

Impact Factor: 17.6

5-Year Impact Factor: 41.8

Contents

 PDF/EPUB

More

## Abstract

Effectively designed data visualizations allow viewers to use their powerful visual systems to understand patterns in data across science, education, health, and public policy. But ineffectively designed visualizations can cause confusion, misunderstanding, or even distrust—especially among viewers with low graphical literacy. We review research-backed guidelines for creating effective and intuitive visualizations oriented toward communicating data to students, coworkers, and the general public. We describe how the visual system can quickly extract broad statistics from a display, whereas poorly designed displays can lead to misperceptions and illusions. Extracting global statistics is fast, but comparing between subsets of values is slow. Effective graphics avoid taxing working memory, guide attention, and respect familiar conventions. Data visualizations can play a critical role in teaching and communication, provided that designers tailor those visualizations to their audience.

This report presents research-backed guidelines for creating powerful and intuitive visualizations oriented toward communicating data to students, coworkers, and the general public. We begin by reviewing guidelines for helping viewers extract data from visualizations in precise and unbiased ways, avoiding a set of known illusions and distortions. We then describe when visual processing of visualizations is

powerful (processing broad statistics) versus where it slows to a crawl (making individual comparisons), and we provide a tool kit for avoiding that slowdown. We review guidelines for ensuring that a viewer properly maps visualized values to the right concepts in the world (e.g., viewers can extract the size of an error bar on a graph, but do they understand what it means?), allowing viewers to use visualizations as effective tools for reasoning. We then review guidelines for conveying uncertainty and risk (e.g., how could a physician express survival odds for a treatment to a patient?). Finally, we summarize a set of guidelines for creating visualizations that communicate clearly and suggest resources for readers interested in learning more.

Data visualizations range from simple graphs in elementary school classrooms, to depictions of uncertainty in election forecasts in news media, to complex data displays used by scientists and analysts. When designed effectively, these displays leverage the human visual system's massive processing power, allowing rapid foraging through patterns in data and intuitive communication of those patterns to other viewers. But when designed ineffectively, these displays leave critical patterns opaque or leave viewers confused about how to navigate unfamiliar displays.

We review methods, empirical findings, theories, and prescriptions across the fields of visual perception, graph comprehension, information visualization, data-based reasoning, uncertainty representation, and health risk communication. These research communities study similar questions and use complementary expertise and styles of inquiry, yet they too rarely connect. We ignore artificial boundaries among these research fields, and instead integrate across them.

## The Importance of Visualization Design and Literacy

Thinking and communicating with data visualizations is critical for an educated public ([Börner et al., 2019](#)). Science education standards require students to use visualizations to understand relationships, to reason about scientific models, and to communicate data to others ([National Governors Association Center for Best Practices and Council of Chief State School Officers, 2010](#); [National Research](#)

[Council, 2013](#)). Evidence-based public policy prescriptions about climate change, vaccines, and policing are argued to be most effectively built ([Kohlhammer et al., 2012](#)) and communicated to the public ([Otten et al., 2015](#)) with visualizations. Journalists at *The New York Times* Upshot, FiveThirtyEight, *The Economist*, and *The Washington Post* use visualizations to communicate data and evidence about statistics and policy. Data visualizations are ubiquitous in the workplace—in data-analysis software, in data-overview dashboards, and in millions of slide presentations created each day ([Berinato, 2016](#); [Parker, 2001](#)). Physicians rely on them to show data about the risks of medical procedures, and meteorologists use them to show the uncertainty in a hurricane's potential path ([Ancker et al., 2006](#); [Ruginski et al., 2016](#)).

In each of these domains, low graphical literacy and ineffective design lead many viewers to struggle to understand these otherwise powerful thinking tools. Many students can find textbook visualizations too challenging to understand or integrate with nearby text ([Nistal et al., 2009](#); [Shah & Hoeffner, 2002](#)). Public policy visualizations can be counterintuitively designed, leading many viewers to draw a conclusion opposite the one suggested by the depicted data ([Engel, 2014](#)). Dozens of best-selling guides have decried the state of visualizations in the workplace and offered prescriptions for more powerful, clear, and persuasive graphs (see the Recommended Practitioner Books section at the end of this article and [Ajani et al., 2021](#), for a more exhaustive list). Medical-risk visualizations can lead patients to fundamentally misunderstand the base rates or risk factors for diseases or medical procedures ([Ancker et al., 2006](#)). When a prediction has a high level of uncertainty that is not intuitively conveyed, the public can lose trust in scientists and analysts. For example, when a hurricane's path deviates somewhat from the most likely trajectory, or when a politician with a 20% predicted chance to win an election prevails, these outcomes may be consistent with the uncertainty inherent to the predictions. But if the forecaster does not effectively visually communicate that uncertainty, their reputation can suffer when their prediction is “wrong” ([Padilla et al., 2021](#)).

# Who Studies the Design and Comprehension of Visualizations?

Research on the design and pedagogy of data visualizations takes place in several communities. A psychologist focusing on perception might study the mapping between a color value in a heat map and the abstract magnitude that an observer extracts from it ([Stevens, 1957](#)). A cognitive psychologist might explore how working memory limits the complexity of the statistical relationships that a viewer might extract ([Halford et al., 2007](#); [Padilla et al., 2018](#)). An education researcher might try to remove roadblocks for students struggling to translate visual depictions to their underlying concepts ([Börner & Polley, 2014](#); [Shah & Hoeffner, 2002](#)) or seek multimedia design principles for designing effective graphics and integrating them with text (e.g., [Mayer & Fiorella, in press](#)). Researchers in public policy communication or political science might study why viewers find some visualizations to be more trustworthy or persuasive than others ([Nyhan & Reifler, 2019](#)). Health communication researchers evaluate how to effectively communicate the risk of a medical procedure to patients with low *numeracy* (i.e., ability to work with numbers and mathematics; [Ancker et al., 2006](#)). Specialists in statistical cognition and communication seek ways to communicate uncertainty across election outcomes, bus arrival times, and hurricane paths ([Hullman, 2019](#)). Finally, a research community housed in computer and information sciences studies data visualization at multiple levels, from data types and algorithms to the creation of user task taxonomies, to design prescriptions for visually powerful displays and fluid interaction ([Munzner, 2014](#)).

In this article, we also draw advice from communities of practitioners who might not engage in empirical research but use extensive in-context experience to generate prescriptions for powerful and intuitive visualizations. At the end of this review, we include a list of recommended visualization-design guidebooks. Although many of these guides are oriented toward business analysts, their prescriptions extrapolate directly to science, education, and public policy visualizations. We also discuss

design techniques used by a new wave of journalists focused on communicating data analysis to the general public.

## The Structure of Our Review

This review focuses on how to effectively design visualizations that communicate data to students and the general public. We review evidence-based prescriptions for designing visualizations that help people understand and reason about the patterns, models, and uncertainties carried by a data set. Another important topic, which we do not cover systematically here, is how to measure visualization literacy and the effectiveness of teaching techniques that improve it (see [Börner et al., 2019](#); [Lee et al., 2016](#)). We also restrict our scope to quantitative visualizations, omitting discussion of qualitative visualizations of text data, diagrams, and processes (see [Hegarty, 2011](#); [Henderson & Segal, 2013](#), for review). We focus on research and prescriptions that are most relevant for communication to nonspecialist audiences, instead of the design of powerful tools for data analysis within expert communities.

We first illustrate why visualizations can be such powerful tools for thinking about data. Because the human visual system is highly developed for rapid parallel extraction of behavior-relevant features and relationships, visualizations allow us to process some types of patterns across an entire two-dimensional array of values at once. We describe the limited set of visual channels that can effectively depict magnitudes to a viewer, such as the *position* of a value in a dot plot, the *size* of a circle hovering over a map, or the *color intensity* of an activation pattern in a functional MRI (fMRI) image.

We then discuss design guidelines for ensuring that the human eye accurately decodes those depicted values. We review evidence for a ranking of some visual channels (e.g., position) as more precise than others (e.g., color intensity) for at least one common task but also discuss how new work has begun to dismantle that ranking for a broader array of tasks. We list a set of common errors and illusions that cause viewers to extract the underlying values from visual channels incorrectly—for example, *y*-axis manipulations that exaggerate differences among values,

confusion about whether circles depict values with their size or diameter (which can change the extracted value by an order of magnitude), a common illusion produced by line graphs, and other illusions and categorical distortions that can arise when depicting value with color intensity. We also include a brief review of accessibility considerations for viewers with color blindness. Finally, we discuss best practices for distinguishing between groups of data (say, two groups of points on a scatterplot) by marking them with different shapes or colors.

Next, we discuss an important dissociation in visual processing power: Whereas computing statistics across an image is broad and instantaneous, making comparisons among subsets of values is slow and limited to two or three comparisons per second. We review the types of grouping cues that loosely control what information is compared by a typical viewer and further techniques for precisely guiding a viewer to the right comparison. We discuss the importance of respecting a viewer's limited working memory, including avoiding legends and animated displays that can engage but also confuse. Finally, we review evaluations of whether visualizations should have rich and memorable designs, as opposed to a minimalist and clean aesthetic.

The next section introduces visualization *schemas*, or knowledge structures that include default expectations, rules, and associations that a viewer uses to extract conceptual information from a data visualization. We illustrate the importance of schemas by introducing the reader to a small set of new visualization designs that will likely be unfamiliar. We then provide examples of common schema elements that are known to more graphically literate audiences (but not always respected by designers), such as the assumption that larger values are plotted upward. We shift to a brief review of human reasoning about visualizations, including formal models that draw links from visual depictions, to numeric values, to their underlying concepts and the designer's intended message. We then explore two case studies: reasoning about graphs illustrating scientific concepts and reasoning about graphs of mathematical functions.

The subsequent sections review research on visualizing uncertainty or risk. Communication failures can start with a lack of understanding of critical statistical

concepts, even among scientists. We give examples of how viewers tend to misread error bars as depicting the edges of a range of data instead of correctly understanding them as parameters of a distribution. Probability information expressed as risk is critical for people such as patients considering a medical procedure and potential evacuees who may be in the path of a hurricane, but depictions of risk are frequently misunderstood. We review guidelines for showing uncertainty or risk more intuitively, including depicting samples of discrete outcomes, showing probability density functions, and depicting data with arrays of icons.

Finally, we summarize a set of evidence-based prescriptions for creating powerful visualizations for intuitive communication of data and provide a list of recommended practitioner guides ([Box 1](#)), websites, and data-journalism outlets for further reading and inspiration (for a concise review of similar guidelines, see also [Zacks & Franconeri, 2020](#)).

### **Box 1**

#### **Recommended Reading**

The following books offer excellent, concise guides to designing effective visualizations for communication of data and analyses:

Camões, J. (2016). *Data at work: Best practices for creating effective charts and information graphics in Microsoft Excel*. New Riders.

Evergreen, S. D. H. (2017). *Presenting data effectively: Communicating your findings for maximum impact* (2nd ed.). SAGE Publications.

Knafllic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Wiley.

Schwabish, J. (2021). *Better data visualizations: A guide for scholars, researchers, and wonks*. Columbia University Press.

For a review of the ethics of persuasion with visualizations, we recommend the following books:

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. New Riders.

Cairo, A. (2019). *How charts lie: Getting smarter about visual information*. W.W. Norton.

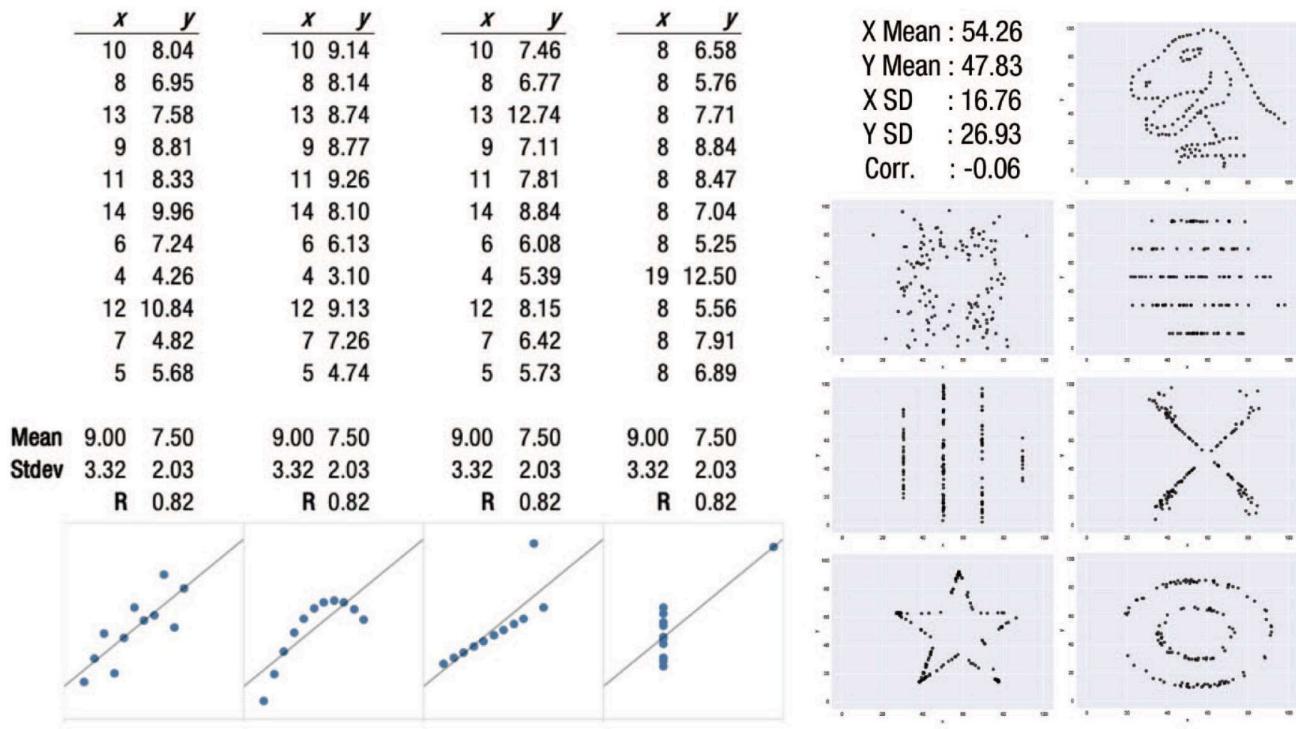
For a review of accessibility considerations in data visualization, we recommend the following article:

Kim, N. W., Joyner, S. C., Riegelhuth, A., & Kim, Y. (2021). Accessible visualization: Design space, opportunities, and challenges. *Computer Graphics Forum*, 40(3), 173–188.

## The Power of Visualization

Visualizations let viewers see beyond summary statistics

Visualizations allow powerful processing of an entire two-dimensional rectangle of information at once, in stark contrast to the limitation of reading handfuls of symbolic numbers per second. As a demonstration, [Figure 1](#) (top left) contains four sets of 11 pairs of values. Take a moment to compare those columns, and notice that reading symbolically represented numbers takes time. As you seek patterns within each set, or make comparisons among the four sets, progressively processing more pairs of values becomes increasingly difficult. Worse, these tasks quickly exhaust memory capacity, such that new numbers or patterns tend to displace ones that were previously seen. These limitations on symbolic processing of numbers lead viewers to rely instead on summary statistics that compress data sets into a single group of numbers. For the four sets of numbers in [Figure 1](#), those statistics on the bottom row—means, standard deviations, and correlation coefficients—are identical, which might lead you to believe that the numbers contributing to the statistics are similar ([Anscombe, 1973](#)). However, because statistics summarize larger sets of numbers by abstracting over them and making assumptions about the patterns that they might contain, many sets of numbers can generate the same statistics. For these four sets of numbers, relying on statistics turns out to be dangerous.



**Fig. 1.** Examples of how visualizations can let viewers see beyond summary statistics. At left, four sets of 11 numbers have identical statistics but dramatically different patterns, as revealed by the scatterplots below each column. At right is a more extreme example of nine dramatically different scatterplots (including one that looks suspiciously like a dinosaur) depicting data with identical statistics, down to the second decimal place. The graphs on the right are adapted with permission of the Association for Computing Machinery, from “Same Stats, Different Graphs: Generating Datasets With Varied Appearance and Identical Statistics Through Simulated Annealing,” by J. Matejka and G. Fitzmaurice, *CHI ’17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (<https://doi.org/10.1145/3025453.3025912>). Copyright 2017 Association for Computing Machinery.

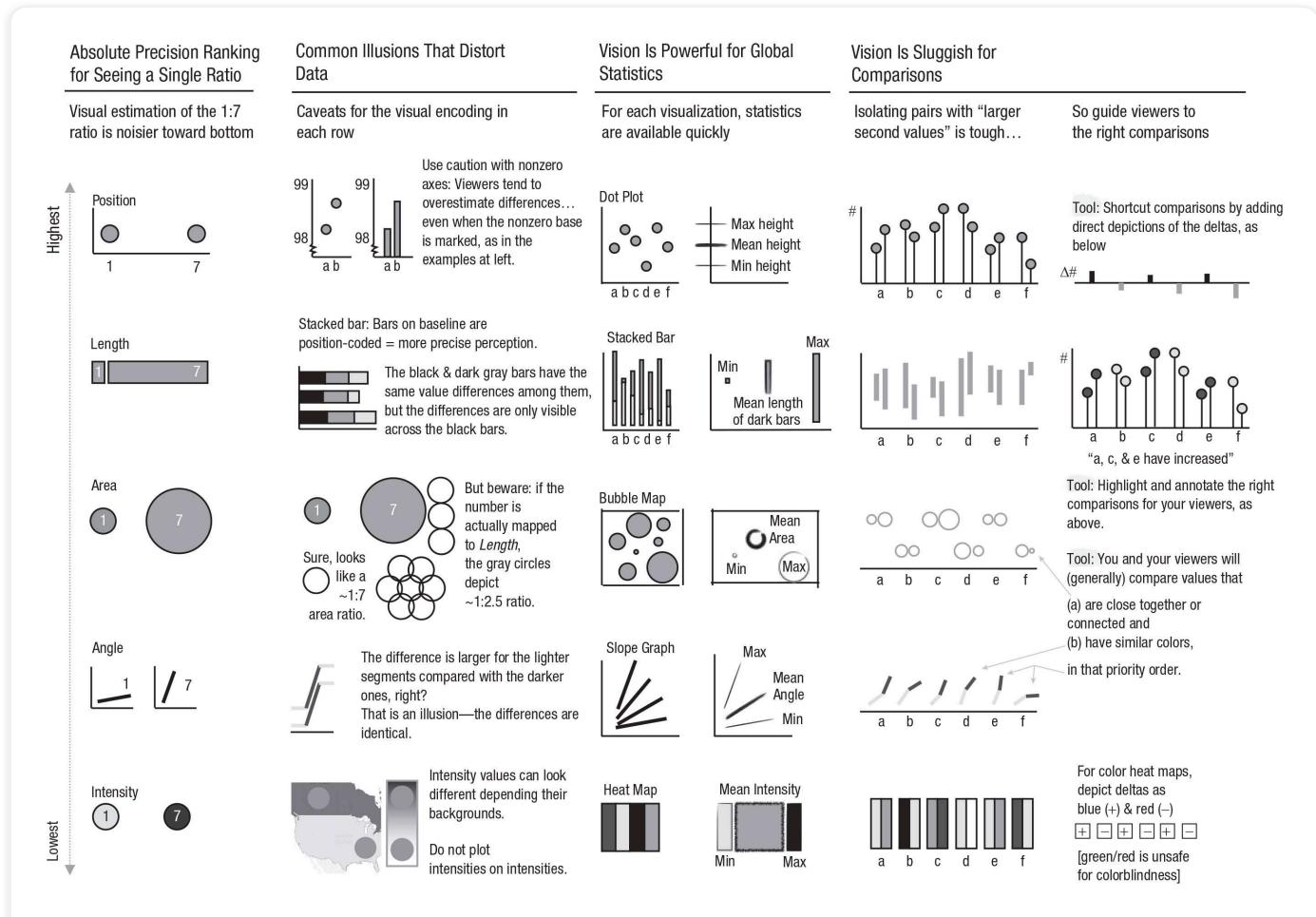
The differences between the four sets of numbers in [Figure 1](#) are immediately visible when translated into images in the form of scatterplots (bottom left). These images allow you to leverage the two-dimensional processing power of your visual system, the largest single processing system of the brain ([Van Essen et al., 1992](#)). If you are familiar with statistics, then the first image at the bottom left likely matches

what you assumed the numbers should look like given the statistics on the bottom rows: an orderly positive relationship. But the other sets are clearly different in important ways. The right side of [Figure 1](#) depicts frames of an animation that provides a more sophisticated example ([Matejka & Fitzmaurice, 2017](#)), in which visual processing allows one to immediately see that, despite identical statistics (to the second decimal place) for each scatterplot, the nine plots contain saliently different patterns. When exploring a new data set, it is good practice to visualize every column of data with a histogram, and every potentially interesting pairing of columns with scatterplots, before turning to statistical summaries ([Moore et al., 2017](#); [Wongsuphasawat et al., 2015](#)).

## Visual channels translate numbers into images

Visualizations rely on several visual channels to transform numbers into images that the visual system can efficiently process ([Bertin, 1983](#); [Mackinlay, 1986](#); see [Munzner, 2014](#), for a more complete list). Knowing these channels allows a designer to consider which might be best suited for a given data set and context—particularly given that each is associated with differential levels of precision and potential illusions, as described in the following sections. The first column of [Figure 2](#) depicts five of the more frequently used channels. Dot plots and scatterplots, such as those in [Figure 1](#), represent values as *position*. Bar graphs represent values not only as positions (of the tips of the bars) but also as one-dimensional *lengths* (and, some argue, even two-dimensional *areas*; [Yuan et al., 2019](#)). If two bars do not rest on the same baseline, such as segments within the same bar in a stacked bar graph, the comparison relies only on length or area. Next, two circles code numbers exclusively as two-dimensional areas (typically circles), a technique often used to overlay values across maps. *Angle* typically emerges when points are connected to form a line graph, organically allowing an encoding of the difference between adjacent points (a bigger difference creates a steeper slope and, typically, a longer line). Outside of pie charts, angle is less frequently used to depict numbers directly—perhaps on local areas of a map to represent wind directions. *Numerosity* is omitted from the figure, but it often implicitly shows higher-level attributes of data.

For example, you can immediately estimate the number of points in a scatterplot, segments in a stacked bar chart, or icons in an infographic. *Intensity* is an umbrella term (often also called *lightness* or *value*) for either luminance contrast or color saturation, as used in a heat map or fMRI activation map. *Motion* is also not included in the figure, but animating a scatterplot to show values changing over time can encode the rate and direction of change in the speed and direction of the dots' motion.



**Fig. 2.** A summary of the power and limits of visual data processing. The two columns on the left show a quick reference guide to channels that can depict data visually and common illusions for each channel. The column in the center presents a summary of how visual statistics are powerful. The two columns on the right illustrate how comparisons are severely limited and present a set of design techniques that focus viewers on the "right" ones.

# How to Design a Perceptually Accurate Visualization

Understand how to leverage visual channels

**Visual channels are ranked by their perceptual accuracy**

These channels differ in how precisely they convey numeric values to a viewer, and knowing the ranking of these channels allows a designer to prioritize what information to show most precisely. The leftmost column of [Figure 2](#) presents five of the channels that can depict metric data to the human visual system. This list is ordered by the typical precision with which a viewer can verbally state the ratios between the two values shown; more precise ways of communicating numbers are at the top and less precise ways are at the bottom ([Cleveland & McGill, 1984, 1985](#); [Heer & Bostock, 2010](#)). It should be clear from the figure that the 1:7 ratio can be relatively precisely extracted for position, but that task is a bit tougher for area, and far more difficult for intensity, at the bottom of the list.

Because position is the clear winner for precision, visualization designers often prioritize the vertical and horizontal dimensions of two-dimensional space when depicting or organizing quantitative data. Faced with a single column of numbers in a spreadsheet, a visualization designer might depict those data vertically with position (in a bar or line graph) and rely on horizontal position to organize the values into categories, as in a typical bar chart. If faced with two columns of numbers, a designer might simply create two of those same types of graphs or organize each set of numbers along both the vertical and horizontal axes of position, as in a scatterplot.

The advantage of position over length for precisely depicting ratios between numbers is demonstrated in the second column of [Figure 2](#), which shows a horizontally oriented stacked bar graph in its second row of examples. Because the black segments of the bars are aligned on a common axis at their left, their right tips provide a precise position code, allowing the viewer to see the delicate 0.9:1 ratio between the bars. However, the next set of medium-gray segments are tougher to distinguish because the positions of their right tips are no longer useful,

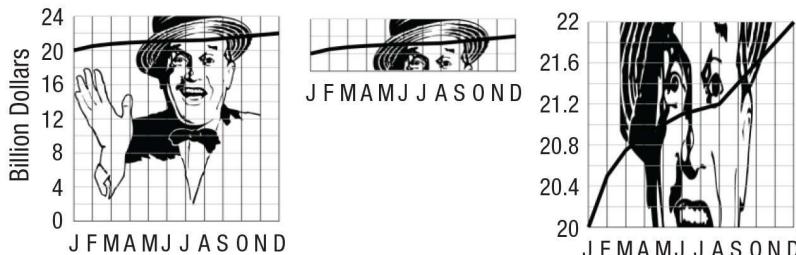
so the viewer must rely on length—a lower-precision channel—to extract the same ratio.

