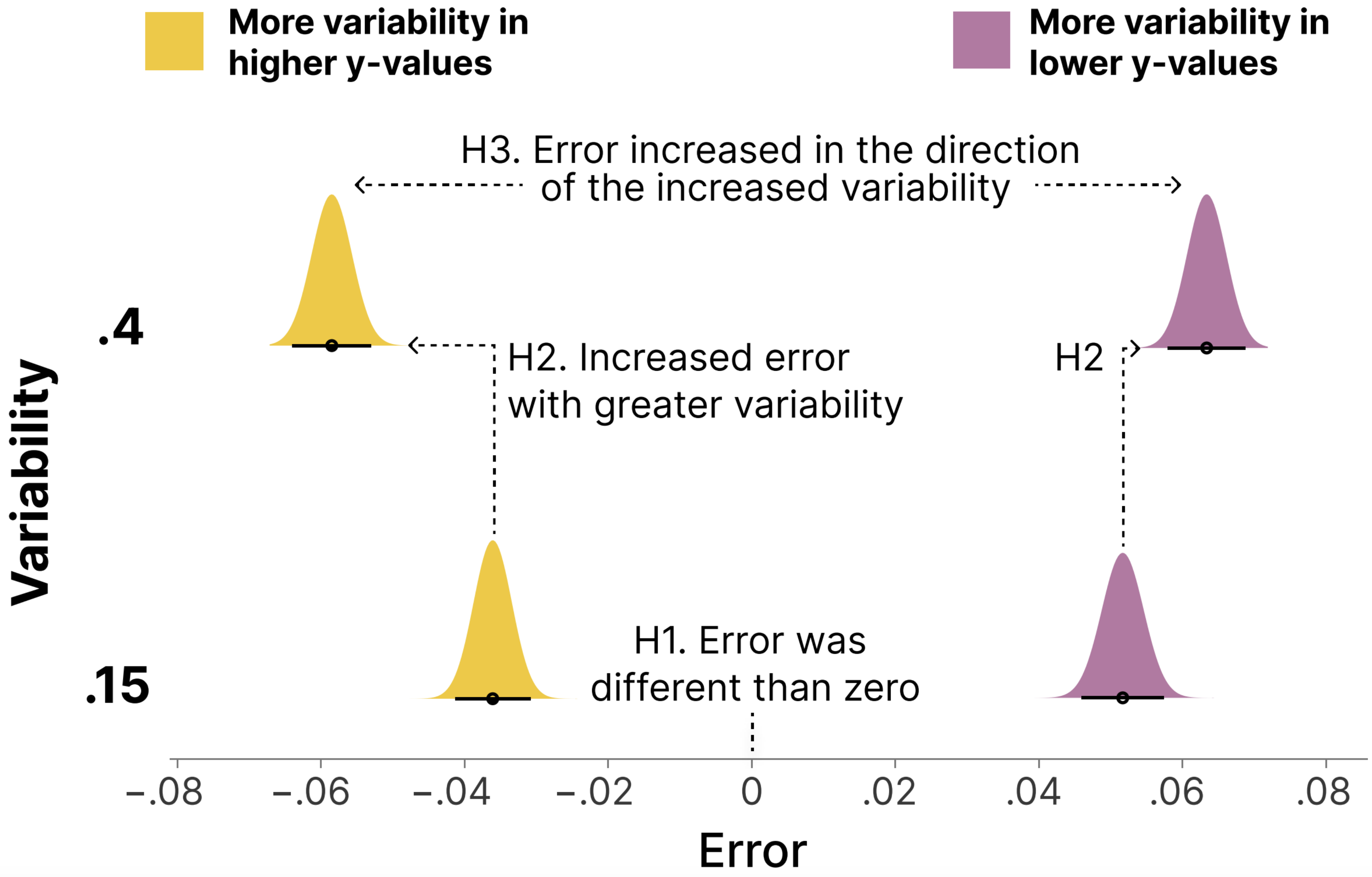


1. 150 points from geometric Brownian motion process
2. Add noise with $\mu = 0$ and $\sigma = 1$ increasing with y-value from 0.15 to 0.4
3. Apply scaling factor to scale the averages

Estimation error is consistent with variability: *e.g.*, higher *y*-variability = overestimation.



Linear regression shows that error increases with variability and in the same direction.



OUTLINE

Line graphs and bias?

The visual weight of a line depends on its arc length, drawing attention.

Examining variability overweighting

Is there bias in perceptual line graph average estimation?