### Mapping visual ratios back to numbers can cause perceptual errors

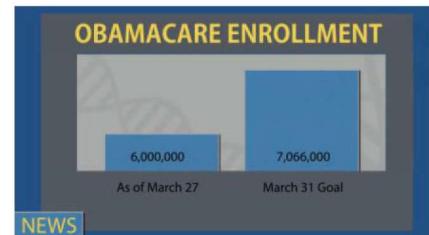
Using visualizations can unlock powerful data-pattern processing. However, a designer must be aware of several perceptual illusions that can lead viewers to map visual depictions back to their original numeric values incorrectly ([Huff, 1954](#); [Tufte, 1983](#)). If two plotted values have a 1:7 ratio, then the visualization should cause a typical viewer to see that 1:7 ratio veridically. Even for a precise visual channel such as position, this requirement can be tougher than anticipated. For example, see the dot plot and bar graph at the top of the second column of [Figure 2](#). The dot plot uses position as its visual channel, and the bar graph depicts the same data with both position and length. The second value appears to be roughly double the first value. Look more closely at the y-axis: The second value is only about 1% bigger than the first; the difference appears greater because the axis baseline does not start at zero. In theory, the data are transparently depicted—but in reality, such graphs are frequently misinterpreted.

[Figure 3](#) illustrates some real-world examples of this problem. The line graphs in the upper left, adapted from Darrell Huff's classic 1954 book *How to Lie With Statistics*, show how a line graph's scale can be stretched to make a trend appear steeper ([Huff, 1954](#)). In March 2014, a version of the bar graph at the upper right appeared on *Fox News*, a network with an avowedly opposite political orientation to Barack Obama, the U.S. president at the time. Around 6 million U.S. citizens had signed up for a new health-care program sponsored by the president, and the government specified a goal of 7 million sign-ups by March 31. Although the numbers presented are honest (a 6:7 ratio), the visualization's truncation of the y-axis tells a different story (a 1:3 ratio) to the viewer's visual system, suggesting a failure of the president's plan.

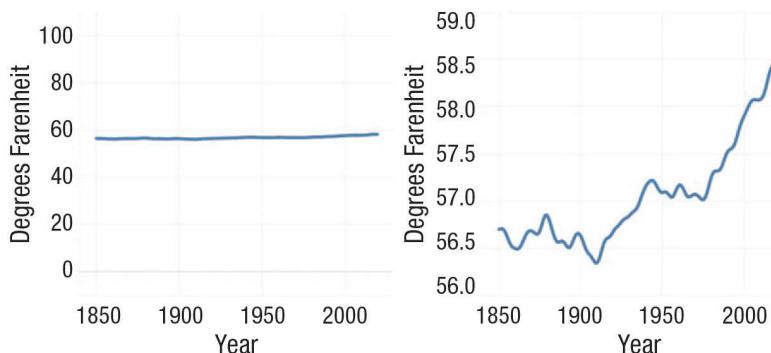
## The Difficulty of Mapping Numbers to Visual Channels in Honest Ways



Stretching the  $y$ -axis scale of the left graph drastically increases the slope of the perceived trend at right, which feels dishonest.



The same axis stretch in a bar graph, from a real-world example.



Here, taking a small visual increase (left) and stretching it (right) is the “honest” way to show climate-change data.



This color mapping similarly amplifies a relatively small temperature increase.

**Fig. 3.** Deceptive axis manipulations across a line graph (top left) and a bar graph (top right). So, should data always be plotted relative to zero? The graph on the bottom left depicts climate change by plotting temperature data from a baseline of 0 °F, yet most would agree that graph is less informative than the version to its right, which allows the differences in temperatures to be seen. At the bottom right of the figure, a cover of *The Economist* from September 2019 maps the same restricted range of data to the full range of a blue-to-red color scale. The graphs at top left are inspired by [Huff \(1954\)](#); those at bottom left are inspired by [Correll et al. \(2020\)](#). The magazine cover at bottom right is reprinted, with permission, from *The Economist* (September 19, 2019).

Will people not simply read the  $y$ -axis labels and easily overcome this initial misperception? Unfortunately, both real-world anecdotes and laboratory studies

suggest that this practice can be deeply misleading ([Correll et al., 2020](#); [Hofman et al., 2020](#); [Pandey et al., 2014](#)). For example, using a visualization similar to the *Fox News* example, researchers asked crowdsourced workers to rate the contrast between the two depicted values on a 1-to-5 scale. Ratings in a zero-baseline condition averaged around 1.5, whereas ratings in a deceptive-baseline condition averaged 2.8 ([Pandey et al., 2015](#)). In that study, stretching the *y*-axis also strongly affected ratings of the strength of the trend. The crowdsourced workers were not simply inattentive: Only participants who passed attention-check trials were included. Moreover, the deceptive effect persisted when participants were asked to type the numeric values represented by each bar before making their effect-size rating and were reminded of the *y*-axis's truncation, which was indicated by a "broken axis" symbol (such as that shown in [Fig. 2](#), top of second column) at the base of the *y*-axis ([Correll et al., 2020](#)).

The prevalence of this deceptive effect has led to quantitative prescriptions for how to set *y*-axis boundaries to produce accurate measures of statistical effect sizes by typical viewers ([Witt, 2019](#); [B. W. Yang et al., 2021](#)). For example, if the relevant data are far from zero, starting the *y*-axis at zero can make effect sizes illegible. One approach to increase legibility is to center the *y*-axis on the data's mean, then extend the *y*-axis 0.75 *SD* above and below the mean ([Witt, 2019](#)). This approach can be appropriate when the overall scale is not essential.

Visually conveying an appropriate difference often depends on what "appropriate" means. Many visualization designers subscribe to the principle that a *y*-axis must always start at zero so that the visually depicted ratios match the ratios in the data. By contrast, others argue that this guideline must be subject to context (for a distillation of arguments by designers, see [Correll et al., 2020](#)). For example, it seems clear that designers should consider the practically relevant range of data values when small but essential differences would be tough to see on a zero-baseline graph. A now-infamous (and now-deleted) story in the *National Review* used a graph similar to one at the bottom left of [Figure 3](#) to suggest that the recent global temperature rise is inconsequential ([Correll et al., 2020](#)). It even went beyond using a zero baseline for the axis, intriguingly using  $-10^{\circ}\text{F}$  as a lower bound. The

graphic to its right shows the pattern of data widely agreed by the scientific community to be more honest: a “hockey stick” pattern indicating a recent rapid rise.

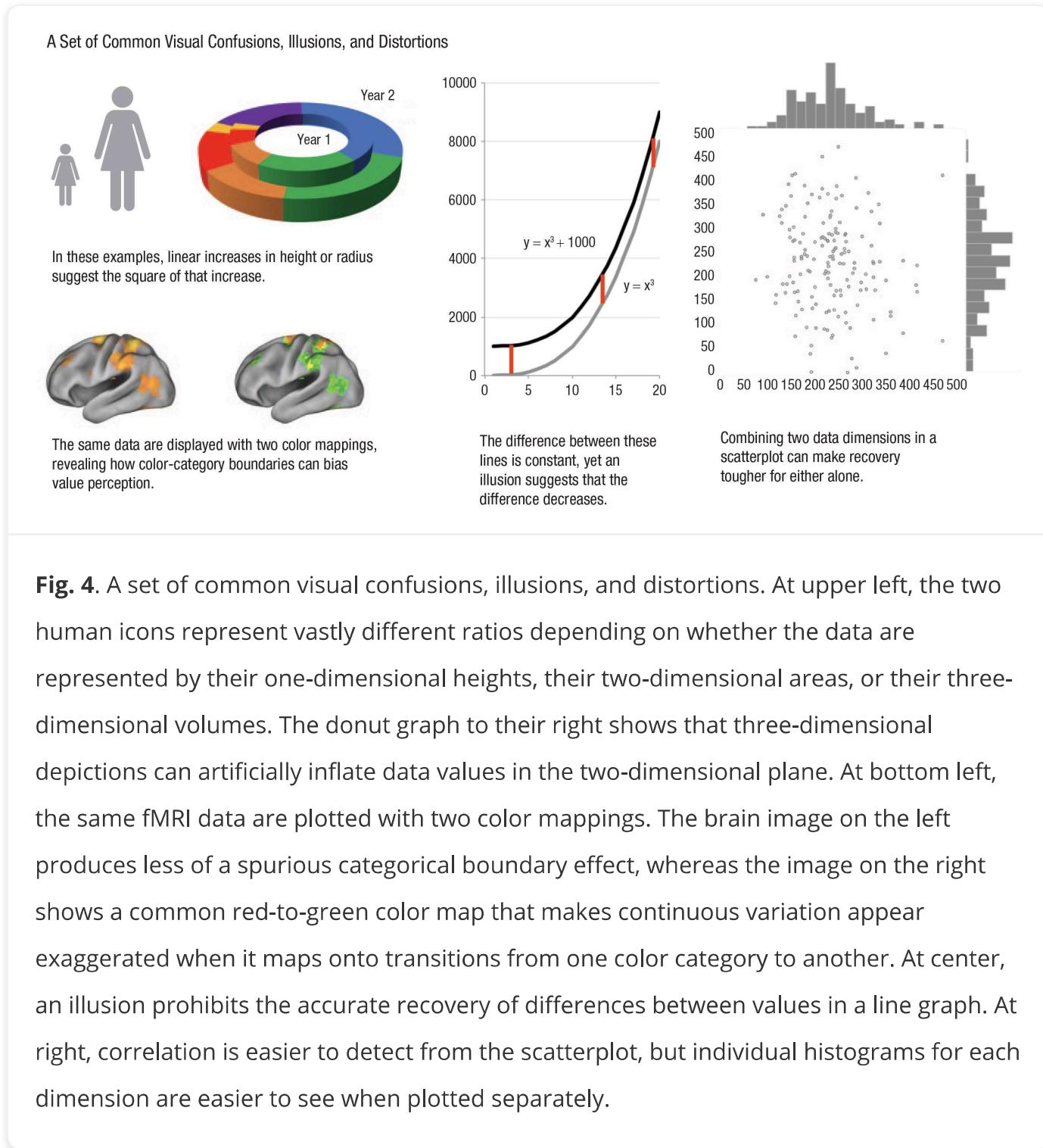
The image at the bottom right of [Figure 3](#) is a now-famous cover of *The Economist* magazine from September 2019, which similarly mapped a restricted range of data to the full range of a blue-to-red color scale. Another issue highlighted by these examples of temperature data is that not all variables have a single natural zero point. Zero degrees Fahrenheit is not a meaningful value for ratio comparisons. Temperature in Celsius has a somewhat meaningful zero point (the typical freezing point of water on earth), but that baseline is not relevant to climate change.

These mismappings of visually depicted ratios are not isolated to the position and length channels. One can imagine similar differences in interpretations arising for size or intensity. Imagine a map of a country with a circle over each major city. One might depict crime rates with the intensity of the circles’ color, but with what link to the original data? If zero were mapped to gray and 100 violent crimes per capita per year (a very high rate) to gray with a small amount of red, the country might seem safe regardless of the underlying data. Another depiction might link 100 violent crimes per year to a bright red. Side by side, those two depictions would suggest a huge difference in crime rates despite showing the same data. One could imagine the same trick, but this time mapping crime rates to the circles’ size. If 100 violent crimes were linked to a 1 mm circle, the country would seem safe, but if the same number of crimes were linked to a 10 mm circle, it would seem more dangerous.

### Which visual channel is linked to the data values?

In other cases, viewers can misunderstand which visual property encodes the data. In a classic example at the top left of [Figure 4](#), values are encoded by one-dimensional length (the height of each person), producing a 1:2 ratio of the two numbers. However, you might find that your estimate of the depicted values is determined instead by the area taken up by each person, leading to something closer to a 1:4 ratio (or even a 1:16 ratio, if the icons suggest three-dimensional

volume). People do indeed make this error, even when the numeric data are printed saliently near the visual representation ([Pandey et al., 2015](#)).



In the donut plot to the right of the human icons in [Figure 4](#), there are two ways that numbers are potentially miscommunicated. First, although the data are mapped onto angle, viewer judgments are more strongly determined by two-

dimensional area (with some potential contribution from arc length, which is tough to dissociate from area). Because the two donuts are different sizes, this means that the larger donut's value will be artificially inflated. Second, the graph is depicted in 3D. If the viewer can recover the actual three-dimensional geometry from the two-dimensional depiction, then the values should be accurately perceived. By contrast, if the viewer pulls values from the two-dimensional image (the amount of green or purple pixels on the screen), the values in the “front” will be inflated because of the perspective projection. Unfortunately, this technique is indeed substantially misleading because static two-dimensional projections do not typically lead to effective recovery of three-dimensional structures ([Tittle et al., 2001](#); but see [Brath, 2014](#)).

## Avoid common illusions and misperceptions

### A common optical illusion in line graphs

In the line graph in the middle of [Figure 4](#), the two curves are identical ( $y = x^3$ ), but the darker line is translated vertically upward by a constant of 1,000. Even if the viewer knows that the two shapes are identical, it is difficult to see that the vertical distance between the two lines is the same across their entire horizontal span. Instead, given any point on the dark line, viewers tend to see its distance from the closest point on the gray line, which becomes progressively smaller as both lines increase. This illusion makes it difficult to visually estimate differences between lines, especially lines with steep slopes ([Cleveland & McGill, 1984](#)). A similar example is depicted in the second column of [Figure 2](#). This illusion is well known to electrophysiology researchers: When faced with visualizing the difference between two measured waves, they will explicitly plot a “difference wave” that shows the difference as a single line ([Luck & Kappenman, 2012](#)).

### Illusory contrast effects for intensity

The bottom of the second column of [Figure 2](#) shows a final illusion that can warp our perception of visualized data. Both on the map and in the rectangle, the two vertically separated circles have the same luminance value. However, the lower

circle is subjectively darker to the eye because it has been placed on a lighter background and has a higher contrast with its surroundings. In the real world, converting luminance to contrast is a critically important mechanism for seeing accurate luminance and color despite changes in the brightness and color profile of light in the environment ([Purves et al., 2004](#)). However, this correction leads to misperceptions of intensity-coded values in the artificial world of data visualizations ([Szafir, 2018](#)). One rule of thumb is never to plot intensities on top of other intensities that vary, as in the map in [Figure 2](#).

### Misleading illusions that combine separate values

When plotting two sets of numbers on a map, designers typically rely on intensity for one set of values and map the other set of values to the area of circles. This design solution works because, rather than plotting one intensity on top of another intensity, which creates an *integral* representation of contrast, intensity and area are relatively separable representations ([Garner, 1974](#)). Other examples of integral representations include encoding two sets of data in rectangles—one set in their widths and one set in their heights. Instead of seeing these values separately, the eye is tempted to translate them into the aspect ratio and the area of each rectangle. The eye then focuses on the ratios and multiplication of each pair of values ([Shechter & Hochstein, 1992](#)). As an extreme case, it is unwise to attempt to use the red component of a single color (imagine using the RGB sliders to change the color of an object in presentation software) to depict one number and the green component for another. Red and green will combine in an integral fashion when both are at their highest value, and the viewer will see a single integral percept of yellow ([Ware, 2019](#)).

A final example is shown in the scatterplot in [Figure 4](#). Once two sets of numbers are combined into a single two-dimensional plot, new integral percepts emerge, such as the distance between any two points across both their x and y values, points that are outliers on both axes, or the global shape of all points that we can easily interpret as a correlation ([E. Yang et al., 2019](#)). However, there is a trade-off. The distribution of values of either set in isolation is now tougher to disentangle

([Mackinlay, 1986](#)). This is why data scientists often pair a scatterplot with “marginal histograms” that allow them to see those data in a separable way.

### The biasing effect of categorical perception

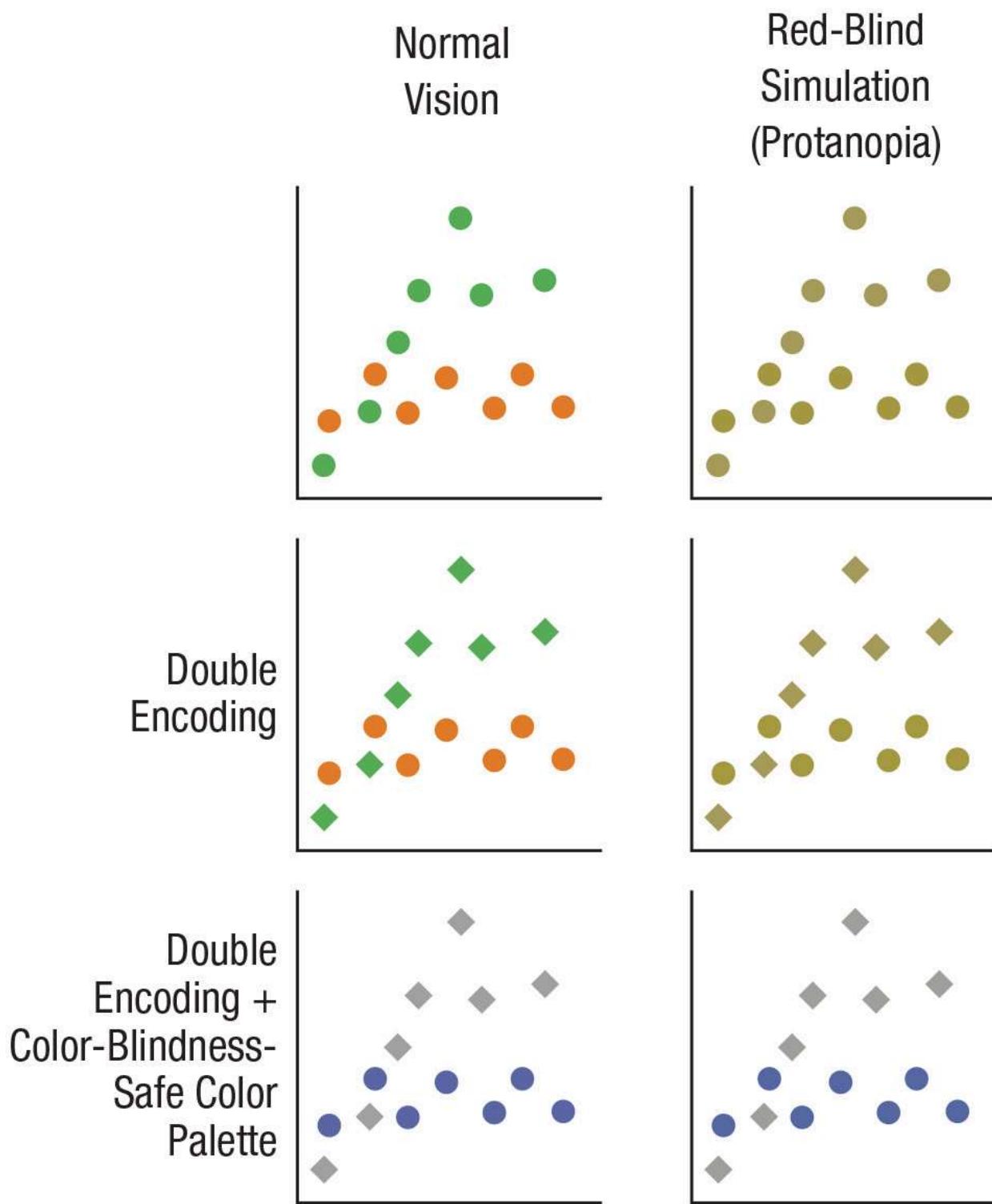
When continuous values are encoded through visual channels, those values can be warped by categorical perception. A classic example is the seven discrete colors that we see in a rainbow, which are not present in the rainbow itself. Those color categories are created by an automatic process that systematically bins continuous wavelengths into one of several perceptual categories, exaggerating metric differences among values that straddle those boundaries and shifting percepts toward prototypes for the categories ([Goldstone & Hendrickson, 2010](#); [Newcombe et al., 1999](#)). This same phenomenon occurs when data are depicted by hues, as in the bottom left of [Figure 4](#). Across the two brain images, the one on the right uses a greater variety of hues to depict activation values. These additional hues create new color-category boundaries that can dramatically exaggerate the differences between values that straddle them ([Y. Liu & Heer, 2018](#); [Quinan et al., 2019](#)). In *The Economist* cover shown in [Figure 3](#), the blue-to-red scale creates a salient categorical color boundary at the color transition point, which makes the temperature increase in the past few years especially salient. Similar category boundaries can affect the perception of values depicted by other channels such as position or length. In a pie graph or stacked bar graph, values that are near gridlines or the implicit 50% mark in the middle of the bar or pie are recalled as being farther from that category boundary ([Ceja et al., 2021](#); [McColeman et al., 2021](#); [Spence & Krizel, 1994](#); [Xiong, Ceja, et al., 2020](#)).

### Design for color-vision impairments

Beyond illusions and biases, certain combinations of colors in visualizations can be problematic for people who are color-blind or have other color-vision impairments. Color-vision impairments are estimated to affect 4% of the global population ([Olson & Brewer, 1997](#)), or roughly 300 million people. Further, older adults can have less sensitivity to color ([Silva et al., 2011](#)). Some color blindness results in viewers not being able to distinguish between various colors; protanopia, or red–green color

blindness, is the most common, but various other versions of color blindness exist. Color-vision impairments are highly problematic for visualizations, as a portion of the audience may literally not see important patterns in data.

Numerous online color-blindness simulators allow users to upload an image to learn how someone with color blindness would see it ([Asada, 2019](#)). For example, the first row in [Figure 5](#) shows a scatterplot encoded with two colors, green and orange. People with typical vision can see that the green dots have a steep positive correlation and the orange dots make a flat line. However, when the scatterplot is processed through a color-blindness simulator, the colors look the same, and all the dots appear to show a positive correlation. The simplest way to make visualizations accessible to viewers with color blindness is to avoid using hue as the only encoding channel or allow viewers to change the color palette ([Silva et al., 2011](#)). Designers can also double-encode a variable, using hue and another encoding channel ([Plaisant, 2005](#)), as in the second row of [Figure 5](#). The most thorough and inclusive option is to use color palettes that are safe for people with color-blindness, such as those proposed by [Harrower and Brewer \(2003\)](#), as seen in the bottom row of [Figure 5](#).



**Fig. 5.** Three ways to encode data for two groups in a scatterplot, as seen by observers with typical color perception and those with protanopia, a form of color blindness. Protanopia was simulated in Photoshop.

Design for perceptual accuracy across a broad array of tasks

The ranking of precision for visual channels depicted in the first column of [Figure 2](#) is based on a measure of precision in a particular task: the average error in verbal reports of the ratio between two depicted values (1:7 for each of the examples in the first column; [Cleveland & McGill, 1984](#); [Heer & Bostock, 2010](#)).

Although this task is surely important when comparing two values, it is not the only task (or perhaps even the primary task) that viewers complete when examining visualizations ([Bertini et al., 2020](#)). One proposed taxonomy of graph-interpretation tasks separated them into three levels: elementary, intermediate, and “overall,” roughly corresponding to simple fact retrieval, comparison and identification of trends, and gist understanding ([Bertin, 1981](#)). Additional taxonomies have been developed in the context of cognitive models of graph comprehension. Other taxonomies are based on analyses of cognitive processes involved in different tasks ([Tan & Benbasat, 1990](#)), reliance on local versus global features ([Carswell, 1992](#)), or other criteria ([MacDonald-Ross, 1977](#); [Washburne, 1927](#); [Wickens, 1989](#)). Work in the data-visualization literature has begun to catalog such tasks and to test the types of displays that best support performance for each. One list of tasks is distilled from questions submitted by students who were asked to analyze a large data set ([Amar et al., 2005](#)). This list includes analytic tasks such as “retrieve value,” “sort,” “determine range,” “correlate,” and “characterize distribution.” Another list of tasks focuses on how people generate summary statistics for plotted data ([Szafir et al., 2016](#)).

The ranking derived from performance on two-value ratio judgments does not always extrapolate across these alternative tasks. For example, depicting data with a line graph, which relies on “precise” position coding, can lead to *lower* efficiency in seeing big-picture statistical properties such as means. Intriguingly, the opposite is true for intensity coding. Multiple studies have shown that for identifying particular values, position is far more precise than intensity, but for judging an average across many values, intensity is more precise ([Albers et al., 2014](#); [Correll et al., 2012](#)). The reason for this dissociation is not well understood. It is possible that the lower precision of intensity better allows its values to blend together to construct aggregate values, or perhaps intensity is simply processed by a different

mechanism that affords aggregate judgments ([Szafir et al., 2016](#)). Some visualization designs purposely use the low precision of the intensity channel to focus viewers on the big-picture trend of the data instead being distracted by precise details, as in the visualization from the cover of *The Economist* magazine ([Fig. 3](#)), which focused viewers on the big picture of climate while de-emphasizing more detailed (but less relevant) variability in monthly weather.

Other work has tested whether people can more quickly or accurately complete the types of tasks listed above (judge mean, correlate, etc.) given different graph formats (scatterplots, bar graphs, tables, etc.) and data-set sizes (small vs. large; [Y. Kim & Heer, 2018](#)) or different arrangements of data within a graph ([Jardine et al., 2019](#); [Ondov et al., 2019](#)). Although such studies have found consistent interactions —better performance on Task X with Design Y in Arrangement Z for Data-Set Size S —the existing pattern of results is currently too complex to generalize to novel combinations. To create generalizable guidelines, the research community will likely require a more complex model of the underlying perceptual operations that produce these complex interactions ([Jardine et al., 2019](#); [Ondov et al., 2021](#)).

Among these alternative tasks, the perception of correlation is particularly well studied ([Harrison et al., 2014](#); [Jardine et al., 2019](#); [Rensink & Baldridge, 2010](#); [F. Yang et al., 2019](#); [Yuan et al., 2019](#)), leading to initial suggestions of its underlying perceptual operations. For example, scatterplots allow for more accurate judgments of correlation than pairs of bar charts presenting the same data ([Harrison et al., 2014](#)).

### Effectively distinguish among groups in data

Scatterplots present two metric variables in relation to each other. Imagine that the horizontal position represents the longitude of a collection of oil wells and the horizontal position represents their latitude. This example makes clear that scatterplots and data maps share DNA. Suppose the designer now wants to differentiate the set of points according to a *nominal* (categorical) variable, such as the company that owns the oil well. Because the two dimensions of position (vertical and horizontal) are already being used to represent the metric variables,

designers would typically differentiate points from each group with either different colors or shapes.

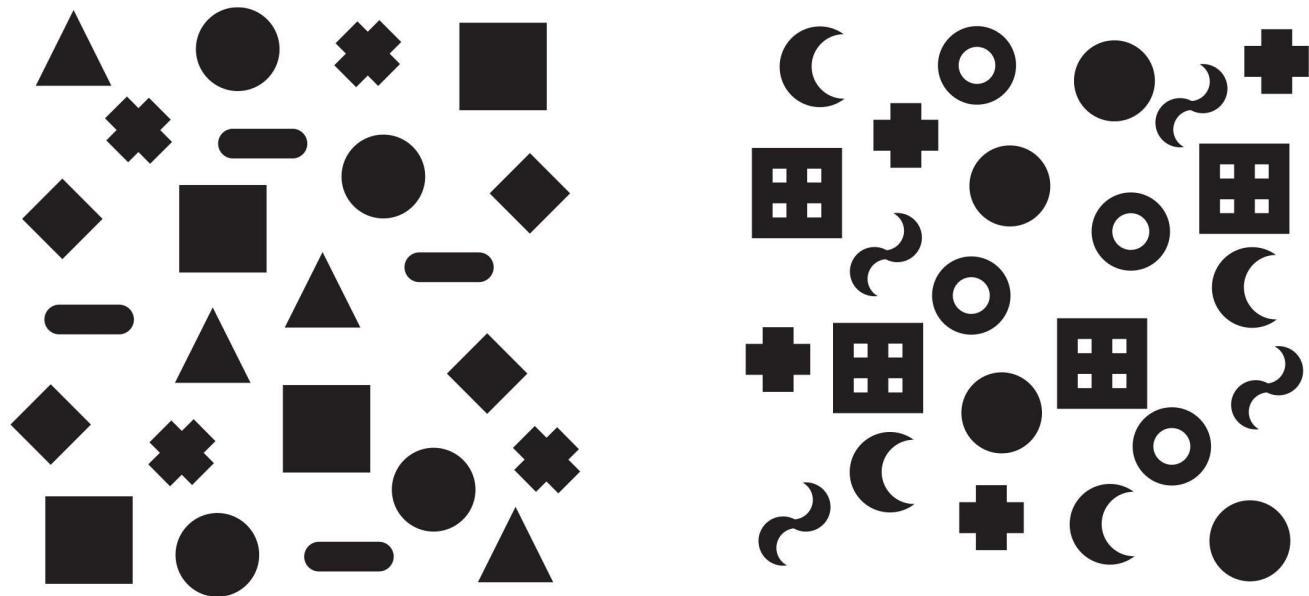
Plotting the groups as categorically different colors is often the first choice because the visual system processes color differences more efficiently than shape differences across the two-dimensional visual plane, as measured by performance in visual-search and texture-segmentation tasks ([Wolfe & Horowitz, 2017](#)). These findings, based on simple displays used in laboratory studies, extrapolated to a data-visualization context in which viewers compared the average heights of multiple color-coded or shape-coded clouds of points in a scatterplot: Color coding produced far better performance ([Correll et al., 2012](#)).

When color differences distinguish data from two separate groups or classes, differentiating those classes is easier if the encoded colors are farther apart in a perceptual color space. For example, it is easier to differentiate red from blue than from orange-red. Researchers have constructed effective palettes from perceptually informed color spaces (e.g., CIELAB; [Y. Liu & Heer, 2018](#)) to suggest colors to use to differentiate  $N$  nominal classes; colors become progressively crowded together as  $N$  increases. One such set that is viewable (and customizable) online is ColorBrewer 2.0 ([Brewer, 1994a, 1994b](#)). Another tool, Colorgorical, balances perceptual differentiation with aesthetic considerations ([Gramazio et al., 2016](#)).

Picking a color for a nominal value should also be constrained by the semantic congruence of the value and the color. If the nominal values are lemons and cherries, it is easier for viewers to answer questions about a chart that labels those values with yellow and red, compared with a standard palette that does not consider semantic congruence ([Lin et al., 2013; Schloss et al., 2018](#)). An algorithm can automatically generate intuitive color choices for a given noun by analyzing the color profile of images pulled from Web searches for the noun and then optimizing color assignments in terms of perceptual spacing and semantic fit. Such algorithms can perform as well as human experts in quickly picking color palettes for nominal data ([Lin et al., 2013; Setlur & Stone, 2015](#)).

Sometimes a visualization designer needs to show a second set of nominal values in the same plot. Thinking back to our map, imagine that we have used a unique color to represent each oil company (A, B, C) but also wish to depict each company's nationality (Canada, Brazil, Mexico). Typically, shape would be used to show that second nominal variable. Because it is less perceptually effective than color, it should be used for the less important variable, or the one with fewer values to differentiate. The shape sets used in commercial software (e.g., Microsoft Excel) gravitate toward intuitive shapes, such as circles, triangles, squares, and diamonds, that are not actually well separated in perceptual space.

Researchers have begun to explore perceptual shape space, and some have extrapolated from pairwise subjective similarity ratings of combinations of candidate shapes ([Demiralp et al., 2014](#)). Other work has relied on objective performance tasks on actual simulated scatterplots ([Burlinson et al., 2017](#); [Huang, 2020](#)). This work is also new, but so far, human shape space (at least for the simple shapes used in visualizations, and at least for the types of tasks tested so far) appears to prioritize the difference between open (circle, square, triangle) and closed (x, +, \*) shapes, such that differentiating points is easier when they differ in that property ([Burlinson et al., 2017](#); [Huang, 2020](#)). An initial full three-dimensional perceptual shape space ([Huang, 2020](#)) adds the additional properties of intersection and spikiness; [Figure 6](#) depicts a clear improvement in shape differentiability compared with the typical sets used even in professional data-visualization software.



**Fig. 6.** The standard shape set for Microsoft Excel (left) compared with a perceptually spaced set (right; inspired by [Huang, 2020](#)). Try to pick out the four instances of each shape in each display—you should find that task easier on the right side.

Finally, some software automatically differentiates nominal variables with both color and shape, under the assumption that more differentiation is better. However, work has shown that color is already so dominant in its effectiveness that *redundant encoding* does not substantially improve visual processing efficiency ([Gleicher et al., 2013](#)) unless the viewer has color-vision impairments or the viewer's task is exceptionally difficult ([Nothelfer et al., 2017](#)). Given anecdotal claims from some expert designers that redundant encoding can cause confusion in viewers, who typically expect color and shape to signal different nominal variables ([Tufte, 1983](#)), the lack of evidence for a perceptual advantage suggests that redundant encoding should be avoided in most cases, except when used to make visualizations accessible for viewers with color-vision impairments.

## How to Design a Perceptually Efficient Visualization

Use visualizations to allow viewers to powerfully compute statistics

One core advantage of visualizations is that they capitalize on our visual system's ability to extract information efficiently. The encoding properties in the first column of [Figure 2](#) are properties that we process in parallel across two-dimensional images ([Treisman, 1998](#); [Wolfe & Horowitz, 2017](#)). Faster than an eyeblink, we can pull statistics from data encoded as positions, one-dimensional lengths, two-dimensional areas, angles, or intensities. In the natural world, we typically use these statistics to recognize the type of scene we are in (e.g., a beach scene containing light brown, blue, and horizontal angles from the sand, sky, and horizon; [Oliva, 2005](#)) and to help provide a summary of an otherwise complex world ([Brady et al., 2009](#); [M. A. Cohen et al., 2016](#)). A substantial literature in visual cognition has examined the power and limits of this ability to explore what types of statistics can be extracted under what conditions (e.g., [Baek & Chong, 2020](#); [Haberman & Whitney, 2012](#)).

Much of this work in visual cognition relies on simplified laboratory displays of colored squares and circles, which are conveniently similar to the structure of the artificial worlds of data visualizations (for a review of insights from this literature for data-visualization design, see [Szafir et al., 2016](#)). The third column of [Figure 2](#) summarizes the types of statistics (minimums, means, maximums, and outliers) that can be pulled from any of the five visual properties in the first column. In the dot plot (encoding numbers as positions), a viewer can easily extract the minimum, mean, or maximum height. In the stacked bar chart below, the lighter segments present numbers through both their lengths and the positions of their tips because their bases are aligned. The darker segments are offset in arbitrary ways by the lighter segments underneath, so only their lengths visually represent values—yet the observer can still pick out minimum, mean, or maximum values quickly. The same is true for the bubble chart and the angles of the slope graph (a specialized line graph whose lines have only two points). Finally, for the heat map at the bottom that represents values as luminance contrasts, the lightest (minimum) and darkest (maximum) values are easy to pick out, and one can also quickly imagine the contrast of the mean value.

## Avoid a visual processing limit: making comparisons

Within their first glance at a data visualization, viewers can immediately pull general statistics from the positions, lengths, areas, slopes, and intensities ([Szafir et al., 2016](#)). This provides viewers with a starting point in understanding the distribution of the data and whether there might be outliers. However, the second step is to extract relations from the data by making a sequence of comparisons, a critical mental operation when viewing visualizations ([Franconeri, 2021](#); [Gleicher et al., 2011](#); [Tufte, 1983](#)). Comparisons include local comparisons between elements (“this point is higher than that point”; “this line segment is shallower than that one”), groups of elements (“these two bars have a greater range than those two”; “these red circles are on average larger than the green ones”), or global trends (“this section of the heat map is more saturated than that section”).

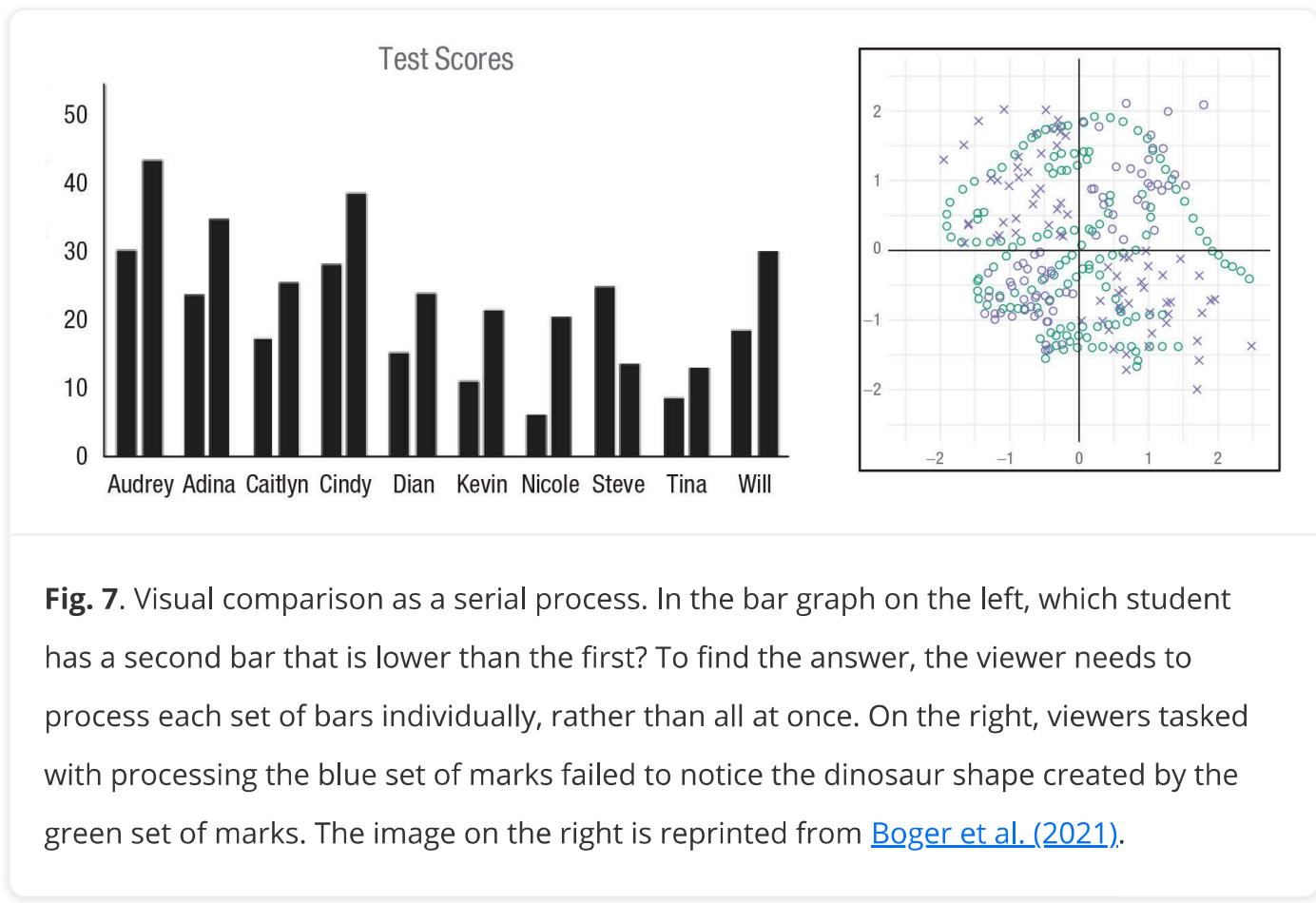
Each of those verbal sentences describes a single visual comparison that must be extracted serially, as part of a sequence ([Michal & Franconeri, 2017](#); [Roth & Franconeri, 2012](#)). These comparisons can each take hundreds of milliseconds to process ([Franconeri et al., 2012](#); [Logan & Compton, 1998](#); [Wolfe, 1998](#)), so that viewers tend to process only a handful per second ([Franconeri et al., 2012](#)).

Extracting the necessary multitude of comparisons from a visualization is therefore not like instantly recognizing a particular face, place, or Pokémon but instead more like the serial and controlled reading of a paragraph ([Carpenter & Shah, 1998](#); [Shah et al., 2005](#); see [Franconeri, 2013](#), for a discussion of the different types of attention demanded by these two scenarios). Each single comparison is limited in its complexity as well—by some estimates, encompassing interactions among approximately four variables at maximum ([Halford et al., 2007](#)).

This severe capacity limit has been confirmed for realistic comparison tasks in data visualizations. When viewers are asked to perform the types of tasks shown in the fourth column of [Figure 2](#)—for example, to locate pairs of graphed values in which the second value is bigger among pairs in which the second value is smaller—the time need to make each of those comparisons cumulates in painfully slow overall performance ([Nothelfer & Franconeri, 2019](#)). Likewise, the bar graph in [Figure 7](#)

(left) shows two test scores for each student, one before and one after an intervention. You can feel the sluggishness of those comparisons by answering the question: Who is the only student who got worse?

As a second example, when viewers are challenged to complete a tough comparison task with the blue dots of the scatterplot shown in [Figure 7](#) (right), they can fail to process the positions of the green dots, such that 93% fail to notice the presence of a dinosaur shape among them ([Boger et al., 2021](#)).



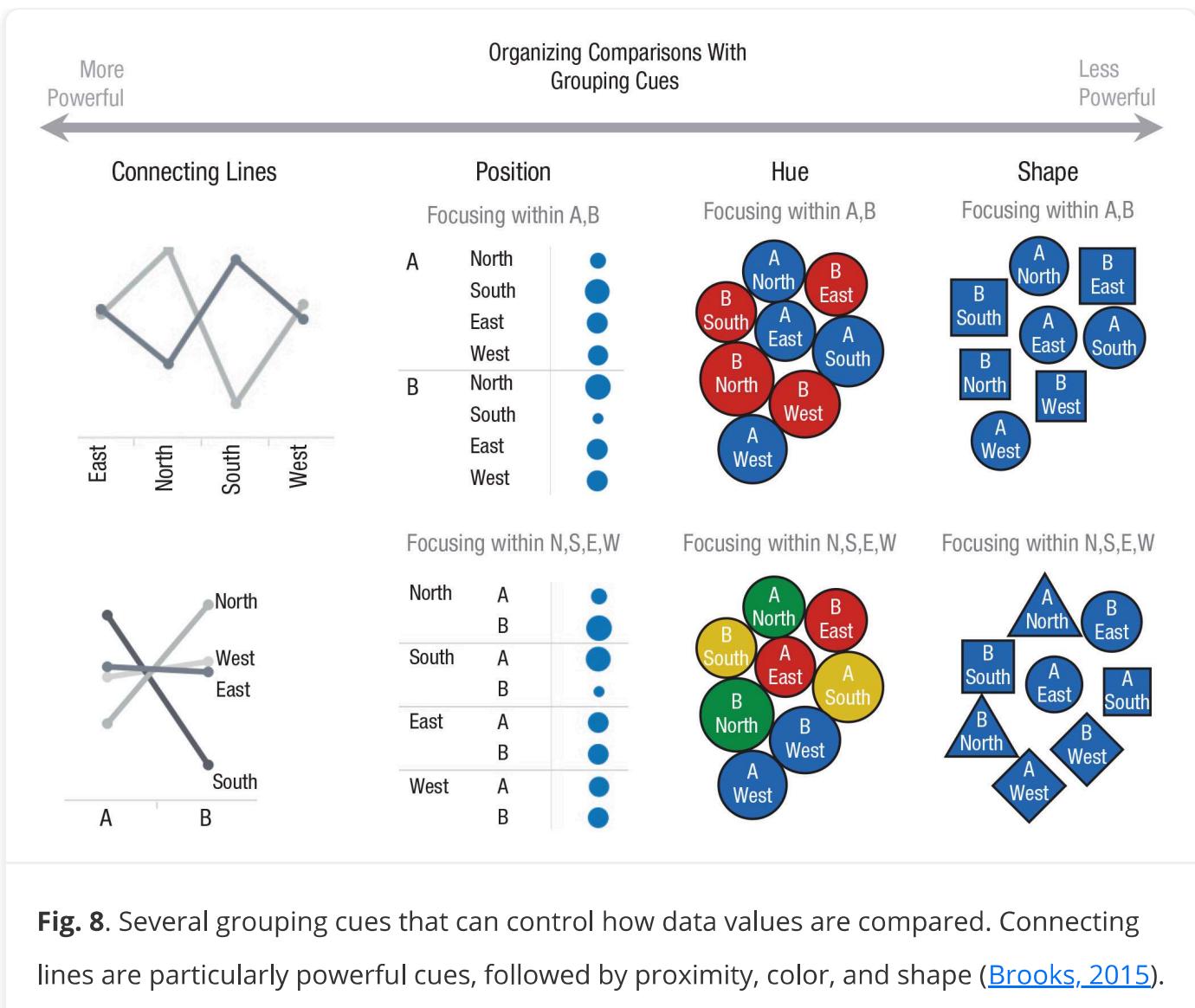
**Fig. 7.** Visual comparison as a serial process. In the bar graph on the left, which student has a second bar that is lower than the first? To find the answer, the viewer needs to process each set of bars individually, rather than all at once. On the right, viewers tasked with processing the blue set of marks failed to notice the dinosaur shape created by the green set of marks. The image on the right is reprinted from [Boger et al. \(2021\)](#).

A few hundreds of milliseconds of processing time for a single comparison is not noticed. However, across many comparisons, that time quickly adds up. Imagine a bar graph showing the performance of two groups, A and B, in both treatment and control conditions, for a total of four bars. Even in this small set, there are six possible pairwise comparisons, plus two main effects, plus multiple ways to look at interactions. In a graph with 12 bars, there are 66 pairwise comparisons alone!

When visualization designers create data-heavy displays that require fast processing, such as a business-metric-monitoring dashboard, many implicitly realize that comparisons are a sluggish visual operation. One common solution is to remove the need for the viewer to explicitly compare values by providing a second view of the data plotting the differences between values ([Fig. 2](#), upper right; [Few, 2009](#)). Known as *directly depicting deltas*, these dashboards commonly show differences between current financial values, such as revenue, and baseline financial values, such as revenue from the same month in the previous year, or between financial and budget values.

### Control comparison with visual grouping cues

A visualization that is designed to guide viewers to make the “right” visual comparisons can lead those viewers to make meaningful insights than they would gain on their own. One major cue for this guidance is to visually group the data values that should be compared with each other or should be compared as a group to other groups. The same classic grouping cues studied in the perception literature can control which values are selected and compared. [Figure 8](#) depicts four of the primary cues roughly in order of strength: connecting lines, position proximity, and similarity in either color or shape ([Brooks, 2015](#); [Palmer, 1995](#)). The two rightmost columns of [Figure 2](#) also show examples of grouping-defined comparisons.

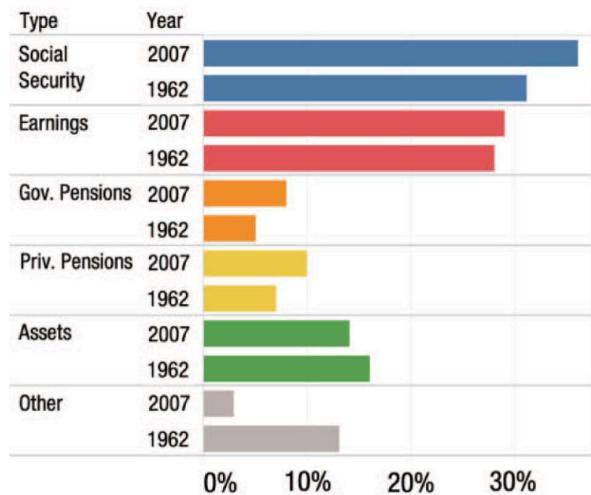
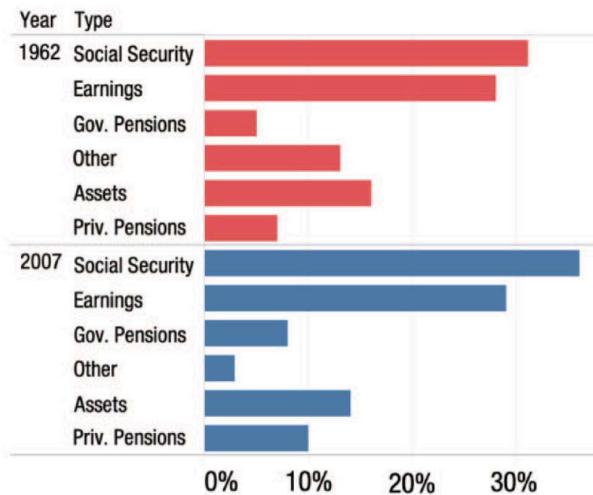


**Fig. 8.** Several grouping cues that can control how data values are compared. Connecting lines are particularly powerful cues, followed by proximity, color, and shape ([Brooks, 2015](#)).

Many of these grouping cues may stem from metaphors about real-world scenarios ([Tversky, 2001](#)). Objects that are spatially nearby are likely to stem from the same source, such as tomatoes from the same plant or birds nesting in the same nest. When three objects fall close to a common line, their spatial arrangement naturally conveys conceptual ordering because of how processes produce spatial ordering: If an animal leaves a trail, the order of the prints naturally reflects the timing of how they were laid down. Lines imply connections because connected things in the world tend to belong to the same objects, such as grapes connected by vines. Closed contours such as circles or blobs define objects because objects in the world tend to have closed, discernible borders or boundaries.

In line graphs, the choice of which variable is grouped by connecting lines (e.g., Fig. 8) has a profound impact on the interpretation of data. For example, one study presented viewers with graphs that showed the effects of two independent variables on a third variable. The data were depicted in line graphs; one variable was plotted as separate lines and the other was plotted as connected lines. Viewers could answer more sophisticated questions about the quantitative relationships depicted by the connected lines but only relatively simple questions about relationships between the separate lines ([Shah & Carpenter, 1995](#)). In many cases, viewers were unable to recognize the same data plotted with a different choice of how points were connected by lines. Relationships *between* values that straddle different lines or different panes need to be integrated with multiple comparisons, which requires controlled processing, taxes working memory, and introduces a risk of error.