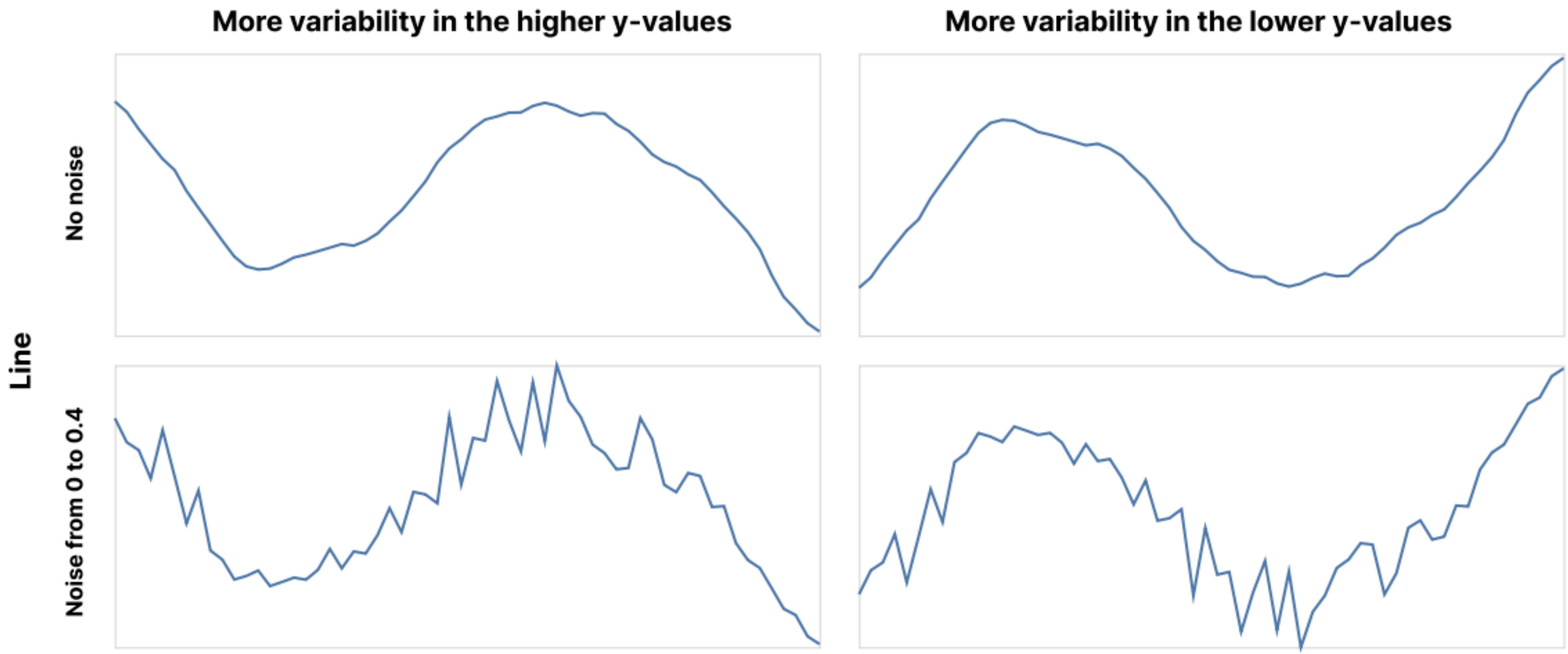
Influence of mark type

Can the degree of variability overweighting be influenced by mark type?

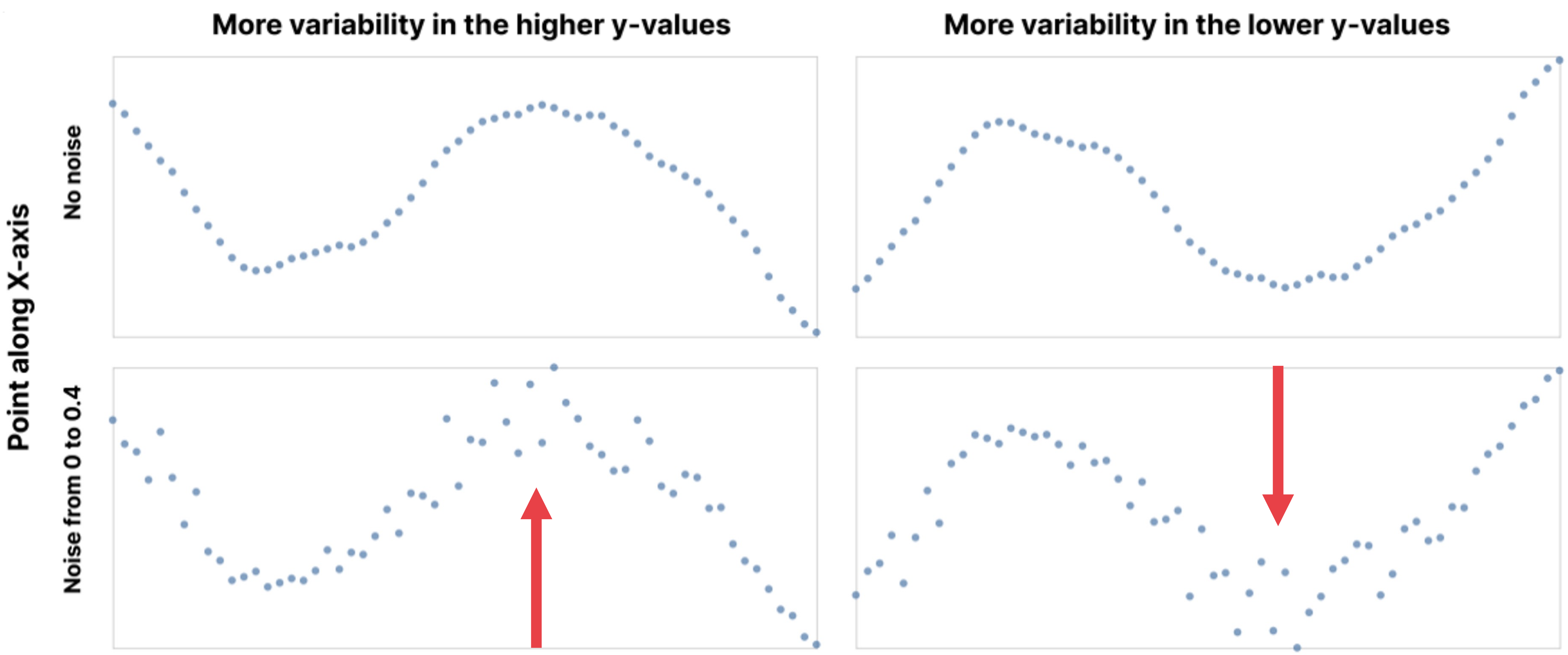
How *not* to lie with visualization

Evaluation in large patient cohorts and patient-derived brain organoids.