The graphs at the top of [Figure 9](#) show how these grouping cues can control which comparisons are prioritized. In the example to the left, both proximity and color facilitate comparisons among categories (Social Security is highest) and overall between years (the yellow bars have a larger range than the green bars). However, you are less likely to compare a single category across the same year, which requires a jump of your eye to find the companion bar. The opposite is true of the example to the right ([Shah & Freedman, 2011](#); [Shah et al., 1999](#)). The word clouds at the bottom of [Figure 9](#) show an example of controlling comparisons with proximity grouping. In the word cloud on the left, the relatively weak color grouping makes it tougher to identify the common theme in each of five groups of words. On the right, identifying the common themes (restaurant, baseball, hands, etc.) is easier because the more powerful cue of proximity grouping has been added ([Hearst et al., 2020](#)).



**Fig. 9.** How visual grouping cues can control visual comparison. At top, a combination of color and proximity grouping lead the viewer to different visual comparisons across the two bar graphs. At the bottom, comparisons in a word cloud are weakly controlled by color grouping, and more strongly controlled with proximity grouping.

### Guide the viewer to the most important comparison

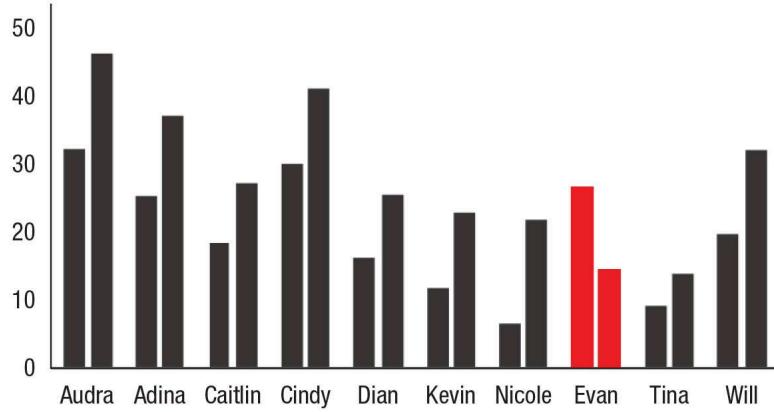
A good visualization relies on the grouping techniques described in the previous section, including connectivity and proximity, to help guide a viewer to compare one set of values or another. However, even within that one set, there are still many possible comparisons to make.

When multiple groups compete for comparison, that competition tends to be won by groups that are different or brighter in color, largest in size, or presented at the

top or left of a display. Such visual salience can be modeled by showing viewers pictures or visualizations, recording their eye movements, and then feeding the images and responses into computational models that predict human attention ([Bylinskii et al., 2017](#)). Many of these models exist for salience in natural scenes, ranging in complexity from simpler weighted linear combinations of relative differences in features (uniqueness in color, or orientation, at various locations and spatial scales) to more complex models that extract object contours or predict salience in movies ([Borji et al., 2013](#)). But many of these models fail to predict salience in the novel context of visualizations because the statistical profiles of those images differ substantially, containing large areas of blank space, text, axes, and titles ([Haass et al., 2016](#)). New models of salience for artificial information displays, trained on eye movements or on mouse-tracking data that are closely correlated with eye movements ([N. W. Kim et al., 2017](#)), can reach higher levels of predictive power ([Bylinskii et al., 2017](#); [Matzen et al., 2018](#)).

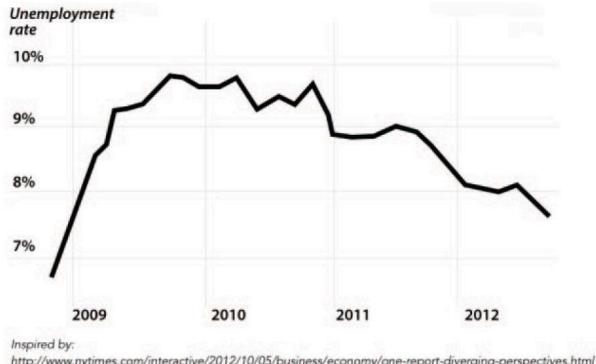
Visualization designers will often deliberately control visual salience to bring the viewer's eye straight to the critical comparison. One technique is to add salient color highlighting a single group of items ([Fig. 10](#)) to ensure that viewers process that comparison first ([Ajani et al., 2021](#); [Grant & Spivey, 2003](#); [Hullman & Diakopoulos, 2011](#); [Mayer & Moreno, 2003](#)). Some practitioner guides also recommend using color-coded accompanying text, as in the graph at the top of [Figure 10](#), to ensure that the viewer will match the pattern in the data to the relevant reference in the visualization designer's argument (e.g., [Knafllic, 2015](#)). More generally, a good visualization should place verbal information near relevant visual information so that viewers do not need to glance back and forth in a time-consuming search to see what text matches what visual pattern ([Moreno & Mayer, 1999](#)). These methods of guiding viewers to the most relevant comparisons are particularly important for low-knowledge readers, who benefit from visualizations that present or highlight only the relevant comparisons ([Canham & Hegarty, 2010](#)). These techniques are less important for experts, who rely on prior experience to guide their attention.

### One Student Got Worse



### Unemployment is higher than stated goals

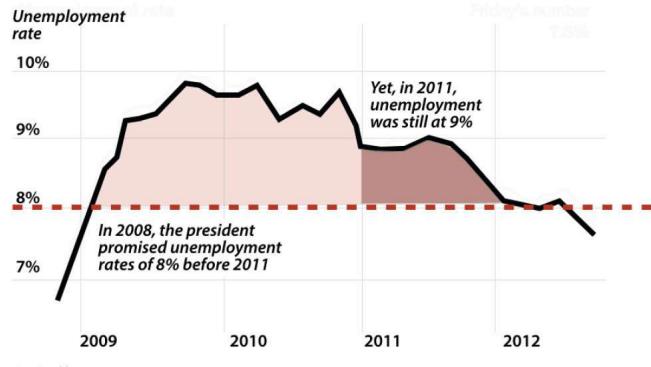
In 2008, the president promised unemployment rates under 8% before 2011. Yet, in 2011, unemployment was still at 9%



Inspired by:

<http://www.nytimes.com/interactive/2012/10/05/business/economy/one-report-diverging-perspectives.html>

### Unemployment is higher than stated goals



Inspired by:

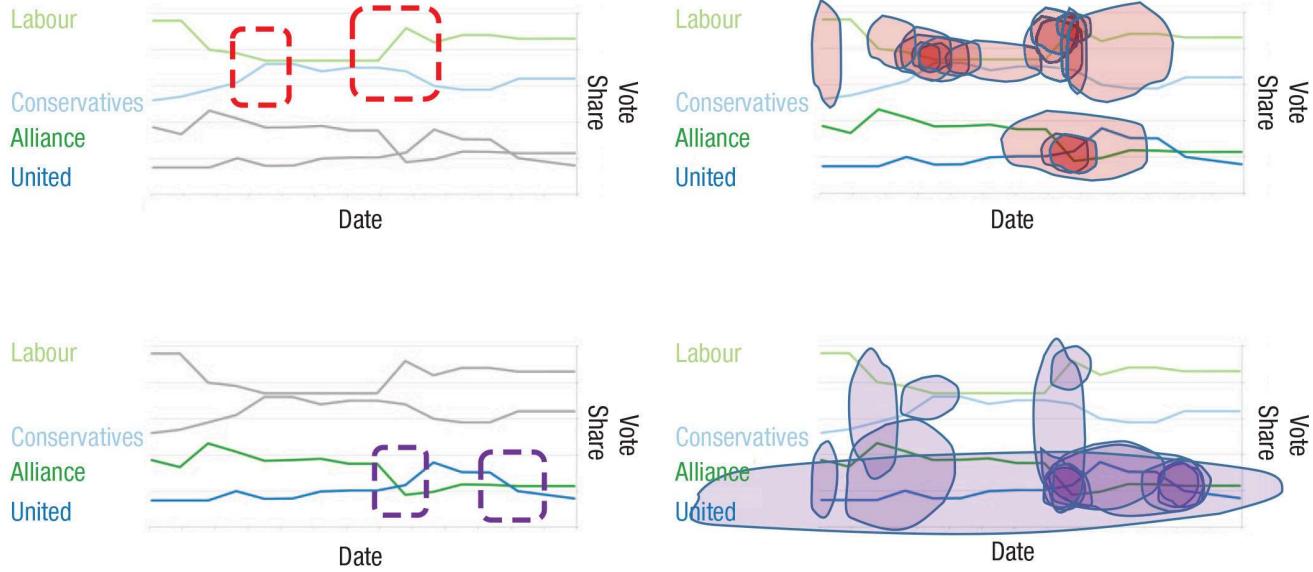
<http://www.nytimes.com/interactive/2012/10/05/business/economy/one-report-diverging-perspectives.html>

**Fig. 10.** Color highlighting and direct annotation to help viewers make the right comparison first and know what conclusion is supported by that pattern in the data. The graphic at the top illustrates a color-highlighting technique suggested in business-oriented practitioner guides (e.g., [Knafllic, 2015](#)). The graphs at the bottom (inspired by [Bostock et al., 2012](#)) are an adaptation of a graph by data journalists using grouping, highlighting, and verbal annotation.

Some research has studied the techniques of data journalists, who are tasked with clearly communicating data-based arguments to nonexperts. This work has cataloged techniques used by news outlets such as *The New York Times* Upshot, *The Washington Post*, *The Economist*, and FiveThirtyEight ([Hullman & Diakopoulos, 2011](#); [Hullman, Diakopoulos, & Adar, 2013](#); [Segel & Heer, 2010](#)) and has used computation to automatically generate visualizations that use particular strategies ([Gao et al.,](#)

[2014](#); [Hullman, Drucker, et al., 2013](#); [N. W. Kim et al., 2017](#)). These outlets employ specially trained reporters who convey data-based political, economic, health, and science stories to the lay public. They will often show a single pattern in the data at a time, relying on the viewer to step or scroll ([Amabili, 2019](#)) through a sequence of patterns. In each step, a pattern of data is highlighted to ensure that the viewer's limited processing capacity is directed to the values simultaneously described in a verbal annotation. The bottom row of [Figure 10](#) illustrates how data journalists might redesign the example on the left, which requires viewers to navigate text that is placed far from the patterns that it describes and to use their imagination to fill in the patterns referred to by the text. The example on the right addresses both of these issues, leading to more effective communication of a key data pattern ([Ajani et al., 2021](#)). The two rightmost columns of [Figure 2](#) summarize these techniques, showing that grouping, highlighting, and annotating can help viewers quickly make the right comparisons.

If helping viewers prioritize critical comparisons in a visualization is so important, then why do so few presenters do it? One likely reason is that presenter have a *curse of knowledge*—an inability to simulate the perspective of the naive viewer because they cannot ignore what they know and see ([Birch & Bloom, 2007](#); [Camerer et al., 1989](#)). [Figure 11](#) presents an empirical demonstration of this curse from a lab study ([Xiong, Van Weelden, & Franconeri, 2020](#)). Participants heard an intriguing story about a dip and rise of a political candidate's popularity in the polls (the top line in the top left graph); the story made those patterns stick out to the participants. They were then told to forget the story and predict what patterns another person, naive to the story, would find interesting or salient in the graph. The graph at the upper right presents their collective predictions—and makes it clear that people think that others will see what they see, even when they know that others do not have the same expertise. The graphs in the bottom row show how, when the original story concerned a different candidate, the pattern of results changed accordingly. It is impossible to “turn off” your own expertise, which makes it difficult to see through the eyes of nonexperts.

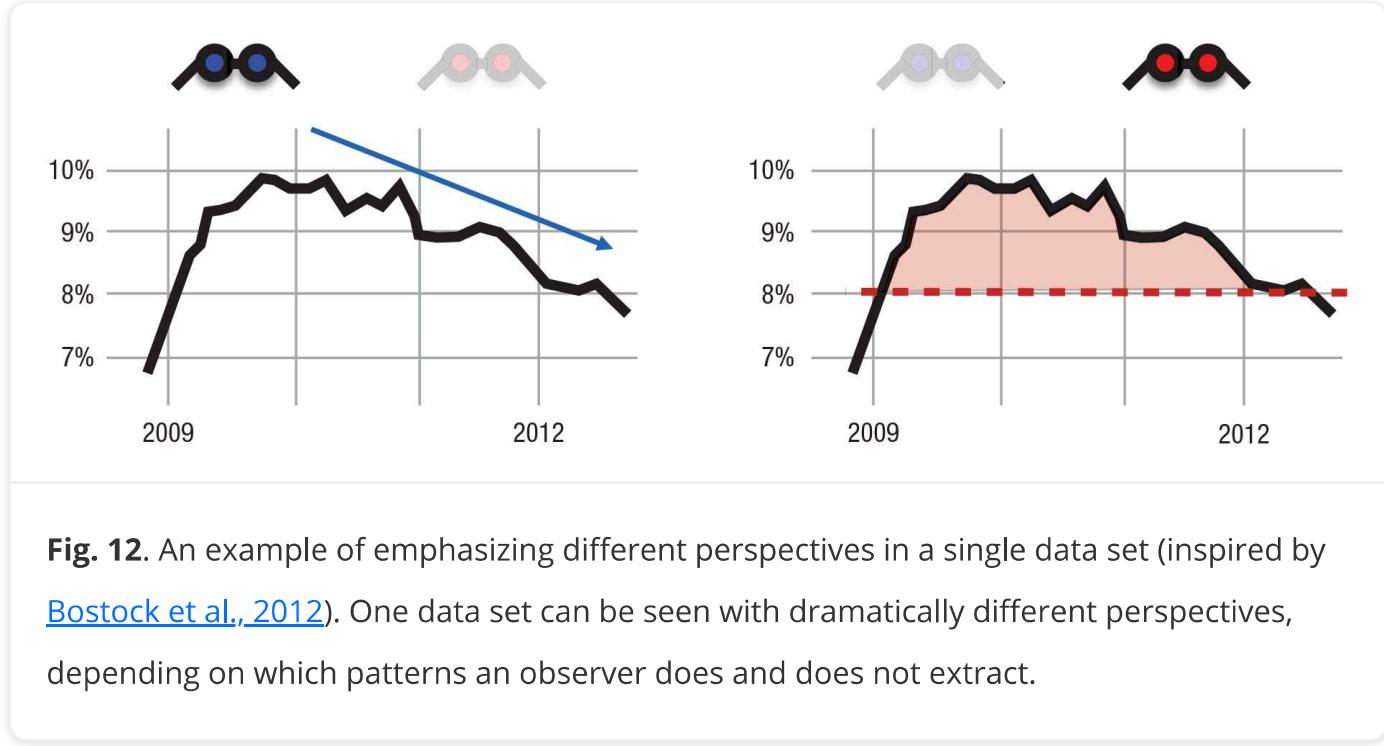


**Fig. 11.** A demonstration of the curse of knowledge in data visualizations. Participants were told a story involving the data patterns highlighted in the top two lines (top left) or the bottom two lines (bottom left) of a line graph. They were then told to forget the story and to circle the patterns that a naive viewer would notice first on an unannotated version of the graph. Their predictions (bottom right) mirrored the story they had been told, showing a “curse of knowledge,” or an inability to inhibit relevant expertise. Adapted with permission from [C. Xiong, L. Van Weelden, and S. Franconeri \(2020\)](#), “The Curse of Knowledge in Visual Data Communication,” *IEEE Transactions on Visualization and Computer Graphics*, 26(10), p. 3051–3062, <https://doi.org/10.1109/TVCG.2019.2917689>. Copyright 2020 by IEEE.

## Persuade with visualizations

If grouping, highlighting, and annotation can guide viewers toward a certain comparison or pattern, then for better or worse, a visualization designer has some control over what their viewers see in a data set. [Figure 12](#) shows an example, adapted from *The New York Times*, in which the same unemployment data might be seen—or designed to be seen—in different ways. During Barack Obama’s presidency, a member of his party (the blue glasses) might see a drop in unemployment rates and highlight that downward-sloping pattern in the orientation channel. But a member of the rival party (the red glasses) might see a

failure to meet a goal of 8% unemployment and emphasize the information in the area channel by highlighting the area under the curve ([Bostock et al., 2012](#)). These visual manipulations are similar in spirit to choosing which of those features to highlight in a verbal argument.



An analysis of distortions and biased annotations in news media visualizations showed that rhetorical biases are pervasive in data graphics and that labels and framing may be skewed toward progressive or conservative positions, depending on the news outlet ([Mehta & Guzmán, 2018](#)). A taxonomy of “visualization rhetoric” likens visualization design to an editorial process in which decisions about what data to include, how to encode them, how to use titles and labels, how to describe the data’s provenance, and what interactions to allow represent rhetorical choices aimed at guiding viewers toward preferred interpretations ([Hullman & Diakopoulos, 2011](#)). People’s perceptions of a visualization’s message, and their ability to recall it, are particularly influenced by the visualization’s title ([Kong et al., 2018, 2019](#)). For an engaging tour of the complexities of truth and deception in visualization with an emphasis on a journalism perspective, we point the curious reader to the books of [Alberto Cairo \(2016, 2019\)](#).

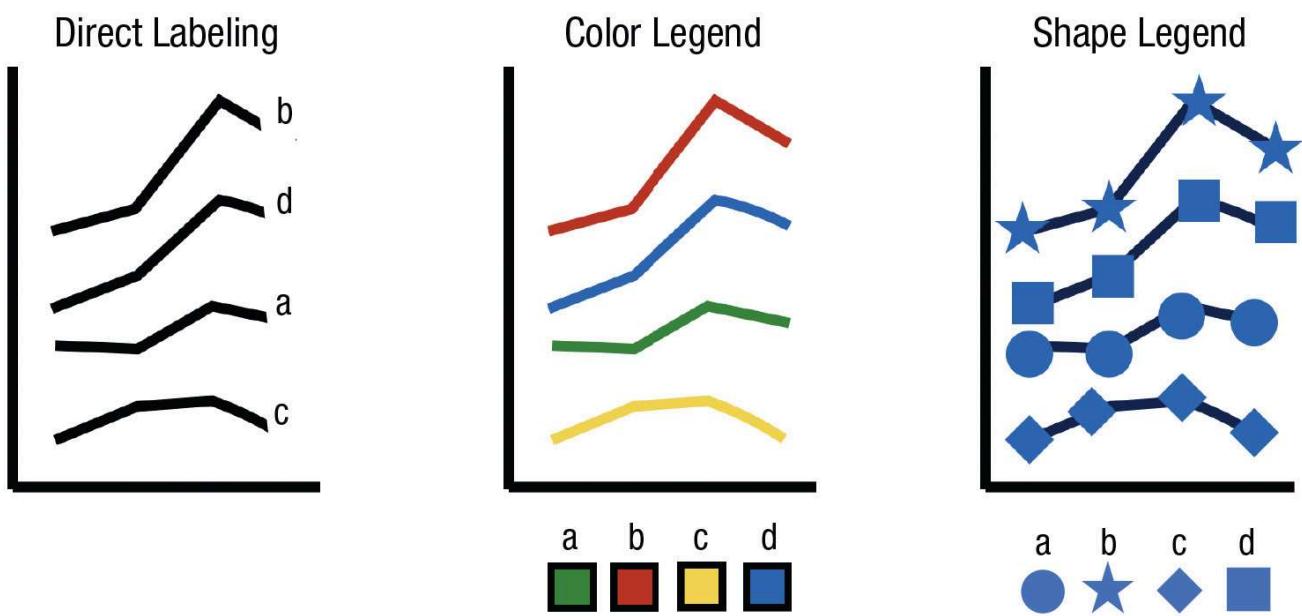
Viewers can also be influenced by expectations or social pressures, even for relatively low-level visual judgments. For example, when participants made judgments about correlations in scatterplots, their estimates were higher when the data were labeled as personality variables, which one would expect to be correlated, than when the data were unlabeled ([Freedman & Smith, 1996](#)). In another study, participants were asked to make judgments about visualizations (e.g., rating the linear association in a scatterplot) presented either alone or with a histogram plotting other people's ratings. The other ratings were either true ratings or a distribution of the true ratings shifted by 1 *SD*. When participants were provided with the manipulated histograms, their judgments were biased in the direction of the social influence ([Hullman et al., 2011](#)).

Finally, the format of a visualization can also guide the types of conclusions that viewers draw from the underlying data. Imagine data showing that students who eat breakfast more often tend to have higher GPAs. A viewer might see this correlation and assume a causal relationship whereby a good breakfast causes better grades. Although plausible, this conclusion cannot be drawn from these data. When shown visualizations like these, viewers made unwarranted claims about similar correlational data, and they did so more often when the visualizations aggregated the data into fewer groups (e.g., a two-bar graph), compared with more groups (e.g., a scatterplot showing all of the individual data values; [Xiong, Shapiro, et al., 2020](#)), perhaps because seeing the data in fewer groups is implicitly associated with those data being gathered by an experimental manipulation.

### Avoid taxing limited working memory

Given that comparisons are already highly capacity limited, any extraneous demands on working memory due to the design of visualizations should be avoided. Interpreting the graphs in the middle and right sides of [Figure 13](#) requires individuals to map the symbols and colors in the graphs to their referents in the legends below. This task is highly demanding of limited working memory resources. If information is lost in interpreting a graph, viewers might make interpretation errors or require extra time to reinspect the legend. Indeed, one study found that

people answered questions about data both faster and more accurately when the data were directly labeled in graphs compared with when there was a legend ([Lohse, 1993](#)). Therefore, instead of legends, use direct labels whenever possible (e.g., [Wong, 2010](#)). Note that there may be exceptions to this recommendation, if legends provide a way to index data values that labels cannot. For example, a map of an amusement park might list locations under the map to allow the viewer to browse the locations alphabetically or clustered by type (rides, food, etc.), which is not possible for map locations that are already organized in space.



**Fig. 13.** A demonstration of the advantage of direct labels over legends. Take a moment to state the names of the four groups shown in the line graph at left in top-to-bottom order. (Answer: b, d, a, c.) Now do the same for the graphs at center and right, which require coordination with color and shape legends. You should notice a substantial slowdown because of the need to frequently look back and forth between the graph and the legend. If you attempt to memorize the legend first, you will experience the capacity limit of your working memory.

When visualizations are well designed, they can help viewers overcome working memory limits by offloading information storage and some types of processing to a

page or screen ([Kirsh, 2005](#); [Tversky, Heiser, et al., 2002](#)). Compared with the slow difficulty of reading and comparing symbolic numbers, a visualization can allow these steps to unfold far more quickly and efficiently. For example, older adults with normal cognitive decline were asked to compare multiple health-care plans. They were given plan information (e.g., about monthly premiums, deductibles, and gap coverage [supplementary insurance to cover medical costs incurred before reaching the deductible]) either in a table full of text and numbers or in a table with visual categorical encodings of the information (e.g., green circles for the best gap coverage and red circles for no coverage). The adults using the visually encoded table made better and, in some cases, faster decisions about which health-care plan to select because they found it easier to make comparisons ([Price et al., 2016](#)).

### Beware the working memory load of animation

Limits on working memory can also be strained by animation. Some kinds of visual motion, such as patterns of translation or expansion that accompany moving the head or walking, can be tracked efficiently and automatically ([Gibson, 1979](#)). However, our capacity for tracking the motion of objects that move in arbitrary directions is highly limited, to as few as one or two objects at a time ([Alvarez & Thompson, 2009](#); [Scimeca & Franconeri, 2014](#); [Xu & Franconeri, 2015](#)), including in moving scatterplots ([Chevalier et al., 2014](#)). If a viewer's capacity is overwhelmed during an animation, they may not retain the motion information before the animation is over. Education researchers have termed this the *transient information effect*, a "loss of learning due to information disappearing before the learner has time to adequately process it or link it with new information" ([Sweller et al., 2011](#), p. 220).

Examples of both of these problems can be seen in studies of mechanical diagrams. In a review of roughly 100 studies on the use of animated diagrams for teaching complex mechanical, biological, or computational systems, researchers found that students' descriptions of processes were no more accurate with animations than with labeled static diagrams ([Tversky, Morrison, & Betrancourt, 2002](#)). In one experiment, students saw static or animated diagrams of the mechanical processes

involved in flushing toilet tanks. Both groups could identify the number of sequential stages. However, students who saw the animated diagram made more errors (20% for animated vs. 5% for static) about individual stages, such as whether air or water ends the flushing process ([Kriz & Hegarty, 2007](#)). Even when viewers can extract global patterns from an animated diagram, the capacity limitation in processing animations can generate a cost for other information in the scene.

The mere presence of animation can also induce an illusion of understanding, such that it erroneously inflates observers' confidence in their percept. For example, observers who saw animated diagrams of a toilet's flushing mechanism were not only less accurate at recalling the names and functional roles of parts relative to observers who saw static diagrams; they also reported less perceived difficulty and higher engagement than did the more successful static-diagram learners ([Paik & Schraw, 2013](#)).