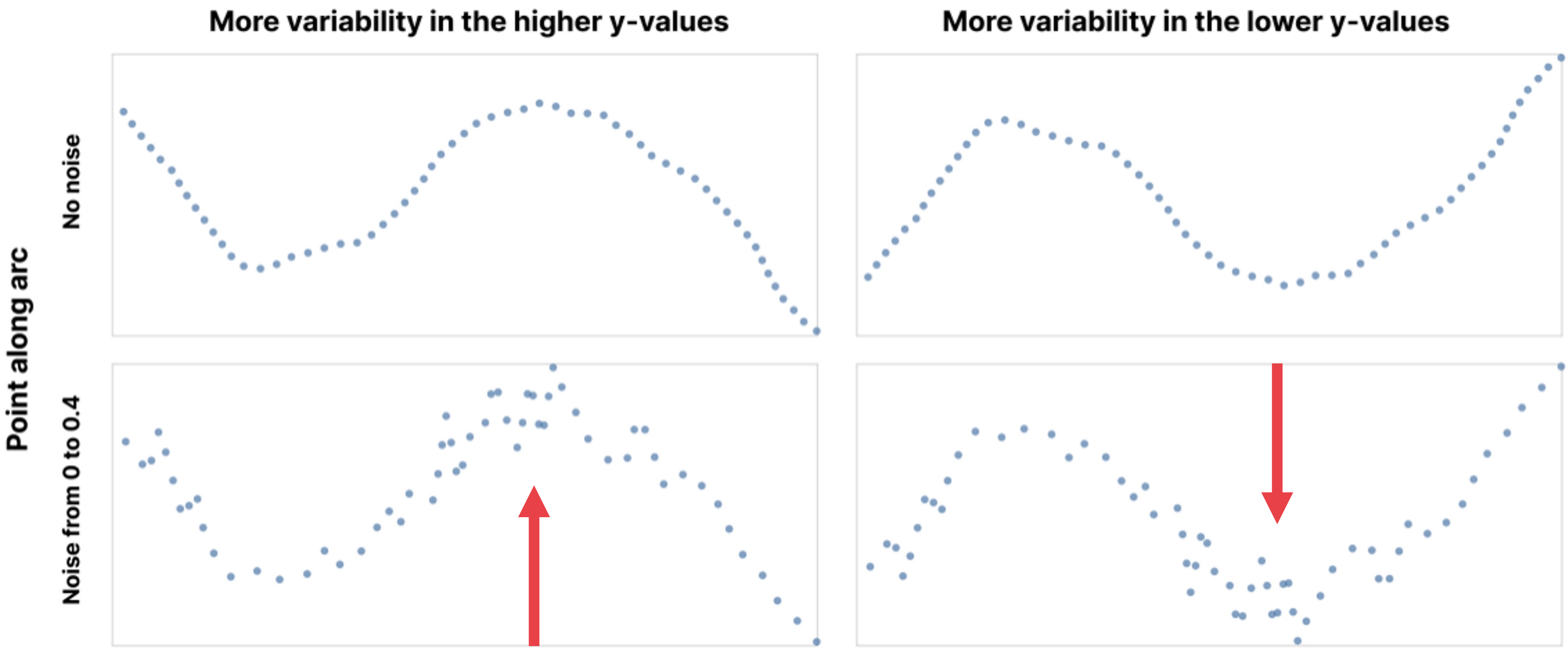
Second experiment tested lines, points sampled along *x*-axis, and points sampled along arc.



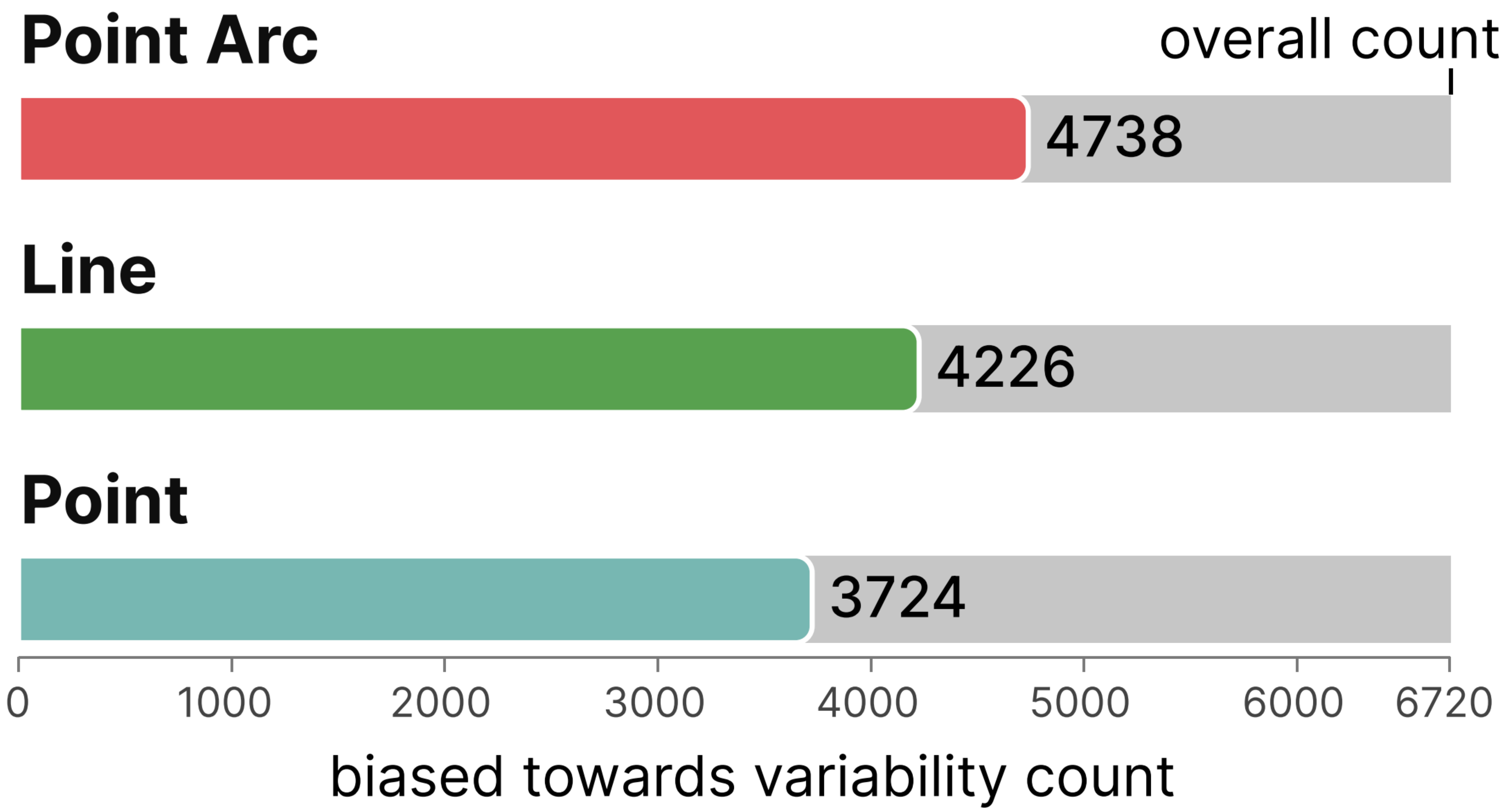
Second experiment tested lines, points sampled along *x*-axis, and points sampled along arc.