There is little evidence that animation facilitates understanding of information displays, but there is one important exception: animation used to convey probabilistic processes and uncertainty, in which draws from a distribution provide a metaphor for random sampling ([Hofman et al., 2020](#); [Hullman et al., 2015](#); [Kale et al., 2019](#)). Such animations are effective in conveying sampling uncertainty. One reason appears to be automatic (i.e., not requiring conscious attention) processing of frequency information ([Hasher & Zacks, 1984](#)). Using animation to convey uncertainty does not tax working memory because neither the sequence of samples nor the specific properties of individual samples are important for understanding that variability.

Allowing a viewer to control an animation manually may prepare them to focus their limited capacity on the right subsets of information at the right time and to replay critical portions of an animation. Viewer-controlled animation has shown some success ([ChanLin, 1998](#); [Faraday & Sutcliffe, 1997](#); [Mayer & Moreno, 2003](#); [Schwan & Riempp, 2004](#); [Tversky, Morrison, & Betrancourt, 2002](#)). However, empirical evaluation shows that interactivity does not always improve performance. One famous example of an animated data visualization is Gapminder's Trendalyzer ([Gapminder Foundation, 2007](#)), a scatterplot containing circles for countries and

plotting, for example, the countries' gross domestic product (GDP) on the y-axis, life expectancy on the x-axis, and population as the size of their circle. The Trendalyzer uses animation to show trends over time, moving the dots to show how each of these factors changes. When participants were shown a similar display with interactive controls for the animation, they were better able to answer precise questions about each country than participants without interactive controls, but they also became worse at extracting global trends across countries compared with participants shown static alternatives ([Abukhodair et al., 2013](#); [Robertson et al., 2008](#)). Furthermore, user-controlled interactivity is not always possible for some systems or audiences.

Nevertheless, viewers frequently report that animated diagrams and data visualizations are more engaging and enjoyable than static versions, which may explain animation's continued use ([Abukhodair et al., 2013](#); [Robertson et al., 2008](#); [Tversky, Morrison, & Betrancourt, 2002](#)). In a communication context, anecdotal evidence suggests that an engaging animation can still communicate patterns in data when carefully deployed. The example in [Figure 14](#) depicts a TED talk by [Hans Rosling \(2006\)](#), in which Rosling used the moving display to depict changing world health statistics over time. Rosling made his dynamic data story easy to understand by carefully using language to guide his audience ("look at this cluster, it's moving up . . ."), paired with exaggerated gestural cues to help the audience focus on the relevant data values. Such cues have been shown to help students integrate verbal information when interpreting diagrams ([Mautone & Mayer, 2007](#)), graphs ([Michal et al., 2018](#)), animations ([de Koning & Tabbers, 2011](#)), and other educational materials ([Goldin-Meadow, 1999](#)). Future research inspired by Rosling's case study might help outline concrete rules for using animation in ways that would allow his success to be replicated.



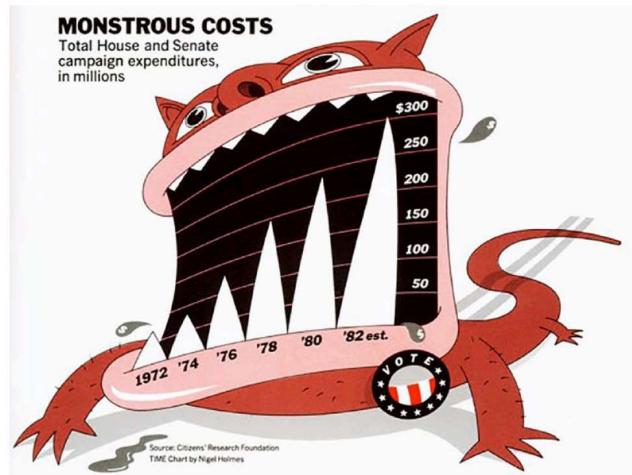
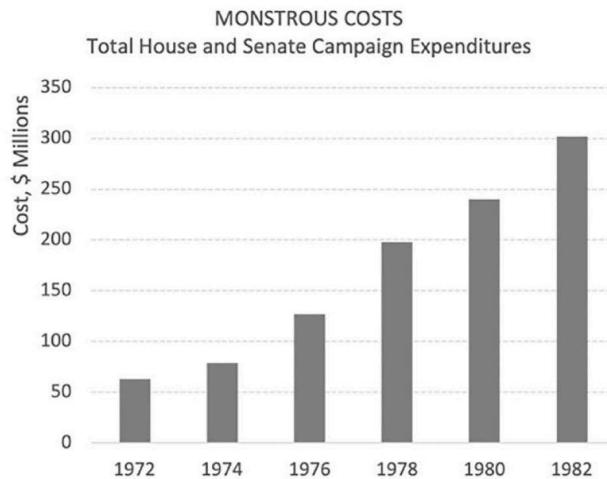
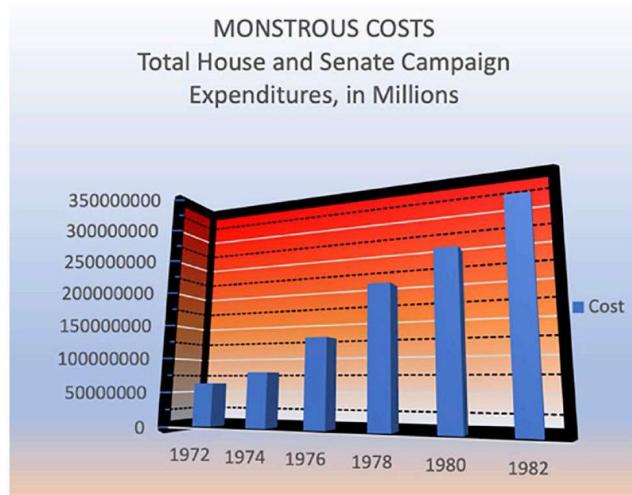
**Fig. 14.** A screenshot from [Hans Rosling's \(2006\)](#) TED talk on the power of visualized data.

Hans Rosling helped viewers see relevant patterns in a complex animated visualization with exaggerated gestures and clear linguistic guidance toward the critical visual comparisons that supported his arguments.

Should your visualizations be rich or minimalistic?

Given the limited working memory resources of visualization viewers, designers often recommend a minimalist aesthetic that strips away any design elements that are not critically needed ([Few, 2004](#); [Knafllic, 2015](#); [Tufte, 1983](#)). One popular mantra is to keep a maximal “data–ink ratio” ([Tufte, 1983](#)), though the definition of “ink” can be frustratingly vague ([Correll & Gleicher, 2014a](#)). [Figure 15](#) depicts variants of a visualization based on this prescription. The visualization at the top is filled with “clutter”: grid lines, a background pattern, and varied colors across the bars. The

middle image is a “decluttered” visualization that omits these elements. Although newer editions of Microsoft Excel have eliminated some forms of clutter shown in the top image, critics of the 2007 edition argued that the software encouraged users to create graphs containing dense grid lines, unnecessary labels, unneeded color variation, and even three-dimensional effects that transformed simple bars into cylinders or pyramids ([Kirk, 2012](#); [Kosslyn, 2010](#); [Su, 2008](#); [Ware, 2010, 2019](#)).



**Fig. 15.** A “cluttered” visualization (top), a minimalist “decluttered” version (middle), and a version that incorporates pictorial embellishment (bottom). The graph at the bottom was created by Nigel Holmes for *TIME Magazine* and was reprinted in his [1984](#) book, *Designer’s Guide to Creating Charts & Diagrams*. Used with permission.

Despite strong calls to declutter visualizations (e.g., [Tufte, 1983](#)), there is only mixed evidence that this practice improves aesthetic ratings and little evidence that the prescription affects objective performance. Several studies have measured aesthetic ratings for cluttered versus decluttered charts, and some have shown clear preferences for decluttered versions ([Ajani et al., 2021](#)) and others, surprisingly, showing the opposite ([Hill et al., 2017](#); [Inbar et al., 2007](#)). Researchers who have found the opposite have typically argued either that viewers' higher level of familiarity with cluttered charts make those charts more attractive or that decluttered charts that are *too* minimalistic become boring. Another possibility is that users may prefer particular depiction styles for particular purposes, mindful of their audience and goals ([Levy et al., 1996](#)). Objective performance measures, such as the speed with which viewers can compute means across values in a bar graph, also present mixed evidence. For example, that speed can be slightly faster when some forms of "clutter," such as axis tick marks, are removed but slower when other elements are removed ([Gillan & Richman, 1994](#)). Given the large number of design elements that could count as clutter, combined with the large number of tasks that one could complete on a visualization, some have argued that a simple rule for whether to declutter is unlikely to arise and have discouraged further empirical testing given the small and mixed effects found so far ([Ajani et al., 2021](#)).

The bottom of [Figure 15](#) shows the same bar graph embedded within a cartoonish monster. The addition of such pictorial elements that embellish data with anthropomorphic or metaphorical elements—intended to enhance engagement or memory—has been demonized as “chartjunk” (e.g., [Tufte, 1983](#)). Various studies have shown that adding these elements leads to no improvement in memory for the data ([Helgeson & Moriarty, 1993](#); [Kelly, 1989](#)), mixed results depending on the details of the task and context ([Gillan & Richman, 1994](#); [Li & Moacdieh, 2014](#)), or better memory for the data content or message ([Bateman et al., 2010](#); [Borkin et al., 2016](#); [Haroz et al., 2015b](#)). Like animation, these visual embellishments can increase ratings of engagement and aesthetic value ([Li & Moacdieh, 2014](#)). And despite mixed evidence as to whether their presence improves memory for the data,

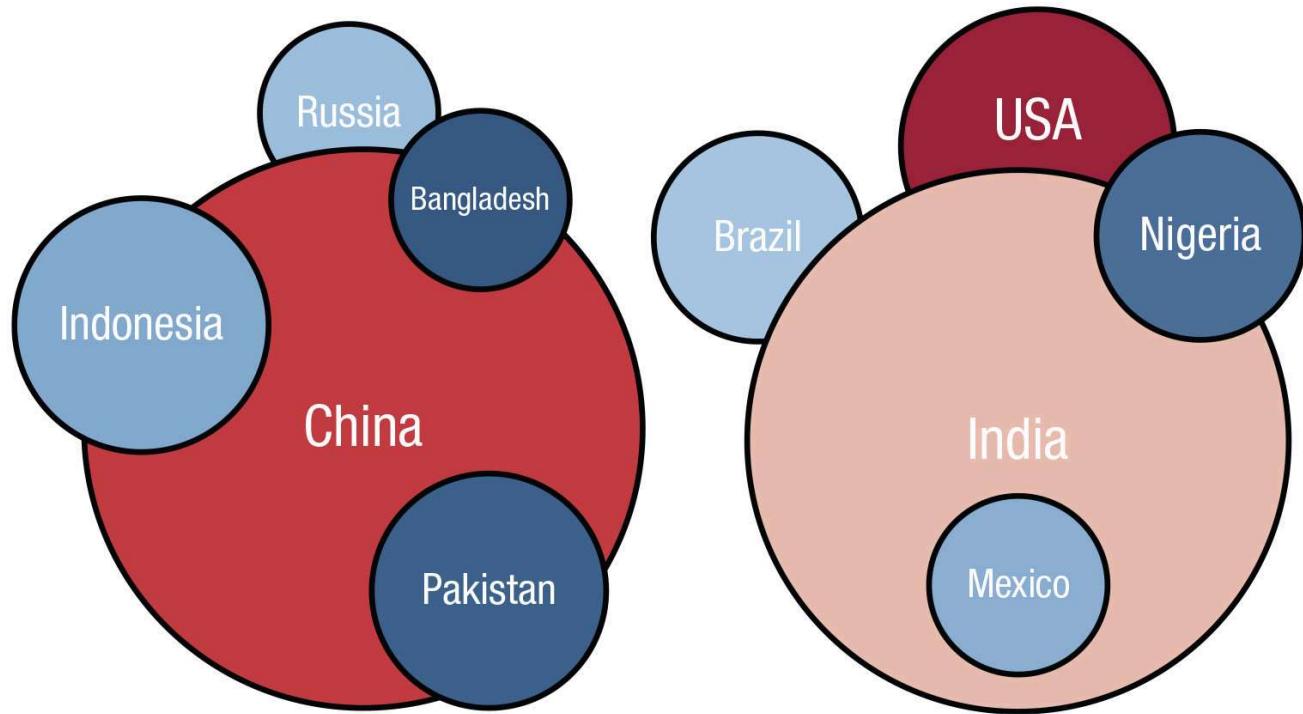
pictorial elements do improve memory for the fact that a visualization was previously seen, both in the short and the longer term ([Borkin et al., 2013](#)).

## How to Design an Understandable Visualization

Use familiar designs to show data intuitively

Visualizations can be powerful, but a poorly designed visualization can easily confuse or even mislead ([Burns et al., 2020](#); [Cairo, 2019](#); [Szafir, 2018](#)). Because the interpretation of visualized data is in the eye and mind of the human beholder, we must consider the psychology of the observer as the translator of images into an understanding of the original data and the patterns that they hold. Below, we outline a set of common translation errors that can confuse and mislead.

Understanding a visualization can depend on a *graph schema*: a knowledge structure that includes default expectations, rules, and associations that a viewer uses to extract conceptual information from a data visualization. [Figure 16](#) serves as an example of why a graph schema is often needed to interpret a data visualization. It depicts the GDP (on a log scale) and population of the 10 most populous countries. Take a moment to interpret the data.



**Fig. 16.** An example of a visualization with an unclear schema.

If you are having trouble extracting the data from this visualization, it is not your fault—you do not have the needed schema. First, if you have never seen this type of visualization, you cannot know which aspects of its variation are meaningful. The bubbles differ in their areas, hues, horizontal positions, vertical positions, proximity and enclosure relationships, and depth planes, which causes confusion. Why are Russia, Brazil, and the USA “hiding” behind the other countries? Why does India enclose Mexico? Why are there two main clusters? Is it meaningful that Bangladesh and Pakistan have the same horizontal position? Which of these many sources of variability should the viewer translate back to the original data?

You are experiencing the same confusion felt by a novice viewer encountering a given type of visualization for the first time. A young student might see rectangles with widths, heights, tops, bottoms, left sides, right sides, legend entries, and gridlines that vary in vertical position. Reading a bar graph requires a schema that allows the viewer to do what you do when viewing a bar graph: to focus instantly on the positions of the bar tops, because they signal the numbers depicted by the bars.

Here is a hint for [Figure 16](#): Only the size and hues matter; the rest of the variation is irrelevant. Areas map to population, and hues to GDP. Now, how does the variability in those channels map back to numeric values? You will guess that larger sizes are linearly related to a larger population because you have implicitly memorized that association from previous experience with visualizations. However, how do the hues map to GDP? It looks like the variance scales between red and blue, but is red high because it is “hot” or low because it is “bad”? You may have already inferred that red maps to higher by stepping back to your knowledge of the data to realize that China has a high GDP, so red probably maps to high. However, if you did not have that knowledge, your existing schema would not have allowed you to recover that relationship confidently.

Humans develop similar schemas for many other domains. A face schema, for example, specifies what features are typically in a face (eyes, ears, mouth, nose . . .), how those features tend to vary (many eyes have brown irises, but some are blue, green, or gray), what their relations tend to be (noses are between, and below, the eyes; [Palmer, 1975](#)). Having well-developed schemas enables a viewer to more effectively perceive, understand, and remember—without a face schema, a face presents a set of tens of thousands of independent pixels that vary in unpredictable ways. When one views a graph, one may depend on graph schemas that encode knowledge about the graph type, and also schemas that encode knowledge about what the graph is about—for example, understanding a high-low graph of stock prices requires knowledge about that graph type and also about stocks and the stock market.

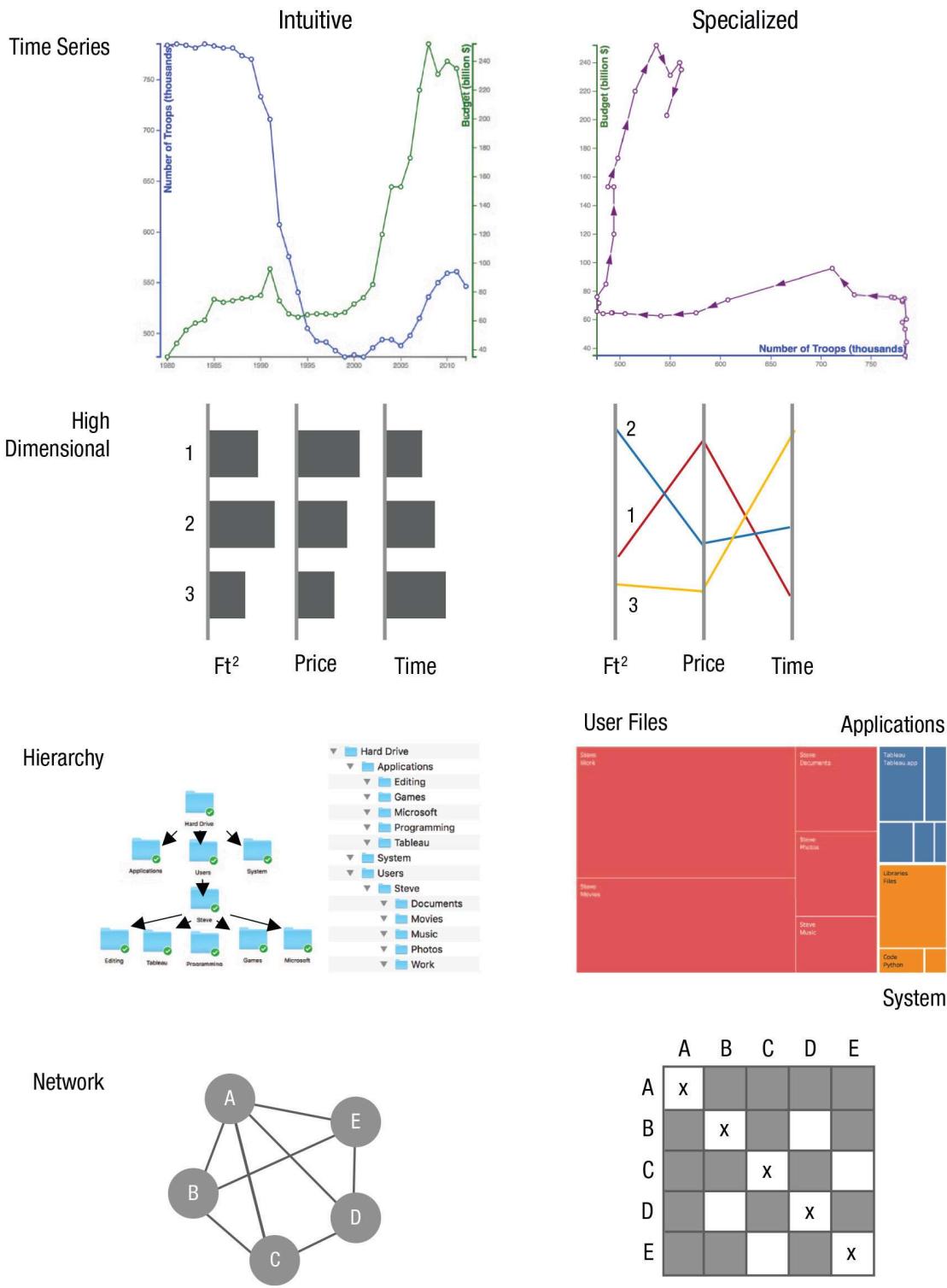
Schemas can bias judgments when a new instance deviates from the schema ([Mandler & Ritchey, 1977](#); [Rumelhart, 1980](#); [Tversky, 2001](#)). In one example of bias in the context of graphs, participants saw simple displays labeled as either “graphs” or “maps” and were asked to draw the displays from memory. When the displays were called “graphs,” participants distorted a central line as being closer to a 45° angle. When these same displays were called “maps,” the participants distorted the lines as being closer to horizontal or vertical, suggesting that their memory for the

image was influenced by canonical representations of these two types of visualizations ([Tversky & Schiano, 1989](#)).

In much of the world, children learn about common graph types such as bar graphs, line graphs, and scatterplots in school—but an audience without such schooling may not appreciate the conventions for how Cartesian space is used and how axes are drawn and labeled. Even relatively common visualizations such as scatterplots are misunderstood by 37% of adults in the United States ([Goo, 2015](#)).

Given that the reader of this article is likely familiar with common visualizations, we remind the reader of the need to learn how a new visualization works, by introducing examples that are less likely to be familiar. Expert communities have developed idiosyncratic visualizations that serve important data-analysis or -communication purposes, but they require learning new graph schemas. This can lead to severe difficulties for new viewers. For a broader tour of a curated “zoo” of novel visualizations, see [Heer et al. \(2010\)](#). For those ready for a true safari of new visualization schemas, we recommend browsing the Xenographics website ([Lambrechts, n.d.](#)).

The left column of [Figure 17](#) depicts four visualizations that are likely familiar to the reader, and the right column shows four equivalent visualizations that likely require a new schema. The first row shows two line graphs, plotted in the same space. An alternative way to plot those data is with a *connected scatterplot* shown at right ([Haroz et al., 2015a](#); [Peebles & Cheng, 2003](#)). For each time point on the x-axis of the two-line graph, plot the point for Line 1 on the x-axis of the scatterplot and Line 2 on the y-axis. Now connect all points with a line, in a temporal order. Engineers and economists use this format because it facilitates views of relationships between two sets of data: The line segments’ slope and direction reflect their ratios and changes in polarity.



**Fig. 17.** Four common types of visualizations and equivalent representations used by expert communities.

The second row depicts standard bar graphs on the left, showing the square footage, price, and time on the market for three houses. It becomes tougher to find

relationships among similar houses in the bar graph as the number of metrics increase or the data set's size grows because both the measures and the houses become more spatially separated. Parallel coordinates, as shown on the right, are a popular visualization format for analysts facing a much larger data set. The same data are now plotted as the height of lines that straddle parallel lines representing each measure. When combined with interactivity, this format allows users to extract individual values and perceive correlations between neighboring measures: flat lines reflect positive relationships, and slanted lines, negative relationships.

The third row shows a simple hierarchy of the content of a hard drive, in the style of an organizational diagram. In this representation, vertical position relies on a “levels of height” metaphor to indicate a folder’s place in the hierarchy, whereas horizontal position is typically irrelevant. The alternative depiction at right is called a tree map: *tree* because it shows a hierarchy, and *map* because of its two-dimensional spatial layout ([Shneiderman, 1992](#)). Reading a tree map requires the schema of knowing that folders are now differently sized rectangles (with area indicating the size of each file and color somewhat arbitrarily mapping to different branches of the hierarchy). Critically, the hierarchy itself is no longer shown through a vertical “levels” metaphor but is instead depicted through an “enclosure” metaphor, as in a physical file drawer. Higher-level folders are now akin to bento boxes enclosing other bento boxes.

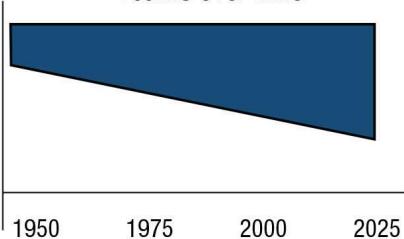
The final row shows network data as a *node-link diagram*, which could depict a social network. A node-link diagram can intuitively show connections (e.g., between people) with lines, making it useful for lay audiences. However, for complex analysis of large data sets, specialists prefer variants of the depiction at the right, in which each “node” is both a row and a column in the matrix, and their intersecting square is filled only if those nodes are connected ([Heer et al., 2010](#)). Similar in spirit to the correlation matrix, it is symmetric across the diagonal if all connections are bidirectional. However, if the connections are directional, then the two halves carry specific information.

Respect assumptions about how visual channels map to “less” and “more”

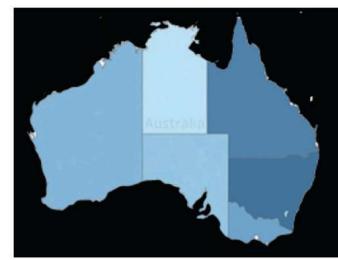
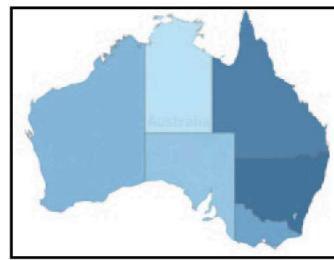
Some elements of schema are relatively universal for moderately experienced visualization users. For example, as illustrated in [Figure 18](#), people have expectations for which “end” of some visual channels (e.g., bottom vs. top for position, light vs. dark for luminance) should map to smaller and larger numerical values. Larger values tend to be mapped to higher vertical positions, perhaps because of the natural metaphor of stacking more objects in higher piles, or larger people’s being taller. Likewise, many communities—particularly those whose languages are read from left to right—use a consistent mapping of larger values as being on the right side of horizontal space, as in the number lines found on the walls of elementary school classrooms ([Tversky, 2000, 2001](#)). Note that you implicitly knew that the phrase “first column” at the start of this paragraph indicated the column on the far left.