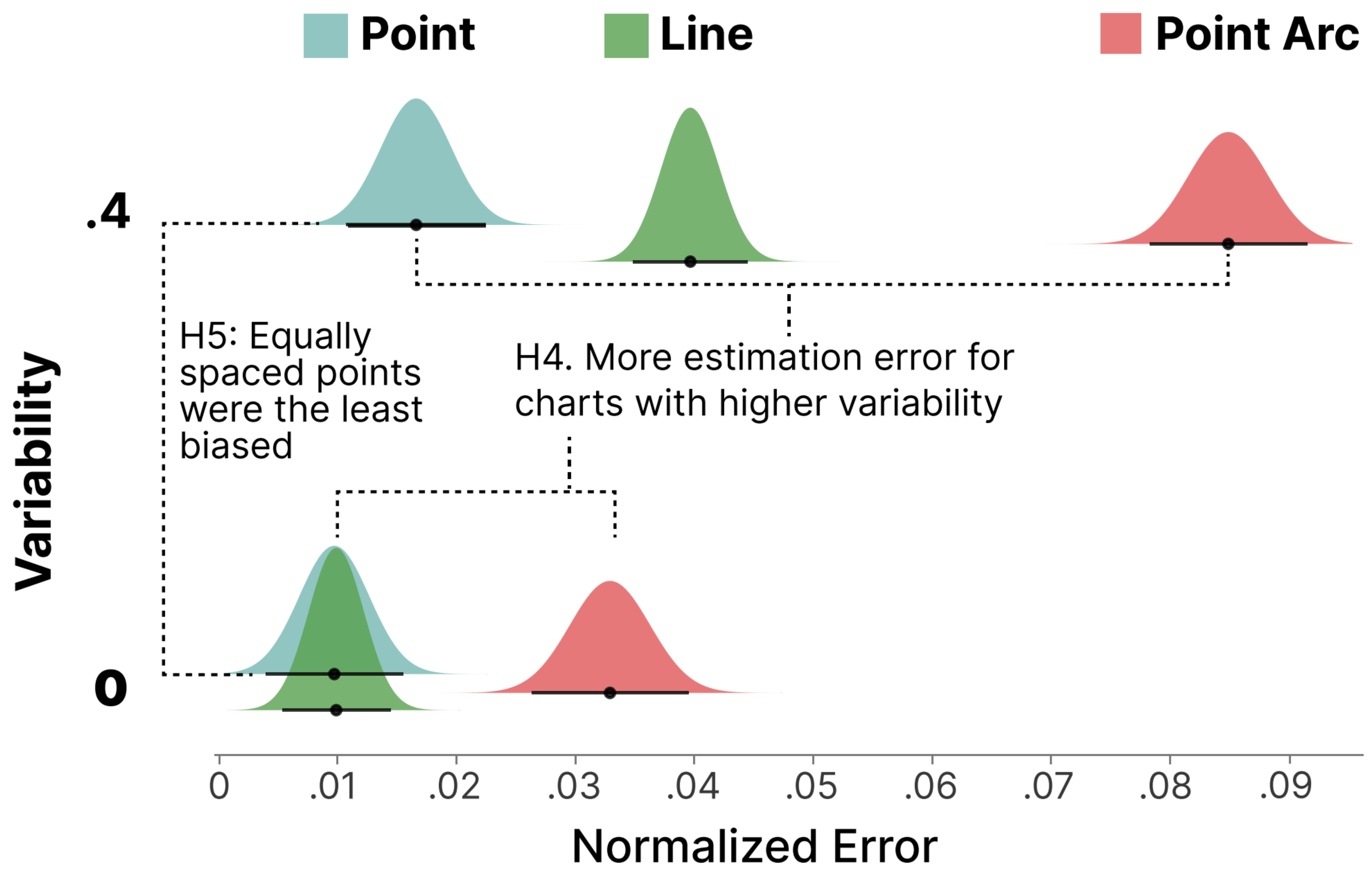
Second experiment tested lines, points sampled along *x*-axis, and points sampled along arc.