#### Common Confusions Caused by How Data Are Mapped to Visuals

Deaths over time

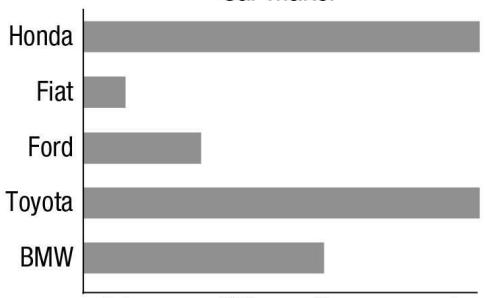


It can be confusing to map increases downward

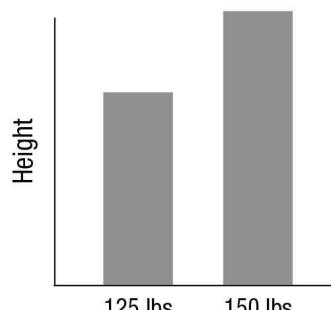


For light backgrounds, darker colors clearly map to higher values. For dark backgrounds, it's not so clear.

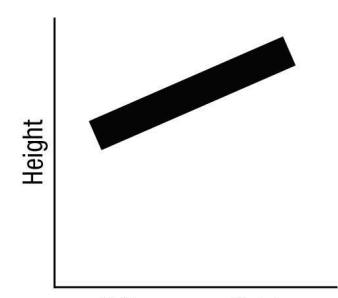
Car maker



It can be confusing to map nominal values to magnitudes



The choice of graph can substantially influence conclusions made from the same data



**Fig. 18.** Four types of counterintuitive graph designs. The graph at the top left shows a confusing flip of the typical mapping: positive numbers are plotted below the y-axis and

increase as they move downward ([Pandey et al., 2015](#)). At right, larger values naturally map to dark values when placed on a light background (recommended), but the mapping is less clear when the colors are placed on a dark background ([Schloss et al., 2018](#)). As shown on the bottom left, a bar graph is a counterintuitive way to plot a nominal variable: Mapping the country of origin of a car company to a bar graph makes the data “feel” metric ([Mackinlay, 1986](#)). At right, bar graphs encourage discrete comparisons between points (“150-pounders are taller”), whereas line graphs encourage descriptions of trends (“One tends to get taller as one becomes more Dutch”); in either case, viewers may extract the wrong conceptual message ([Zacks & Tversky, 1999](#)).

When these mappings are violated, viewers can become confused ([Gattis & Holyoak, 1996](#)). The brain electrophysiology research community shifted from a convention of plotting negative values (for microvolts) up on the  $y$ -axis to a more intuitive convention of plotting positive values up, forcing experienced researchers to relearn how to read familiar patterns that were now inverted ([Handy, 2005](#)). Visualizations from popular news media have been criticized for violating the convention of using higher position to indicate larger numbers, leading to confusion among viewers ([Gattis & Holyoak, 1996](#)). One such visualization depicted an increase in gun deaths in Florida after a permissive gun law was passed ([Engel, 2014](#); see also [Kosara, 2014](#)). Each used a counterintuitively inverted  $y$ -axis, smaller values at the top and larger values at the bottom, so that the designer could evoke a metaphor of downward streams of blood. Critics reported being initially confused by the break in convention, seeing increases in deaths as *decreases*. One study found that when graphs used this unconventional mapping ([Fig. 18](#)), viewers incorrectly assumed that positive values plotted below the  $y$ -axis midpoint were negative values, failing to notice the critical change in the  $y$ -axis labels ([Pandey et al., 2015](#)).

For length and area, the mapping is straightforward: Larger is more. Because these channels cannot intrinsically represent negative numbers (sizes cannot be negative), other channels must pitch in to differentiate negative values: Lengths are placed

with a position under an axis, and negative areas could be red in color. For angle, larger angles correspond to more, perhaps relying on the metaphor of climbing a hill: A steeper slope leads one higher, a flat slope does not change elevation, and a downward slope leads one lower. Of course, all of that assumes that you are “walking” from left to right, relying on a standard mapping of time and narrative as progressing from left to right ([Boroditsky, 2011](#); [Tversky et al., 1991](#)).

For intensity, some previous work had shown that people tend to map larger values to darker colors, whereas other work had suggested the same for colors that appear more opaque. One study appears to have solved this long-standing uncertainty by showing that both factors affect the mapping depending on the background color. In [Figure 18](#), in the map at the upper right with a white background, darker colors also appear more opaque, so darkness and opacity predict the same mapping: Either appears to show the larger number. However, in the map with a dark background, the darker areas can be seen as “see-through” so that the lighter areas seem comparatively opaque. Now, associations with darker as larger compete with associations of more opaque as larger. The clear mapping on the left is preferred, and the unclear mapping on the right must be rescued with adjustments to color combinations ([Schloss et al., 2018](#)).

One could imagine many alternative rules that could govern the mappings from numbers to visual properties. A computer vision system could be programmed to map larger numbers to higher positions, as humans do, or just as easily programmed to map them to lower positions, leftward positions, shorter lengths, or lighter values. Continua need not even be mapped consistently over continuous space: one could map the number 1 to position 1, 2 to 9, 3 to 5, and so on; as long as that translation is available to the computer vision system, it could easily recover the original values. The fixed architecture of the human visual system could not efficiently process those mappings.

### Respect associations between visualization designs and data types

We have so far focused only on *metric* visualization—the visualization of continuous magnitudes that can be positive or negative. There are other types of data ([Stevens,](#)

[1946](#)) that are frequently shown in visualizations ([Bertin, 1983](#); [Mackinlay, 1986](#); [Munzner, 2014](#)). One is nominal data, categorical values that can be counted or can be used to divide a set of metric numbers. Data sets often include both metric and nominal data. For example, a company might have a spreadsheet recording sales with columns of metric values (number of orders, sales in dollars, shipping time in hours) and columns of nominal values (customer name, product ID, country code). One might sum over all of the metric sales numbers or split those numbers by product ID to compute averages for each ID. Another type of data falls in between metric and nominal: *ordinal* values, which represent ranks that can be compared and ordered but do not reflect continuous metric differences. Examples include stars or dollar signs for restaurant ratings or price levels; designations of coach, business, and first class on an airplane; or rankings of colleges in a magazine. These values may have been generated with metric data (e.g., the cost of airplane tickets, numeric ratings of colleges), but they do not contain metric information per se.

Viewers expect particular types of data to be mapped to particular visual properties. Metric data are typically mapped to position, length, area, angle, or intensity. Nominal data can be mapped to position (e.g., a bar graph showing sales for each product ID might spread those IDs from left to right or from top to bottom) or to color or shape (e.g., to identify each point's country code on a scatterplot pitting sales against shipping time). But if nominal data are plotted with lengths or areas, viewers can become confused. An example of this type of confusing mapping is depicted in [Figure 18](#) (adapted from [Mackinlay, 1986](#)). The varied lengths of the bars make their values “feel” metric, such that the viewer wonders why Japan is “larger” than Italy.

[Figure 18](#) also shows that the visual system responds to points or bars as separate objects, whereas a line connecting points “feel like” a single object that has been stretched or has moved over time and left a path ([Tversky, 2001](#)). This has led to a convention of using bars to depict values from different nominal categories and using lines to depict values from different places on a metric continuum. This mapping of nominal data to bars and metric data to lines is commonly described in textbooks and guides on graphing—however, there is more to the story. What

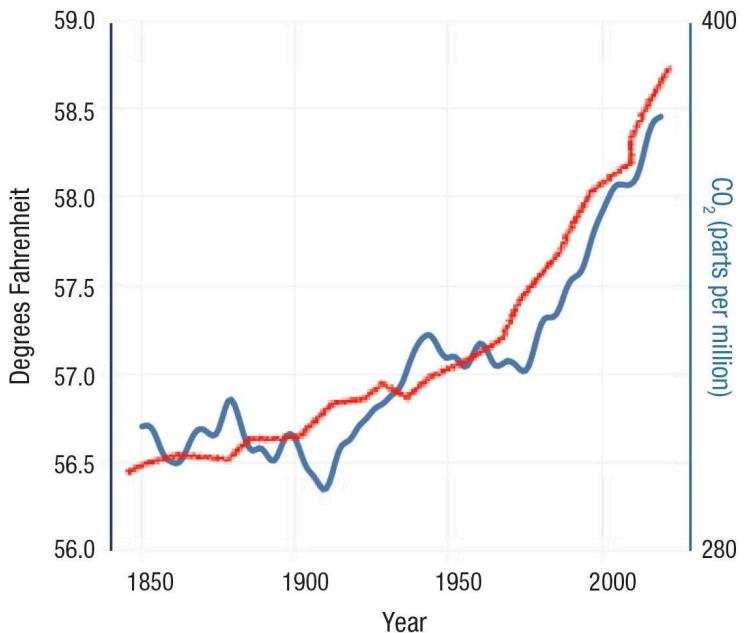
matters more than the class of the data is the intended message to be conveyed by the visualization. Suppose one wants viewers to compare two values along a continuous dimension—for example, mood at 5 p.m. versus 9 a.m. Using bars to represent mood at those two time points would encourage interpretations such as “Mood is more positive at 5 p.m. than 9 a.m.”; this could be the right choice despite the fact that time is a metric variable. Using the wrong visualization for a given conceptual message can lead to odd conclusions: In the height examples at the bottom right [Figure 18](#), the graphs violate both the data-type convention and the intended conceptual convention. Showing values as a function of weight using a bar graph tends to lead people to conclude that “150-pounders are taller than 125-pounders,” which is reasonable but perhaps not as helpful an insight as “Height increases with increasing weight.” Likewise, showing values as a function of nationality using a line graph could lead a substantial number of viewers to conclude that “One tends to become taller as one becomes more Dutch” ([Zacks & Tversky, 1999](#)).

Connect relationships in visualizations to those in the world

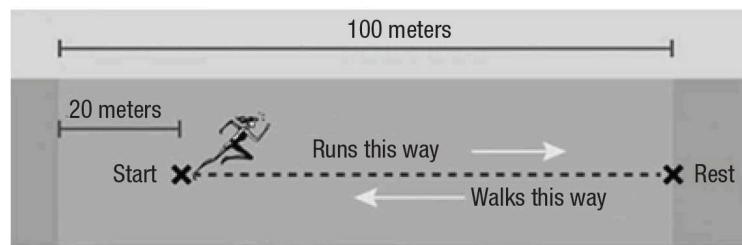
Even if a viewer can extract values and data relationships from a visualization, they can fail to connect those to important real-world concepts in ways that are critical for reasoning. For example, consider the graph and corresponding text at the top of [Figure 19](#), given to 673 high school students by [Farrar \(2012\)](#), as cited in [Whitacre & Saul, 2016](#)). Only 4% of the students noted that the text did not correspond to the graph’s information. In another study, high school students in honors chemistry and environmental science received similar materials on the topic of teen pregnancy, and only 25% noted a discrepancy between the text and graph ([Whitacre & Saul, 2016](#)). In a more complex task ([Fig. 19](#), bottom), sixth through eighth graders were asked to create a position–time graph on the basis of a description of a runner’s training routine. Only 1% of the students did so correctly, and most drew “graphs” representing a map of the runner’s position.

## Changing Global Climate

A friend has asked you to peer edit a presentation. Are all of the facts being presented correctly?



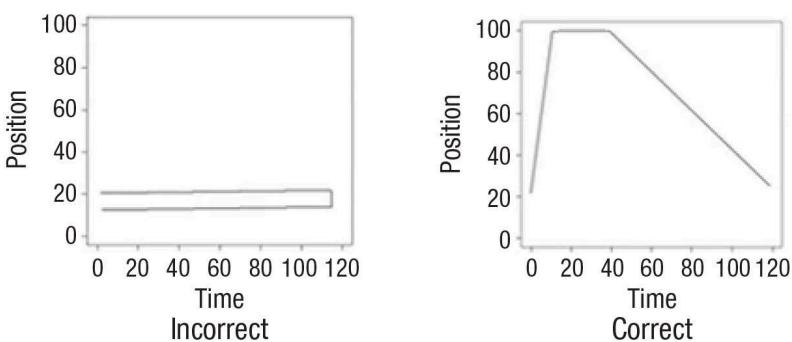
Carbon dioxide (CO<sub>2</sub>) emissions data from ice cores was compared with temperature records, confirming that global temperature and carbon dioxide levels are not related.



Gizelle trains on a field that is 100 m long.

She sprints from a point 20 m distant from one end of the field to the other end of the field, takes a short rest, and walks back slowly to the start.

Draw a graph to represent this routine.



**Fig. 19.** Graphical reasoning problems. When shown the graph at the top, only 4% of high school students noticed inconsistencies with the accompanying text (Farrar, 2012). The bottom shows an example item from Lai et al. (2016); asked to plot a runner's routine on a time–distance graph, most middle-school students drew “maps,” and only 1% drew a correct graph. Bottom figure adapted by permission from Springer: *Journal of Science*

*Education and Technology*, “Measuring Graph Comprehension, Critique, and Construction in Science,” by K. Lai, J. Cabrera, J. M. Vitale, J. Madhok, R. Tinker, and M. C. Linn (<https://doi.org/10.1007/s10956-016-9621-9>). Copyright 2016.

Although students can easily extract individual values and relationships from the climate and running visualizations shown in [Figure 19](#), they have difficulty linking the depicted patterns with the accompanying text because graphing is often taught with a focus on plotting data points and representing quantitative functions. Instead, solving the example problems in [Figure 19](#) requires that they see the underlying link between the visual and verbal representations of the patterns. This problem arises when visualizations must be linked to text, tables, or mathematical equations. One intriguing proposal for helping students see past extracting values, to focus on making external links, is to omit numbers on axes entirely and provide only “qualitative” graphs that distill critical categorical relations. In one study, activities in which students constructed or critiqued qualitative graphs distinctly contributed to their understanding of cancer cell division and treatment impact ([Matuk et al., 2019](#)).

In the classroom, another effective method for helping students see these links is to give them abundant and explicit practice in translating between such representations. One study found that prealgebra students generated better solutions for word problems involving simple linear functions when they first generated graphical, tabular, and equation representations of those problems, compared with a control group who generated and solved equations for similar word problems ([Brenner et al., 1997](#)). The intervention group also gained a substantially better conceptual understanding of functions compared with the control group. Graphing calculators can also help students quickly see how changing features of equations affects visualized functions ([Doerr & Zangor, 2000](#); [Hollar & Norwood, 1999](#); [Mesa, 2008](#)); a meta-analysis of 42 studies of graphing-calculator use found benefits from middle-school mathematics through first-semester college calculus ([Ellington, 2006](#)). Another useful technique was to switch

rapidly (within seconds) among multiple representations of similar concepts (tables, graphs, equations, and verbal rules), which helped students develop an understanding of functions ([Kalchman & Koedinger, 2005](#)).

Another way to facilitate these links is by starting with familiar contexts, such as earning money for each mile walked in a walkathon, allowing students to see abstractions of those functions from concrete meaningful situations ([Kalchman & Koedinger, 2005](#)). Likewise, instructors can provide students with scenarios in which they collect and summarize data and guide them toward inventing visualizations, so that the mappings between the statistical concepts and visualizations are clear ([Lehrer & English, 2018](#); [Lehrer & Romberg, 1996](#); [Lehrer et al., 2000](#)).

## Visually Communicating Uncertainty and Risk

Why communicating uncertainty is critical

Quantifying uncertainty is critical for analysts and scientists who want to make informed decisions from data. Uncertainty is inherent in the generation of data and models, and it comes from multiple sources. Individual measurements contain noise. Measurements of a population of values typically rely on a sample of those values that is incomplete, or even biased. To estimate average human height, for example, we must rely on measuring a subset of all people. Averaging 100 people's heights produces an estimate, but an uncertain one, because we did not definitively measure the height of every person in the world. If we assume that the 100 people are randomly sampled from the population, then our uncertainty can be calculated using statistical theory. However, if we were to sample 100 people who all happen to be NBA players, our sample would not be representative of the world, and standard statistical approaches for random samples would not adequately capture the true uncertainty.

Effective communication of uncertainty is necessary for scientific literacy among the public; without understanding uncertainty, people cannot accurately calibrate their internal sense of how reliable a pattern or claim is. Yet communicating uncertainty

is challenging as a result of many well-documented biases related to decision making, many of which are characterized by misuse of statistical evidence in reasoning (e.g., [Tversky & Kahneman, 1974](#)). Unfortunately, the level of uncertainty in analysts' calculations is often judged to be too complex or time-consuming to communicate to lay readers or busy decision makers ([Fischhoff & Davis, 2014](#)).

Communicating uncertainty is also important for maintaining public trust. In weather forecasts, when uncertainty is presented counterintuitively (or not presented at all), people tend to discount it. Unfortunately, when the forecast turns out to be "wrong," this can lead to a lack of trust in scientific forecasting ([Binder et al., 2016](#); [Joslyn & LeClerc, 2012](#)). Among scientists, underemphasis or misunderstandings of sampling error (a form of uncertainty) and the likelihood of replicating experimental results may contribute to continued use of underpowered studies and the "replication crisis" across many previously accepted findings ([Ioannidis, 2005](#)).

Some common cognitive biases are related to misunderstandings of uncertainty. One reason this may be so is that many core statistical concepts, such as variability, estimation, and sampling, are complex and can be defined clearly only with reference to statistical theory. A key challenge in communicating uncertainty is that lay audiences may not appreciate that estimates are subject to variability and may not be familiar with the statistical abstractions commonly used to express these concepts ([Gal, 2002](#)).

This lack of understanding of statistical constructs is present even among researchers. In a *frequentist* statistical paradigm, it is not possible to make statements about the probability that a specific interval contains the true population parameter value (e.g., the true average height of U.S. males) because probability can be defined only as frequency over a long run of trials. Instead, the frequentist 95% confidence interval refers to the probability that a confidence interval constructed in the same way as the plotted interval contains the population parameter. However, even users with statistical training, such as graduate students, are prone to mistakenly conclude that it is 95% probable that a 95% confidence interval contains the true parameter ([Hoekstra et al., 2012](#)). Likewise, the

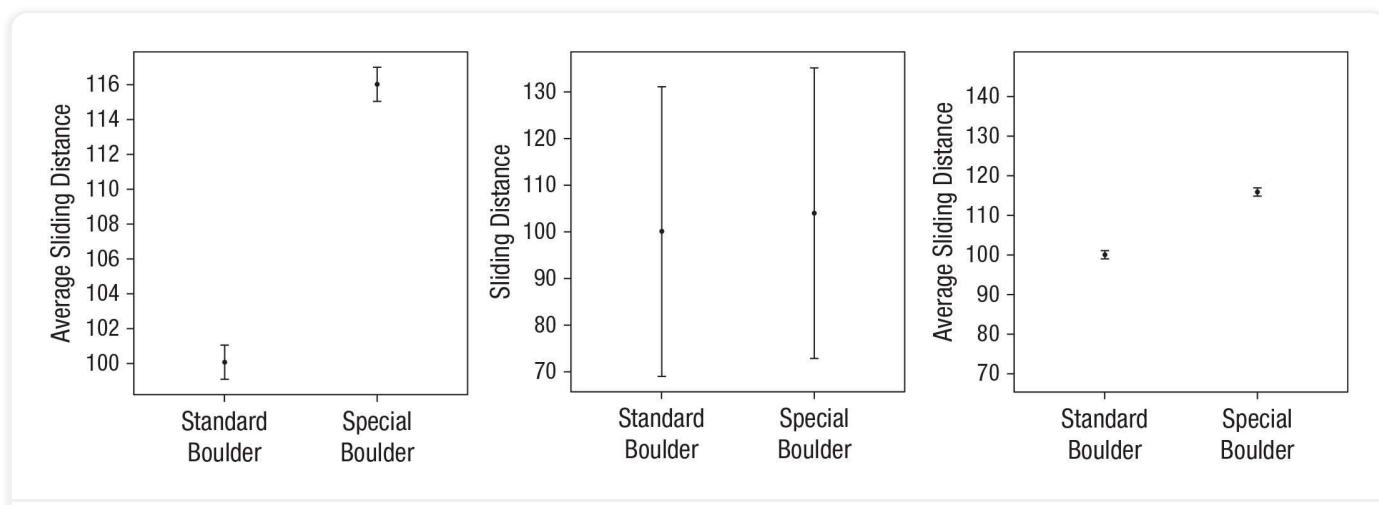
relationship between statistical significance and whether or not two error bars overlap is often misunderstood: When two frequentist 95%-confidence-interval error bars do not overlap, it is correct to assume that there is a significant difference between the two quantities at an alpha level of 0.05. However, when the two intervals do overlap, researchers incorrectly assume that there is no significant difference between the two quantities ([Belia et al., 2005](#)).

Errors such as these likely stem from challenges that students of statistics and others face concerning how to interpret the sampling distribution of the mean implied by a 95% confidence interval or standard-error interval ([Chance et al., 2004](#); [Hullman et al., 2017](#)). The sampling distribution is the distribution of means expected if one were to repeatedly draw samples of a given size  $n$  for a population. Although the sampling distribution is a more natural choice when trying to answer questions about the probability that a difference is not zero (i.e., statistical significance) and is systematically related to the variance in the underlying measurements (and in estimates of the parameter value in the population), this distribution does not directly address the sorts of questions one might ask about the effect of applying a treatment in the world.

Common visualizations of uncertainty are often misinterpreted

Uncertainty is challenging to communicate because the concept of uncertainty is difficult to understand in the first place. In graphs designed for a lay audience, depictions of uncertainty are often simply omitted, even when the graphs are intended to support inferences beyond the data sample that is shown ([Hullman, 2019](#)). When uncertainty *is* presented, it is most frequently shown through error bars. They typically represent either a standard-deviation range, a standard-error range, or a confidence interval (e.g., a frequentist 95% confidence interval or a Bayesian 95% credible interval). Which of these constructs is depicted has a large influence on a viewer's impression of effect size—referring to a set of quantitative measures of the magnitude of a difference, such as the difference in means between two distributions, divided by their pooled standard deviation ([Hofman et al., 2020](#)). For example, when viewing results of an evaluation of a new intervention

(e.g., a special type of equipment one might use to improve one's score in a sports competition) relative to a control, one might wonder how much investing in the new equipment is likely to help increase one's score for a randomly drawn instance of the game. When shown frequentist 95% confidence intervals representing uncertainty on bars representing the control ("Standard Boulder") and treatment ("Special Boulder") outcomes in a fictional boulder-sliding competition ([Hofman et al., 2020](#); see [Fig. 20](#), left), lay viewers were inclined to pay more for the treatment and overestimate effect size relative to when they saw error bars showing a predictive interval representing uncertainty in the underlying measurements ([Fig. 20](#), middle). Although not as effective as showing predictive intervals, rescaling the y-axis of a chart showing 95%-confidenceinterval error bars to display the range required for the predictive interval slightly reduced overpayment and bias in effect-size estimations ([Fig. 20](#), right).



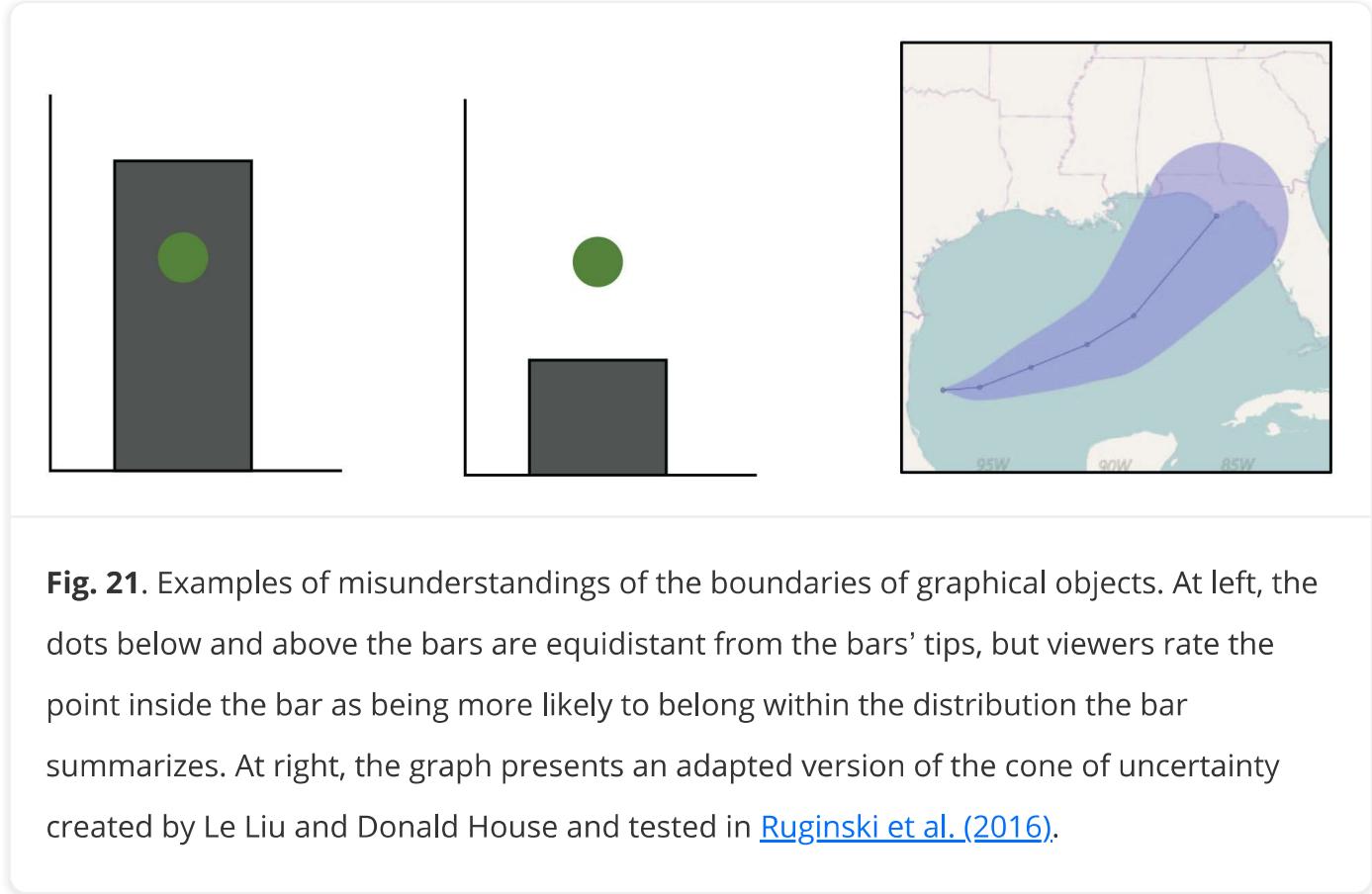
**Fig. 20.** Examples of misunderstandings of variance depictions. Hofman et al. (2020) found that participants in an online study who viewed the chart on the left, showing a range extending  $1.96 SE$  above and below the mean, were more likely to pay more for a new intervention (i.e., "Special Boulder") compared with the standard intervention (i.e., "Standard Boulder") and to overestimate the size of the new intervention's effect than those who saw the chart in the center, showing a range extending  $1.96 SD$  above and below the mean. The chart on the right is the same as the chart on the left, except that it has the y-axis range of the chart in the center. This display slightly reduced overpayment and overestimation, which suggests that viewers are sensitive to superficial visual cues,

such as how big a difference looks within the frame of the chart. Republished with permission of the Association for Computing Machinery, from “How Visualizing Inferential Uncertainty Can Mislead Readers About Treatment Effects in Scientific Results,” by J. M. Hofman, D. G. Goldstein, and J. Hullman, *CHI ’20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (<https://doi.org/10.1145/3313831.3376454>). Copyright 2020 Association for Computing Machinery.

Error bars can lead lay participants to use visual reasoning strategies that result in error-prone perceptions of effect size. For example, when lay participants were asked to evaluate the respective effect size between two distributions shown as bars representing means with error bars, their answers appeared to differ mostly on the basis of the difference between the means and were unaffected by the level of uncertainty around those means ([Hullman et al., 2015](#)). One investigation of the visual reasoning strategies employed by lay participants to make effect-size judgments and decisions from error bars and other common uncertainty visualizations found that many participants used approximations based on visual distances in the chart rather than reasoning more deliberately about the size and uncertainty of effects ([Kale et al., 2021](#)).

If lay viewers are unfamiliar with standard-deviation and confidence intervals, it is perhaps unsurprising that they tend to misunderstand the depictions of those parameters through error bars or confidence envelopes. Error bars are not seen as parameter estimates for a continuous distribution but are instead mistaken as depicting the discrete range of the data. When shown bars representing 95% confidence intervals of a forecasted nighttime low temperature, participants incorrectly believed that the upper and lower bounds represented forecasted high and low temperatures ([Joslyn & LeClerc, 2012](#)). This mistake occurred despite the presence of a highly visible key that depicted the correct interpretation. [Figure 21](#) illustrates how not just error bars but bar graphs themselves can lead to similar misunderstandings of uncertainty: Viewers appear to implicitly believe that a bar contains the full distribution of data values that it summarizes because of its

metaphorical status as a container. When reminded that the tip of a bar graph represents the mean of a distribution of values, viewers still rate points that fall slightly below the tip as more likely to belong to that distribution than points placed slightly above it ([Correll & Gleicher, 2014b](#); [Newman & Scholl, 2012](#)).

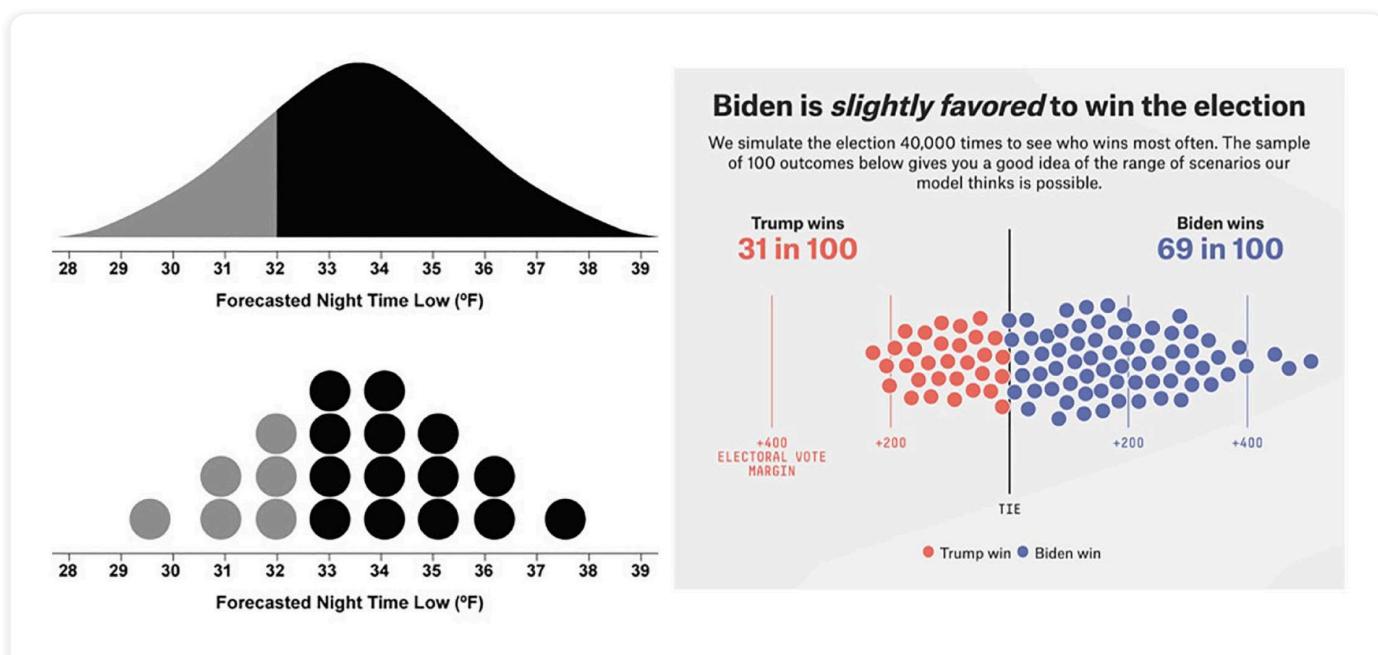


**Fig. 21.** Examples of misunderstandings of the boundaries of graphical objects. At left, the dots below and above the bars are equidistant from the bars' tips, but viewers rate the point inside the bar as being more likely to belong within the distribution the bar summarizes. At right, the graph presents an adapted version of the cone of uncertainty created by Le Liu and Donald House and tested in [Ruginski et al. \(2016\)](#).

Another notable example of this “discrete range” error occurs with hurricane-forecast visualizations. The most popular method for displaying hurricane paths, produced by the National Hurricane Center (see [Fig. 21](#)), uses a “cone of uncertainty” to represent 66% confidence intervals across thousands of storm-track simulations. Note that this means that many predicted storms would travel outside of this cone. Unfortunately, people incorrectly believe that the cone depicts a discrete range, such that areas in the cone are in danger and areas outside are relatively safe (e.g., [Padilla et al., 2017, 2020](#)). Even worse, many viewers do not recognize the cone as reflecting the storm’s potential paths, instead assuming that it depicts growth in the storm’s size over time.

## How to visualize uncertainty more intuitively

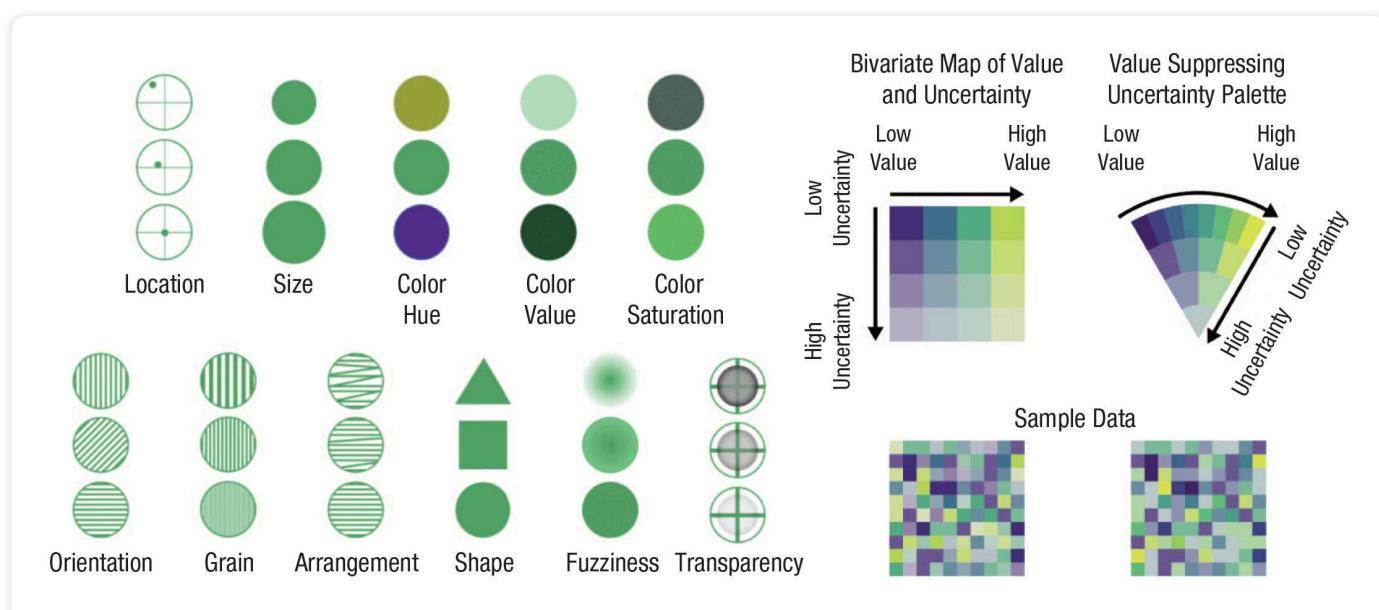
Instead of depicting summary statistics (e.g., confidence intervals) that are rarely well understood by lay audiences, a better approach is to show the underlying probability distribution. Density plots (Fig. 22, top left) and violin plots (similar, but typically mirrored to show a single symmetric shape) show distributions by mapping the probability of a given value to width. These plots can help viewers intuitively understand the predicted variability of values, location (e.g., mean), and distribution shape (e.g., normal, skewed). They can also promote intuitive assessments of how likely or surprising different values are, in ways that are more closely aligned with normative statistical definitions ([Correll & Gleicher, 2014b](#)). Several studies have found that density plots lead to better-quality decisions than error bars showing predictive or confidence intervals ([Fernandes et al., 2018](#); [Hofman et al., 2020](#); [Kale et al., 2021](#)).



**Fig. 22.** Depictions of distributions. At left, a probability density plot is represented as a quantile dot plot that uses a frequency (discrete-outcome) framing. Here, black dots or shaded regions indicate values above freezing temperatures. At right, a jittered variation with a categorical variable added via color shows uncertainty surrounding the 2020 U.S. presidential election outcome. Graph at right adapted by Anna Wiederkehr of

FiveThirtyEight (for more on FiveThirtyEight's creation of similar graphics, see [Wiederkehr, 2020](#)). Used with permission from FiveThirtyEight.

In other work, some researchers have argued that visual features such as fuzziness, color value, and unorderly line arrangement convey probability more naturally than more abstract mappings, such as mappings between uncertainty and size ([MacEachren et al., 2012](#); see [Fig. 23](#), left). However, representations that were judged to be more associated with uncertainty did not necessarily lead to faster or more accurate judgments. These findings illustrate a trade-off whereby some encoding types provide a subjective impression of uncertainty but may also reduce precision, making some tasks hard to complete. Experts might therefore prefer more precise representations because they have no trouble understanding how these visualizations convey uncertainty, whereas lay viewers likely benefit more from less precise visualizations that are more semantically resonant (or more naturally interpreted as uncertainty).



**Fig. 23.** Methods for depicting uncertainty. On the left of the figure are different visual variables for conveying uncertainty tested by [MacEachren et al. \(2012\)](#), who found that fuzziness, location, arrangement, and color value were rated as relatively more logical than other variables for representing uncertainty. The right shows a bivariate color map that

separately encodes value (via hue) and uncertainty (via lightness). A value-suppressing uncertainty palette uses a tree metaphor to collapse value distinctions in proportion to the uncertainty surrounding the values ([Correll et al., 2018](#)). Left figure reprinted with permission from “Visual Semiotics & Uncertainty Visualization: An Empirical Study,” by [A. M. MacEachren, R. E. Roth, J. O’Brien, B. Li, D. Swingley, and M. Gahegan, 2012](#), in *IEEE Transactions on Visualization and Computer Graphics*, 18(12), p. 2497 (<https://doi.org/10.1109/TVCG.2012.279>). Copyright 2012 by IEEE. Right figure republished with permission of the Association for Computing Machinery, “Value-Suppressing Uncertainty Palettes,” by M. Correll, D. Moritz, and J. Heer, *CHI ’18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (<https://doi.org/10.1145/3173574.3174216>). Copyright 2018 Association for Computing Machinery.

Yet another approach for showing statistical uncertainty intuitively is to rely on visual depictions that are inherently *perceptually* uncertain, preventing the viewer from resolving a value precisely. If the location of a point on a map could be blurred proportionally to the uncertainty in the position, then the recovered position would intuitively contain at least that level of uncertainty. Value-suppressing uncertainty palettes attempt this for color, taking advantage of a natural relationship among hue, saturation, and lightness: Hues are easier to tell apart when they are more saturated and darker ([Correll et al., 2018](#); [Fig. 23](#), right). Value is mapped to hue and certainty to saturation and lightness (darker for more certain). As a result, value judgments are more difficult for more uncertain values—the most uncertain values all appear as the same shade of gray. When this approach is applied to maps, users weigh uncertainty more heavily than they do when using conventional bivariate maps separating value and uncertainty.

### Present uncertainty as examples over space or time

Mounting evidence suggests that individuals can more effectively understand probabilities (e.g., 10%) when they are communicated as frequencies (1 out of 10; e.g., [Hoffrage & Gigerenzer, 1998](#); [Peters et al., 2011](#); [Schapira et al., 2001](#)). For

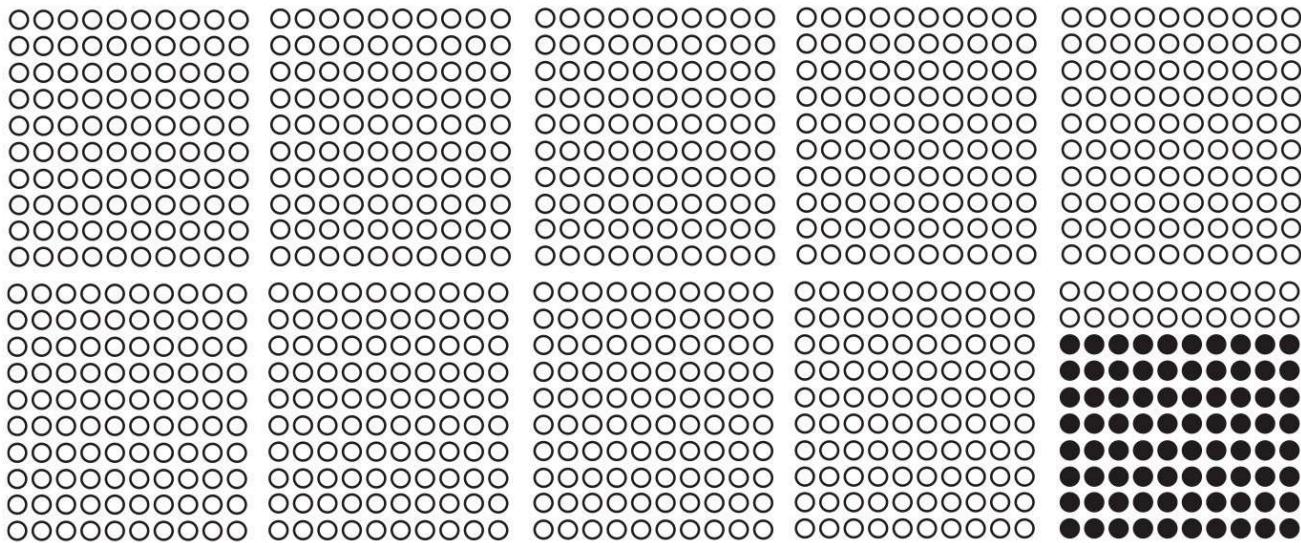
example, individuals were informed about a headache medication's side effects, framed as either a percentage (e.g., 10% of patients got a bad rash from the medication) or a frequency (e.g., 10 out of 100 patients got a bad rash from the medication), before they rated the riskiness of the treatment. Individuals with lower numeracy rated the treatment as less risky when the information was communicated with a percentage, as opposed to a frequency, framing. Individuals with high numeracy did not show a difference in their riskiness ratings as a function of framing. These results imply that the framing influenced individuals with low numeracy because the percentage was harder for them to understand than the frequency, whereas those with high numeracy understood both formats equally well ([Peters et al., 2011](#)). The frequency framing may be more effective because it more naturally maps to how we experience probability in our daily lives ([Hoffrage & Gigerenzer, 1998](#)).

### Convey probability with frequency-based visualizations

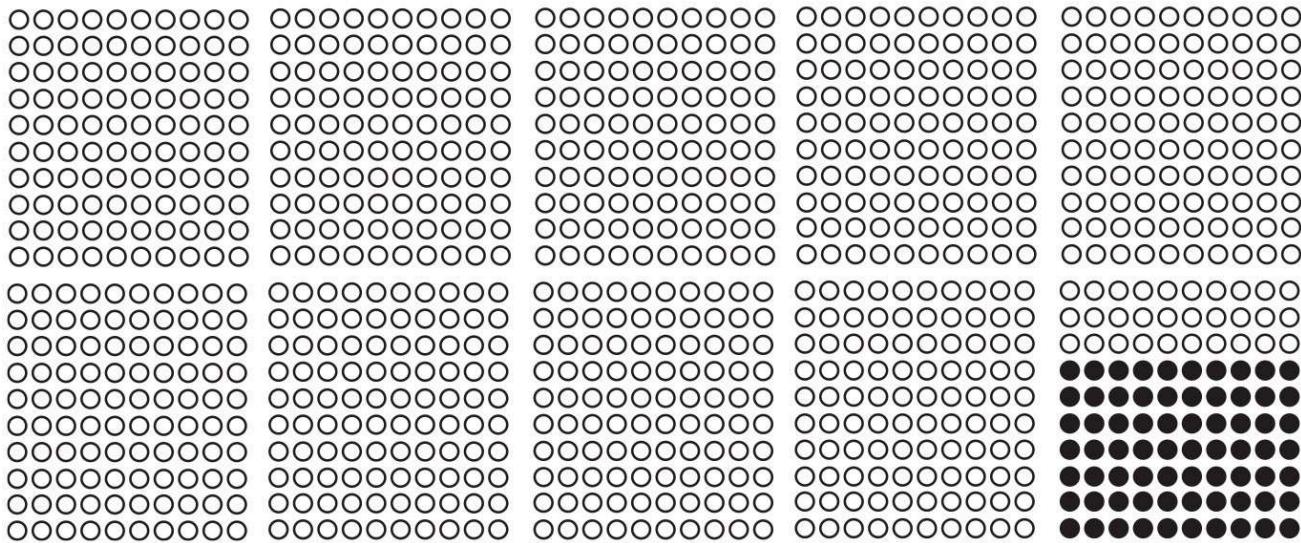
Frequency-based visualizations are those that express probability as frequency or ratio, as in a set of icons that show the numerator and denominator (icon arrays; see [Fig. 24](#)) or dots that represent discrete probabilities in a distribution (quantile dot plots; see [Fig. 22](#), left). The defining characteristic of frequency-framing visualizations is that they allow the viewer to infer probability via the frequency of visual attributes (e.g., position or size) of discrete visual elements such as icons, dots, or lines.

**For people with symptoms of arterial disease, aspirin can reduce the risk of having a stroke or heart attack by 13%.**

### **Without aspirin**



### **With aspirin**



**Fig. 24** Icon arrays from [Galesic et al. \(2009\)](#), which help viewers visually compare the relative rate of stroke or heart attack with and without aspirin. Copyright © 2009 by the American Psychological Association. Adapted with permission from “Using Icon Arrays to Communicate Medical Risks: Overcoming Low Numeracy,” by M. Galesic, R. Garcia-Retamero, and G. Gigerenzer, *Health Psychology*, 28(2), p. 211 (<https://doi.org/10.1037/a0014474>).

For example, a substantial body of research has demonstrated that icon arrays are one of the most effective ways to communicate probability in a health-care context because they use frequency framing ([Fagerlin et al., 2005](#); [Feldman-Stewart et al., 2007](#); [Garcia-Retamero & Galesic, 2009](#); [Garcia-Retamero et al., 2010](#); [Hawley et al., 2008](#); [Tait et al., 2010](#); [Waters et al., 2006, 2016](#)). In one study, students and older adults read descriptions of multiple medical scenarios framing the effects of treatments in terms of either absolute or relative health risks (e.g., “For people with symptoms of arterial disease, aspirin can reduce the risk of having a stroke or heart attack by 13%” vs. “8% of such people who did not take aspirin had a stroke or heart attack, compared with 7% of such people who did take aspirin”; [Galesic et al., 2009](#)). For a second group of participants, each text description was accompanied by two icon arrays that reinforced the numeric information (see [Fig. 24](#)). When asked to estimate outcomes with and without treatment (e.g., how many people out of 1,000 would have a stroke or heart attack if they did and did not take aspirin), participants who had received the additional icon arrays were significantly more accurate than those who had received only text-based probabilities. The benefits of icon arrays were observed for both students and older adults, and individuals with low numeracy drove the effect ([Galesic et al., 2009](#); see also [Garcia-Retamero & Galesic, 2009](#)).

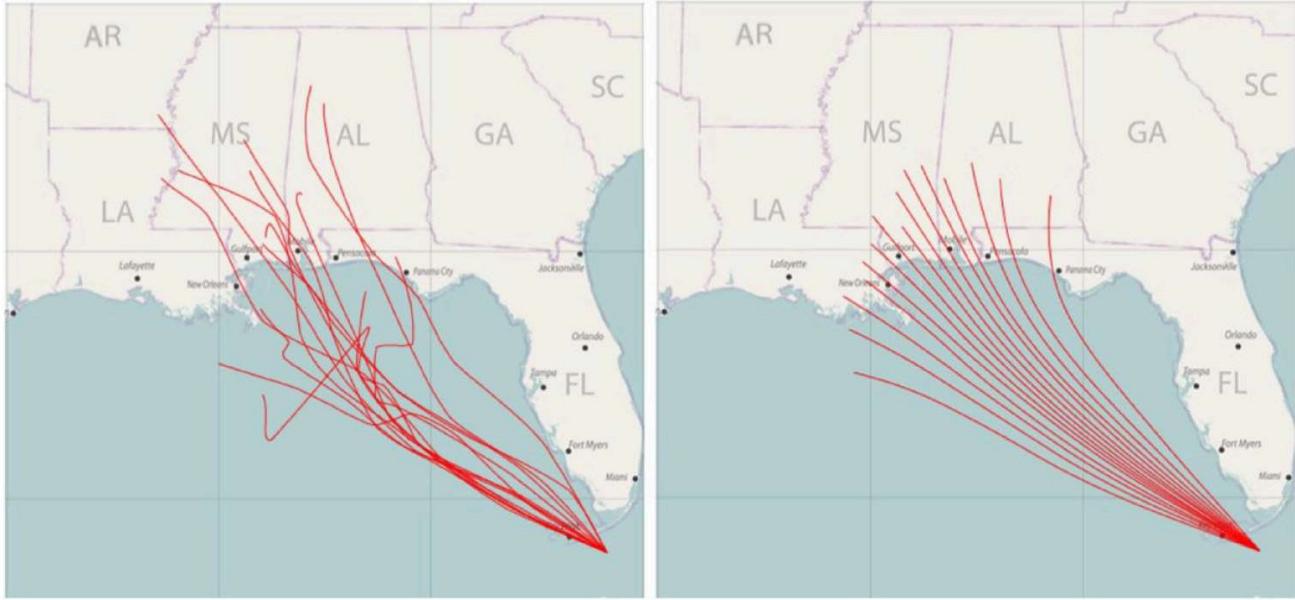
Researchers have conceptually replicated the advantages of communicating ratio data as icon arrays numerous times (e.g., [Garcia-Retamero et al., 2010](#); [Okan et al., 2012](#); [Stone et al., 2003](#); [Zikmund-Fisher et al., 2014](#); for a meta-review, see [Garcia-Retamero & Cokely, 2017](#)). Another study also found that patients trust icon arrays more than other common visualization techniques ([Hawley et al., 2008](#)), and other work has confirmed that icon arrays help patients with low numeracy correctly interpret probabilities (e.g., [Galesic et al., 2009](#); [Garcia-Retamero & Galesic, 2009](#); [Hawley et al., 2008](#)).