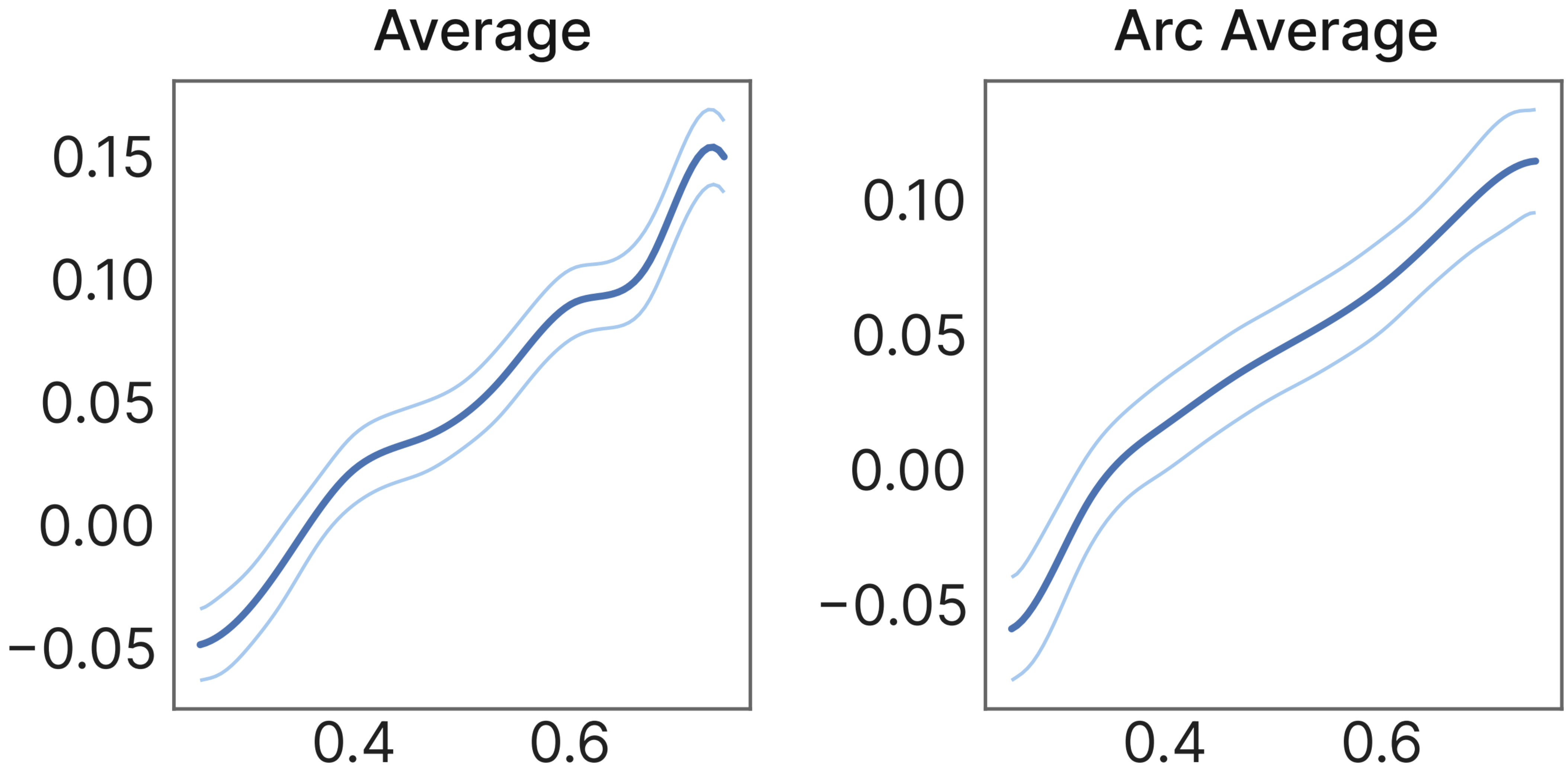
In 2 × 2 × 3 mix design with 420 subjects, more estimation error for charts with higher variability.



Again, linear regression shows more estimation error incurred for charts with higher variability.



Created a model that uses arc average to predict participant bias for a given line chart.



OUTLINE

Line graphs and bias?

The visual weight of a line depends on its arc length, drawing attention.

Examining variability overweighting

Is there bias in perceptual line graph average estimation?

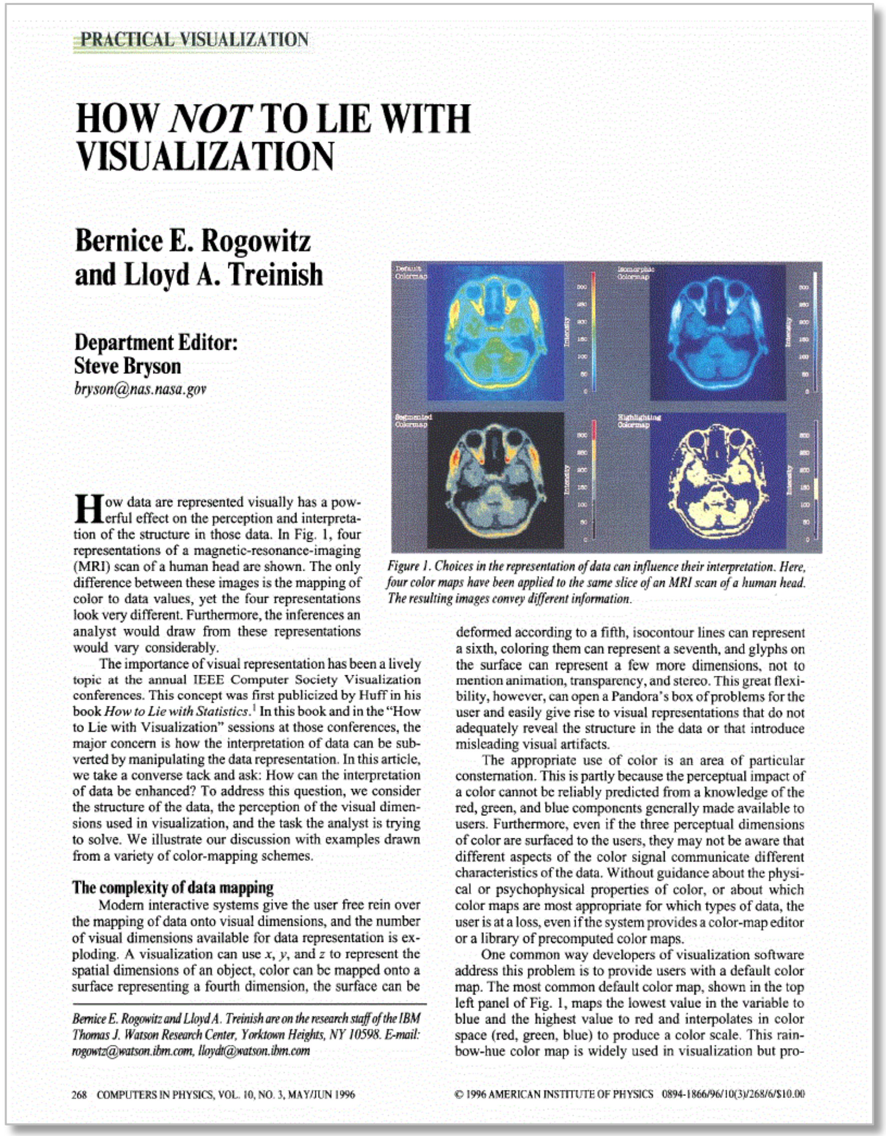
Influence of mark type

Can the degree of variability overweighting be influenced by mark type?

How *not* to lie with visualization

Evaluation in large patient cohorts and patient-derived brain organoids.

How data is visually represented can affect how structure in the data is perceived.



HOW *NOT* TO LIE WITH VISUALIZATION

Bernice E. Rogowitz and Lloyd A. Treinish

Department Editor: Steve Bryson
bryson@nas.nasa.gov

Figure 1. Choices in the representation of data can influence their interpretation. Here, four color maps have been applied to the same slice of an MRI scan of a human head. The resulting images convey different information.

How data are represented visually has a powerful effect on the perception and interpretation of the structure in those data. In Fig. 1, four representations of a magnetic-resonance-imaging (MRI) scan of a human head are shown. The only difference between these images is the mapping of color to data values, yet the four representations look very different. Furthermore, the inferences an analyst would draw from these representations would vary considerably.

The importance of visual representation has been a lively topic at the annual IEEE Computer Society Visualization conferences. This concept was first publicized by Huff in his book *How to Lie with Statistics*.[1] In this book and in the "How to Lie with Visualization" sessions at those conferences, the major concern is how the interpretation of data can be subverted by manipulating the data representation. In this article, we take a converse tack and ask: How can the interpretation of data be enhanced? To address this question, we consider the structure of the data, the perception of the visual dimensions used in visualization, and the task the analyst is trying to solve. We illustrate our discussion with examples drawn from a variety of color-mapping schemes.

The complexity of data mapping

Modern interactive systems give the user free rein over the mapping of data onto visual dimensions, and the number of visual dimensions available for data representation is exploding. A visualization can use *x*, *y*, and *z* to represent the spatial dimensions of an object, color can be mapped onto a surface representing a fourth dimension, the surface can be deformed according to a fifth, isocontour lines can represent a sixth, coloring them can represent a seventh, and glyphs on the surface can represent a few more dimensions, not to mention animation, transparency, and stereo. This great flexibility, however, can open a Pandora's box of problems for the user and easily give rise to visual representations that do not adequately reveal the structure in the data or that introduce misleading visual artifacts.

The appropriate use of color is an area of particular consternation. This is partly because the perceptual impact of a color cannot be reliably predicted from a knowledge of the red, green, and blue components generally made available to users. Furthermore, even if the three perceptual dimensions of color are surfaced to the users, they may not be aware that different aspects of the color signal communicate different characteristics of the data. Without guidance about the physical or psychophysical properties of color, or about which color maps are most appropriate for which types of data, the user is at a loss, even if the system provides a color-map editor or a library of precomputed color maps.

One common way developers of visualization software address this problem is to provide users with a default color map. The most common default color map, shown in the top left panel of Fig. 1, maps the lowest value in the variable to blue and the highest value to red and interpolates in color space (red, green, blue) to produce a color scale. This rainbow-hue color map is widely used in visualization but pro-

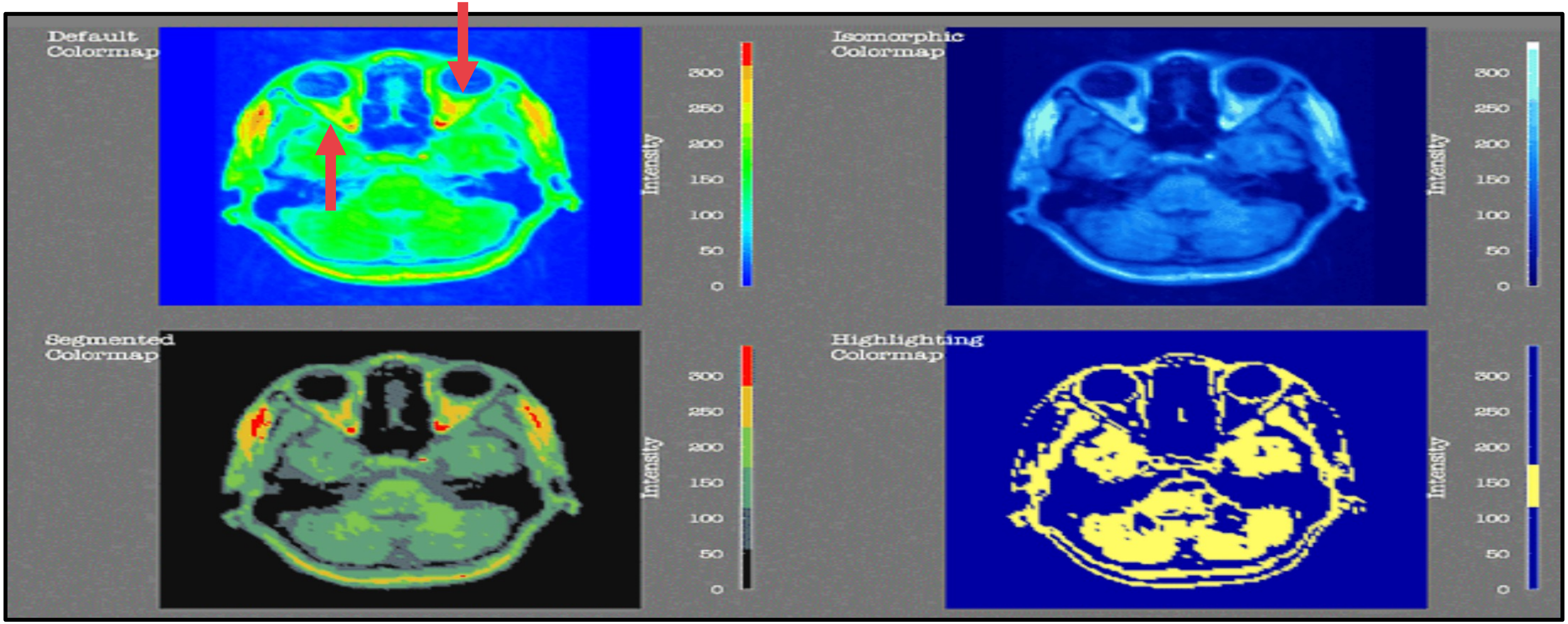
Bernice E. Rogowitz and Lloyd A. Treinish are on the research staff of the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598. E-mail: rogowtz@watson.ibm.com, lloydt@watson.ibm.com

 0894-1866/96/10(3)/268/6/$10.00

How *not* to lie with visualization

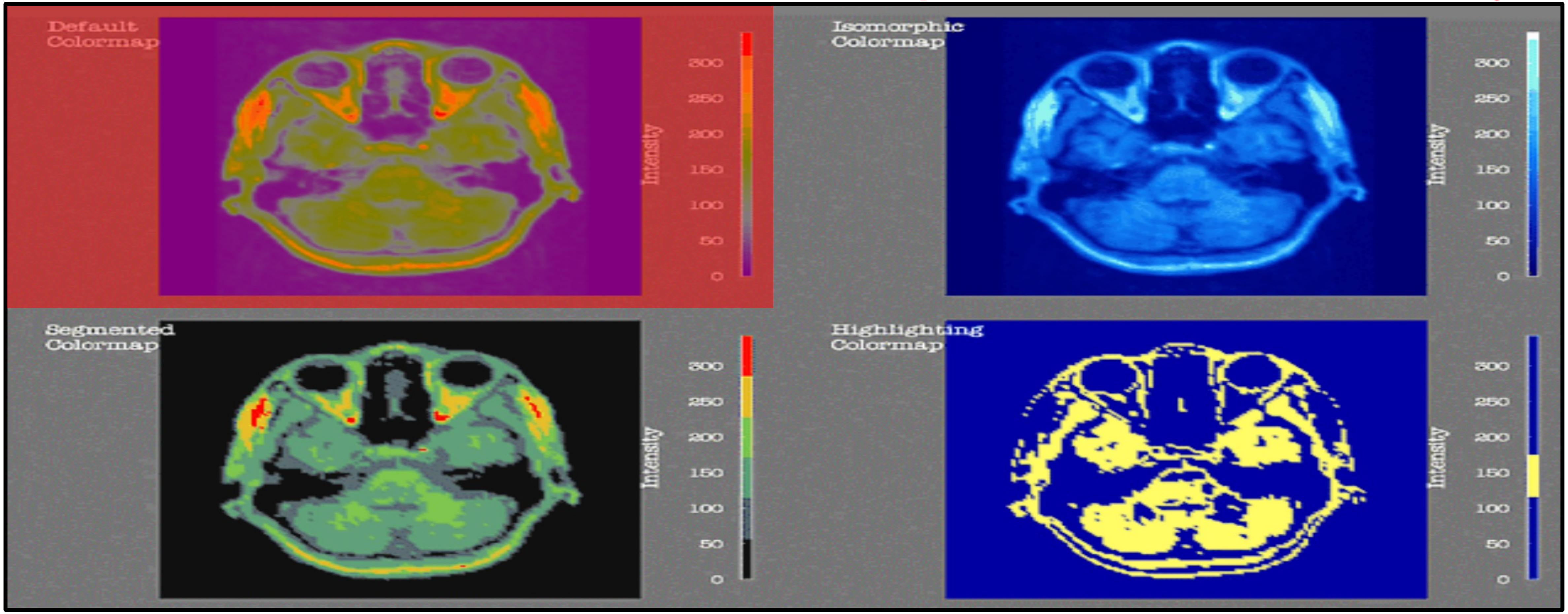
Bernice E. Rogowitz, Lloyd A. Treinish

How data is visually represented can affect how structure in the data is perceived.



Solution? Perceptual rule-based architecture for visualizing data accurately (PRAVDA) software.

Represent structure faithfully.



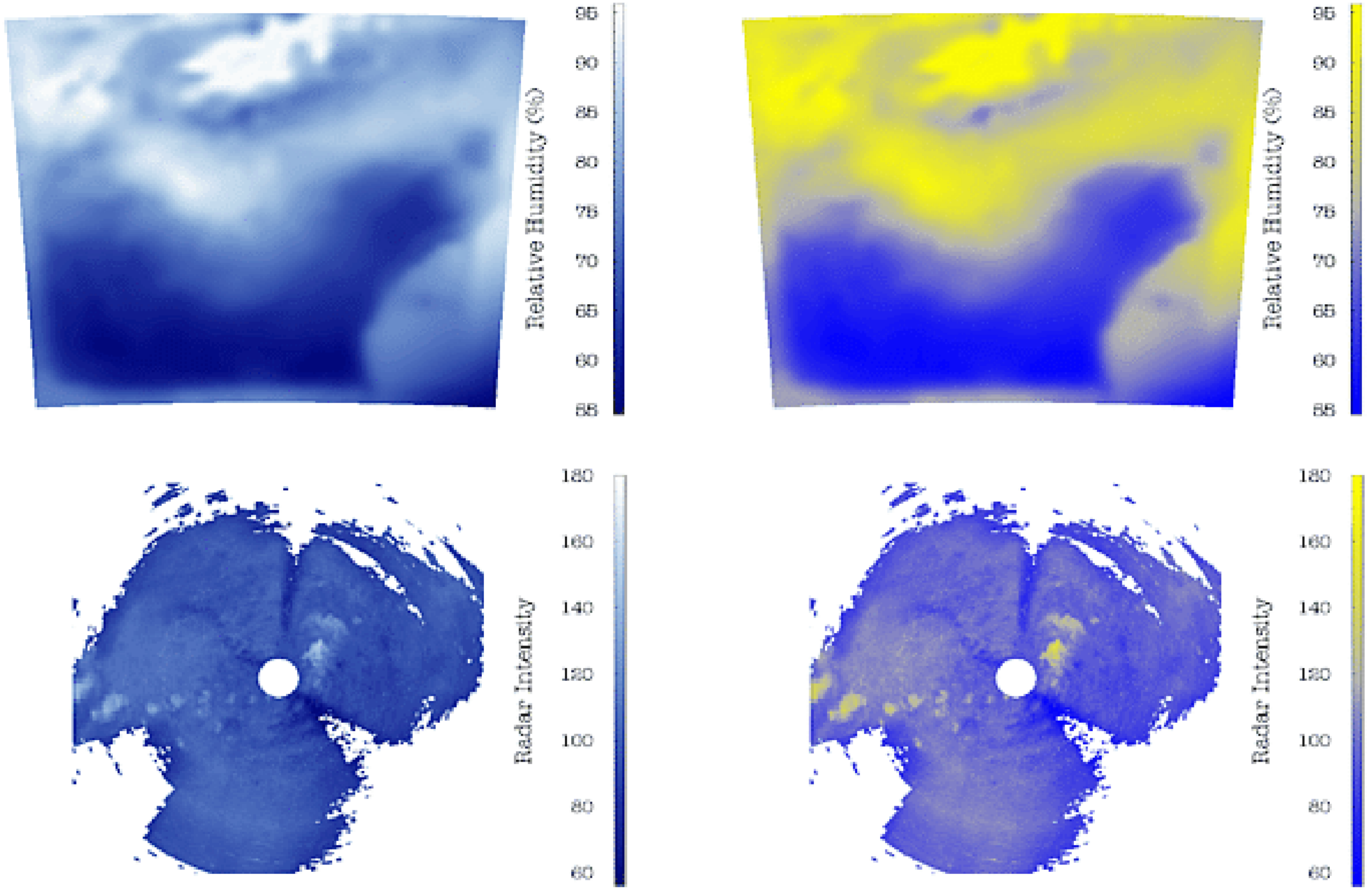
Delineate regions.

Highlight specific features.

In radar data, high frequency colormap reveals detail, low frequency highlights structure.

Low frequency

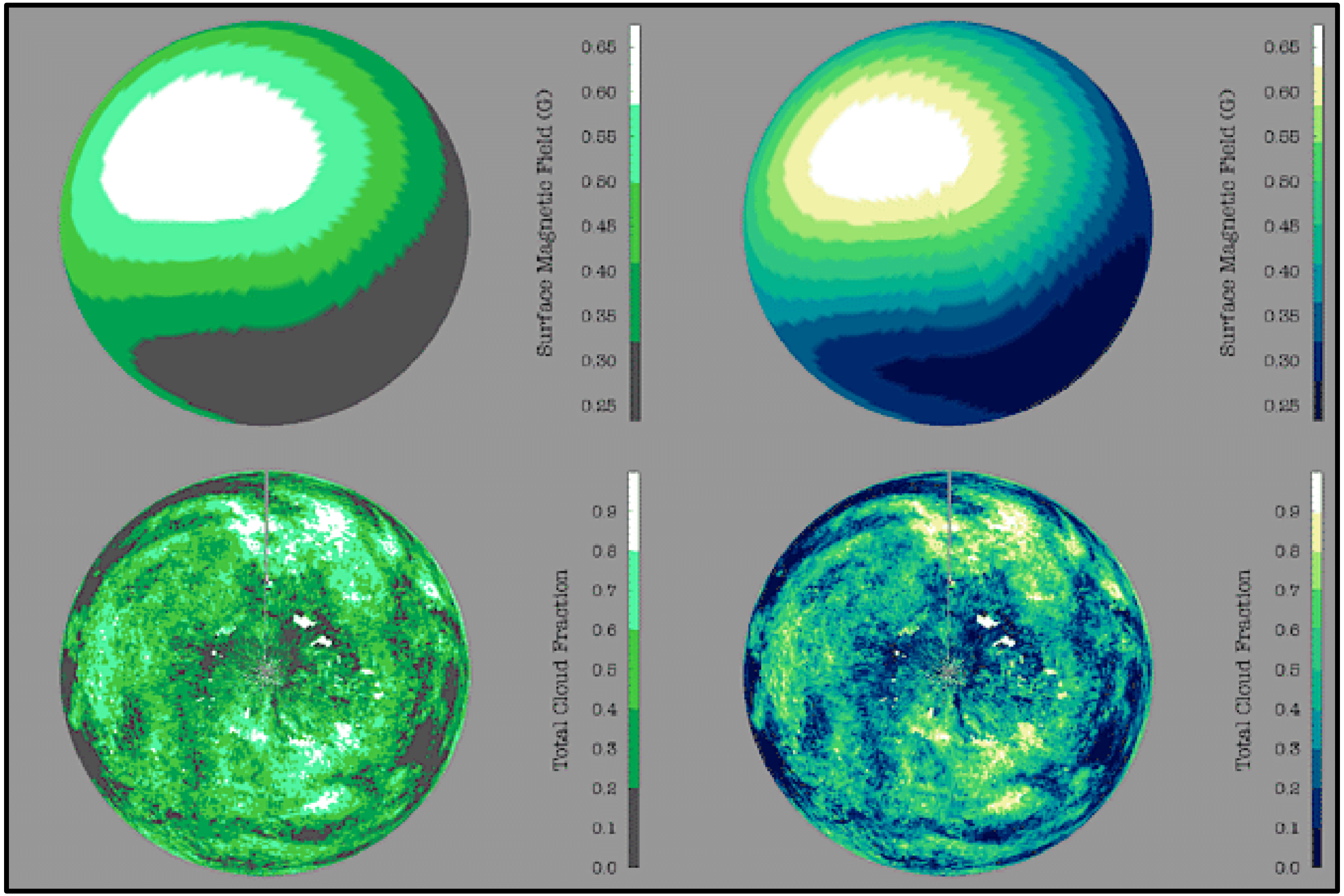
High frequency



In magnetic data, high frequency colormap is useful for high but not low frequency data.

Low frequency

High frequency



We proved variability bias at the start of class! See demo in Colab Notebook.

Aneeshers/
CS271VariabilityBiasExp

bit.ly/
CS271VariabilityBiasExp