A frequency-framing approach is also recommended for conveying distributional information. In the hurricane cone-of-uncertainty example in [Figure 21](#), viewers apply the wrong schema when translating the visualization back to the data that it represents. Viewers misinterpret the edge of the cone, which depicts error bars

showing 66% confidence intervals, as ranges ([Ruginski et al., 2016](#)) or mismap the width of the cone to the storm's size instead of its potential path ([Padilla et al., 2017](#)). Given that policymakers and the general public rely on hurricane-path information to make decisions about preparation and potential evacuation, it is critical that they understand the depicted information immediately and intuitively.

Because it seems unrealistic to attempt to teach lay viewers both the requisite underlying statistical concepts and the schema for how they are mapped to a visualization, a more fruitful route is to redesign the visualization to use a frequency-based approach that pulls from existing metaphors known to lay viewers. One strategy is to simply show numerous examples of the samples that created the summary statistics (or simulated draws from the population estimate). This mimics the experience of collecting samples in the real world: flipping coins to gauge the percentage of tails, trying multiple menu items to gauge a restaurant, or testing how various kicking styles might put spin on a soccer ball. Providing discrete-outcome framings of data can provide a frequency metaphor that reduces probability-processing errors ([Hullman et al., 2015](#); [Kale et al., 2019, 2021](#); [Kay et al., 2016](#); [Padilla et al., 2017](#)).

To apply this alternative approach to hurricane forecasts, one solution to misunderstandings of the cone of uncertainty is to instead plot a sample of possible hurricane paths, allowing the viewer to intuit which areas are in most danger ([Fig. 25](#)). This method stops viewers from confusing uncertainty about the hurricane's path with the size of the storm and helps them understand which areas are most endangered ([L. Liu et al., 2018](#)). When showing uncertainty information using a frequency metaphor, one must ensure that the samples shown are representative of the uncertainty in the data. In hurricane forecasting, track visualizations are sometimes referred to as "spaghetti plots" because they can look messy, like a plate of noodles, when randomly sampled ([Fig. 25](#), left). To increase the usefulness of this technique, researchers have developed reconstruction procedures to ensure that the tracks clearly show the full uncertainty in a storm's trajectory (see [Fig. 25](#), right; [L. Liu et al., 2018](#)).



**Fig. 25.** Techniques for depicting uncertainty in hurricane paths. At left, randomly sampled ensemble hurricane paths, known as spaghetti plots, can look messy. At right, a new procedure for reconstructing ensemble paths can show the distribution's full spread and reduce visual clutter. When paths are easy to distinguish, additional information—such as the storm's size and intensity—can be added to the plot ([L. Liu et al., 2018](#)).

Density and violin plots are useful for conveying the data's shape, but it can be difficult to use them to extract specific *cumulative* values (What is the probability of a randomly selected value's being less than this one?), which requires the viewer to visually calculate an integral. In contrast, quantile dot plots provide a generalizable approach for showing a distribution yet also make values easier to read, by providing probabilities not only through the heights of the function (as in the density plot) but also through the (easily countable) number of dots that contribute to that height. As seen in [Figure 22](#) (left), a quantile dot plot represents a distribution using stacked dots that account for a specific portion of the data; in this case, each dot depicts a 5% probability.

Using the graph at the left of [Figure 22](#) as an illustration, imagine that the viewer's task is to decide whether the nighttime low temperature will drop below freezing. With the quantile dot plot, the viewer could count the dots to determine that there

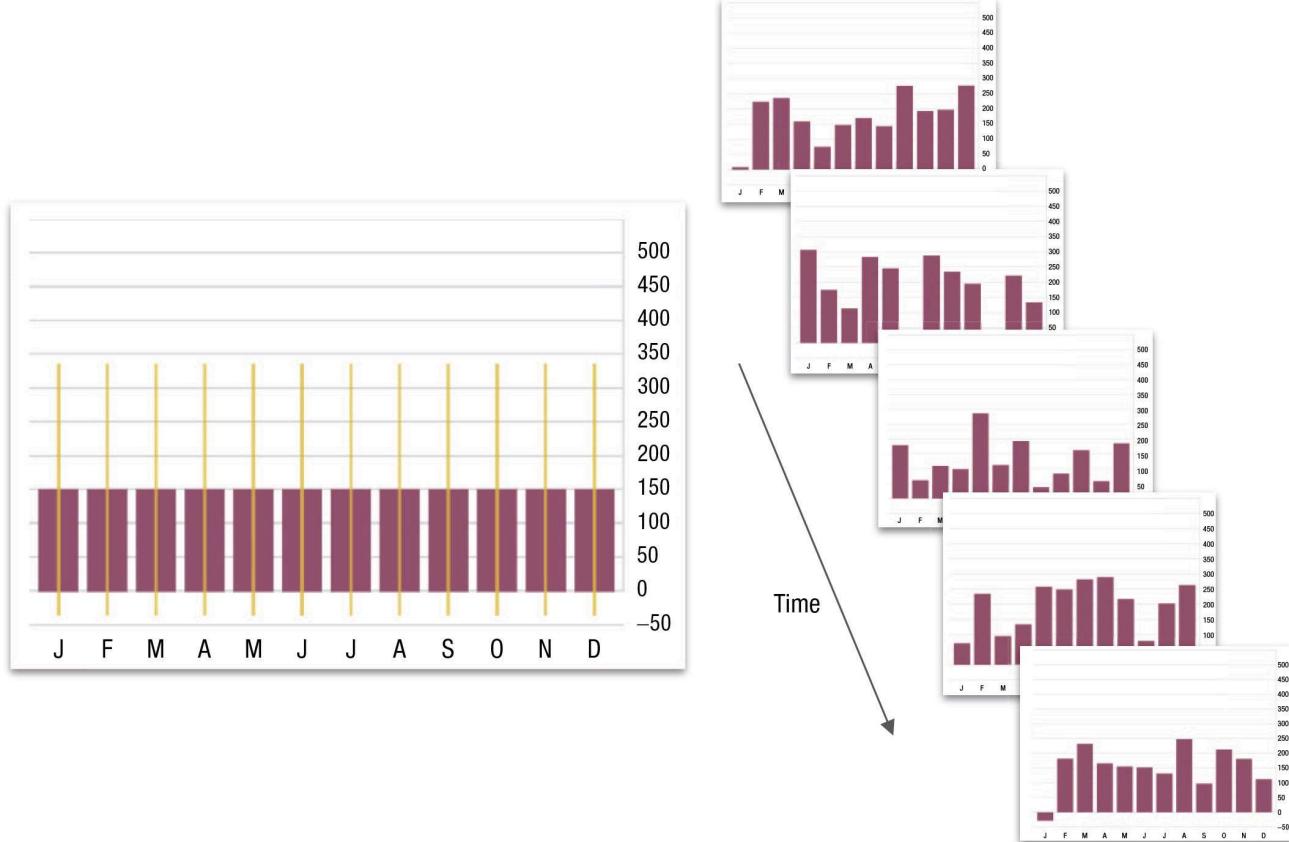
is a 30% chance the temperature will be 32 °F or lower. This same task is much harder with a density plot because it requires the viewer to visually calculate an integral under the curve. Evidence suggests that quantile dot plots improve accuracy and memory compared with density plots ([Hullman et al., 2017](#); [Kay et al., 2016](#)) and outperform summary plots, density plots, and text descriptions of uncertainty for decisions with risk ([Fernandes et al., 2018](#)).

To present their 2020 U.S. presidential forecast, FiveThirtyEight used a variant of a quantile dot plot, which they called a “ball swarm” plot, showing 100 random samples from their model ([Fig. 22](#), right; for details on the creation of these plots, see [Wiederkehr, 2020](#)). Traditionally, dot plots have been used to show distributions of experimental samples ([Wilkinson, 1999](#)). In the ball-swarm plot used by FiveThirtyEight, a total of 100 balls representing 100 hypothetical elections were clustered along an x-axis denoting the margin by which the election could be won by either candidate. Some balls were colored red in proportion to the forecast’s current predictions of Donald Trump’s chance of winning; the remainder were colored blue to depict Joe Biden’s chances.

Showing multiple samples or simulated draws from a distribution can more intuitively communicate that distribution to viewers without training in statistics. Showing those samples simultaneously can lead in many cases to accurate estimates of probability, but in other cases it can lead to new misconceptions. For example, when shown a sample of possible hurricane tracks, viewers often see that set as reflecting the entire population of possibilities (no more and no fewer) instead of merely a subset. When shown visualizations such as those in [Figure 25](#), viewers can overreact if one of the paths directly hits their town ([Padilla et al., 2020](#)), even though the locations between neighboring paths are generally as likely to be hit as the locations underneath a path.

One solution to this problem is to show those samples serially across time instead of simultaneously in space, relying on the limited capacity of working memory to force viewers to naturally abstract those samples into intuitive statistics. Hypothetical outcome plots (HOPs; [Hullman et al., 2015](#); [Kale et al., 2019](#)) animate a set of random draws from a distribution, showing each draw for a short duration (<

500 ms). [Figure 26](#) presents an example set of frames in which each frame is one random draw from the joint distribution represented by the error bars on the left. The frames are shown in a random sequence, which creates an animation that can give viewers an intuitive sense of the uncertainty surrounding the true trend.



**Fig. 26.** Example of a hypothetical outcome plot. An example set of frames (right) in which each frame is one random draw from the distribution on the left. The frames are shown in a random sequence, which creates an animation that can give viewers an intuitive sense of the uncertainty in the true mean ([Hullman et al., 2015](#)).

Presenting animated draws over time has several advantages relative to static approaches. First, joint probabilities (e.g., What is the chance that the rate of disease really is higher in Location A than in Location B?) are naturally expressed as patterns over time, but showing them in a static visualization would require intentionally choosing an encoding capable of presenting joint probability (e.g., a

heat map). The visual system's natural ability to extract summary statistics from temporal frequency ([Hasher & Zacks, 1984](#)) can allow viewers to judge effect size more easily than they could with static depictions of each distribution (e.g., as bars with error bars or violin plots, which do not convey dependence between distributions).

Second, when viewers may be likely to ignore uncertainty in favor of simpler heuristics—for instance, judging an effect size using only the size of the visual difference between the means of two variables, regardless of axis scaling—any static visualization that encodes central tendency runs the risk of allowing users to discount uncertainty in their judgments. In such cases, viewers may ignore uncertainty or use an overly simple strategy, whether a mean is depicted using a direct summary mark, such as a mean temperature forecast presented along with a confidence interval of the mean ([Joslyn & LeClerc, 2012](#)), or implicitly, on the basis of the highest (or widest) point in a density or violin plot of a Gaussian distribution ([Hullman et al., 2015](#); [Kale et al., 2021](#)).

Finally, in visualizations that use the most precise visual channels to show data—for example, a map that uses both vertical and horizontal dimensions of position to show geographic locations—it can be difficult to convey uncertainty because that requires adding a new visual channel (e.g., saturation of position markers) to an already complex visualization. However, as long as a visualization is not already animated, HOPs can be used without requiring the designer to choose a new way to depict uncertainty. This has inspired visualization researchers to use probabilistic animation to show uncertainty in depictions of geospatial data ([Ehlschlaeger et al., 1997](#); [Fisher, 1993](#)) as well as other complex visualizations ([Feng et al., 2010](#)).

Designers might also draw from the best of both worlds by combining animated draws with static displays. Each animated draw of a hypothetical outcome plot could leave a trace that slowly builds into a static display such as a gradient plot, or animated draws could help visually explain a static technique such as a density plot or error bar. *The New York Times* presented animated dots in a simulation to show inequalities in wealth distribution due to race ([Badger et al., 2018](#)). This combination

of animated and static depictions of uncertainty is fertile ground for research at the intersection of psychology, statistical cognition, and data visualization.

## How to communicate health risk

The rapid advancement of technology has made it increasingly easy to collect and analyze massive amounts of health-related data ([Ola & Sedig, 2016](#)). In addition to traditional medical records, doctors and patients now have access to information from population surveys, genomic sequencing, ancestral records, wearable devices, and medical implants. Given the high volume, rapid development, and diversity of this information, it is understandable that making decisions with health data can be complicated. A large body of research is dedicated to examining if visualizations can help doctors and patients incorporate data into health and wellness planning (for reviews, see [Ancker et al., 2006](#); [Garcia-Retamero & Cokely, 2017](#); [Garcia-Retamero et al., 2012](#); [Lipkus, 2007](#); [Lipkus & Hollands, 1999](#)).

A consistent finding from research on the visual communication of health data is that visual aids can clarify numeric information for both patients and doctors. A systematic review of visual aids in health research found that effectively designed visual aids can improve risk assessments, resulting in reductions in health-decision biases, improvements in healthy behaviors, and increased trust in treatments ([Garcia-Retamero & Cokely, 2017](#)). It is noteworthy that the benefits of visual aids are strongest for individuals with low numeracy ([Lipkus et al., 2001](#)) and low risk literacy (i.e., ability to evaluate one's risk; [Ancker & Kaufman, 2007](#)). Low numeracy tends to correlate with less advanced education, which disproportionately affects minoritized groups ([Rodríguez et al., 2013](#))—however, one study with 463 college-educated participants found that 16% to 20% incorrectly answered relative simple questions about risk and probability (e.g., “Which represents the larger risk: 1%, 5%, or 10?”; [Lipkus et al., 2001](#)).

For visual aids to effectively help people reason about health-care data, visualizations must be designed with care ([Fagerlin et al., 2011](#)). The following section provides a summary of what works in the communication of health-care

data, drawing from the most consistent and reliable evidence-based findings to date.

### Convey risks with absolute—not relative—rates

A rational decision maker should not make different treatment decisions based on the way that medical risk is conveyed. Unfortunately, the presentation of medical-risk information has a profound influence on judgments, even for highly educated people. In a study of 235 practicing physicians who were asked to indicate how likely they were to treat patients for hypertension and hypercholesterolemia on the basis of information about the treatments; 41.3% were more likely to recommend a treatment when its effects were framed as a relative reduction in risk compared with an absolute reduction in risk ([Forrow et al., 1992](#)). Absolute risk is the overall probability of experiencing something, such as getting a disease, during some period of time. Relative risk is a comparison between the absolute risk of two groups ([Natter & Berry, 2005](#)). As [Fagerlin et al. \(2011\)](#) illustrated, the change in breast cancer risk from 4% to 2% associated with the drug Tamoxifen can be communicated in a relative format (a reduction of 50%) or an absolute format (an absolute reduction of 2%).

Absolute-risk formats are preferable: Many studies have found that the same treatment outcomes are perceived more favorably when communicated in terms of relative risk as opposed to absolute risk (for a meta-analysis, see [Covey, 2007](#)). Further, relative-risk framing can make changes in risk appear larger than absolute-risk framing (e.g., a 50% relative change incorrectly seems larger than a 2% absolute change; [Akl et al., 2011](#); [Baron, 1997](#); [Forrow et al., 1992](#); [Malenka et al., 1993](#)). This framing bias makes physicians more likely to prescribe interventions when information about their effects is communicated using a relative-risk format (e.g., [Bucher et al., 1994](#); [Lacy et al., 2001](#); [Naylor et al., 1992](#)). A large body of evidence suggests that the absolute-risk format leads to the least biased decisions. Although this has not been tested with visualizations, the evidence points to absolute risk as the best choice.

### Convey ratios with icon arrays and pictographs

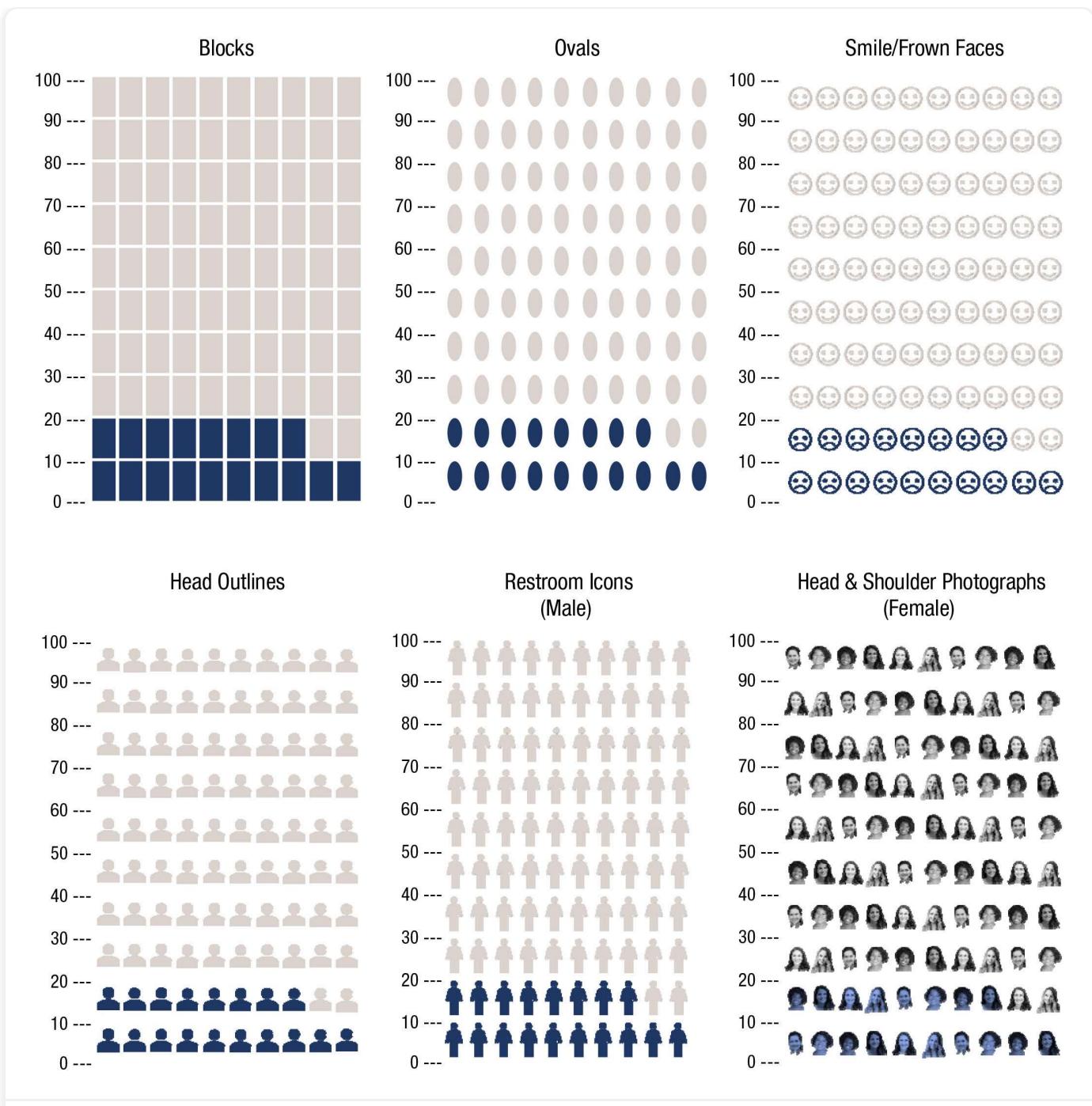
In addition to conferring the previously described benefits of frequency framing, icon arrays help to reduce common errors in health-care communication, such as individuals' focusing on the numerator and neglecting the denominator, an effect called denominator neglect (for review, see, [Garcia-Retamero et al., 2012](#)). For example, when comparing a cancer with a mortality rate of 1,286 out of 10,000 people to a cancer with a mortality rate of 24.14 out of 100 people, undergraduate participants reported that the former cancer was riskier ([Yamagishi, 1997](#)). Researchers have proposed that individuals pay more attention to the relative differences in numerators (in this case, 1,286 vs. 24.14 deaths), even though they should consider the relative ratios (12.86% vs. 24.14% mortality; e.g., [Garcia-Retamero et al., 2012](#); [Yamagishi, 1997](#)). A variety of studies (including [Galesic et al., 2009](#), described above) have shown that icon arrays can reduce denominator neglect by allowing patients to compare relative ratios visually (see [Fig. 24](#); see also [Garcia-Retamero & Galesic, 2009](#)).

Icon arrays can help patients overcome common biases. In one study, participants read statistical information about two treatment options for angina—bypass surgery, which is relatively onerous but 75% successful, and balloon angioplasty, which is less onerous but only 50% successful—and anecdotes about individuals' experiences with each procedure that were either unrepresentative or representative (50% positive and 50% negative for both procedures vs. 75% positive and 25% negative for bypass surgery). Some participants were also provided with icon arrays that reinforced the statistical information. Of the participants who did not see the icon arrays, 41% who received more anecdotes about bypass surgery's positive effects reported that they would likely prefer surgery. Surprisingly, only 20% of those who read fewer positive stories, without the additional icon arrays, responded that they would undergo surgery. When those participants were provided with the icon arrays, the effect of the anecdotes disappeared ([Fagerlin et al., 2005](#)).

To avoid denominator neglect, it is best to use icon arrays that clearly show part-to-whole comparisons by putting both the denominator and the numerator in the same array (as in [Fig. 24](#); [Garcia-Retamero & Cokely, 2017](#)). Designers should avoid

showing only the numerator with icons, and leaving the denominator only in the text, because viewers will be likely to neglect the denominator ([Stone et al., 2003](#)). Icons should also be arranged into a systematic grid that makes them easy to count, because icon arrays that are not arranged systematically are challenging to use ([Feldman-Stewart et al., 2007](#)), particularly for viewers with low numeracy ([Ancker et al., 2011](#); [Zikmund-Fisher et al., 2012](#)).

Designers can also consider using anthropomorphic icons, such as outlines of people. In one study, people saw icon arrays of their personalized cardiovascular risks based on data about their age, weight, and health metrics. Specifically, each individual was shown one of the icon-array formats shown in [Figure 27](#) and was then asked to report their perceived personal risk on a scale from 0 (*extremely small*) to 100 (*extremely big*). People were also asked to report how likely they thought they were to have heart disease or a stroke within 10 years and were later asked how well they remembered the icon array's information. Individuals with higher graph literacy and numeracy who were shown restroom icons had the highest correlation between their perceived risk and actual risk ([Zikmund-Fisher et al., 2014](#)). By contrast, those with low graph literacy and numeracy had the same relationship between their perceived risk and actual risk across all visualizations. In addition, participants' memory for their risk scores was more accurate when they were shown anthropomorphic icons (restroom icons, heads, or photos), particularly for those with low graph literacy and numeracy. Participants also rated the restroom icons as the most preferable. An anthropomorphic font called Wee People provides an easy-to-use set of icons depicting realistic human silhouettes and is available under a Creative Commons license ([Cairo & Klein, 2018](#)).



**Fig. 27.** Icon arrays tested in [Zikmund-Fisher et al. \(2014\)](#). Original figure published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

## Summary of Key Guidelines

- A viewer's visual system can extract broad statistics about the data within a display, such as the mean and extrema, within a fraction of a second. Visualize your data with histograms and scatterplots before trusting statistical summaries.

- Beware common visual illusions and confusions. Failing to start axes at zero can cause viewers to overestimate differences. When plotting data with circles or squares, map the data to their areas, not their diameters. The differences between lines in a line graph are increasingly perceptually distorted as the lines increase in slope. Do not plot intensities on intensities, which causes contrast illusions. Mapping a continuous set of numbers to a spectrum of different hues exaggerates differences that happen to straddle the hue boundaries. For accessibility of color-blind viewers, pair red with blue instead of green.
- Although extracting global statistics is fast, comparisons between subsets of values are slow—limited to only a handful per second. So use visual grouping cues to control which set of comparisons a viewer should make, and use annotation and highlighting to narrow that set to the single most important comparison that supports your message. In a live presentation, rely on language and gesture to illustrate what you see. Do this even when you feel it is not needed: Presenters suffer from a “curse of knowledge” that causes them to overestimate how well others see what they see.
- Avoid taxing working memory by converting legends into direct labels. When possible, integrate relevant text into visualizations as direct annotations. Avoid animations, which typically lead to confusion. Graphical embellishments, sometimes derided as “chart junk,” can distract if unrelated to the data, but if they are related, they can improve viewers’ memory and engagement.
- New visualization formats must be learned, so try to rely on formats that are familiar to your audience. Respect common associations, such as “up” mapping to “more” for vertical position and “more opaque” mapping to “more” for intensity.
- Graph comprehension depends on both bottom-up and top-down factors. Use bottom-up visual salience and top-down direct labels to drive attention to relevant features. Use a graph format that guides viewers to the conceptual message you are trying to convey, respecting their previous experience with graphs.

- When communicating uncertainty to a lay audience, avoid error bars, which can be misinterpreted as data ranges. Instead, show examples of discrete outcomes, either simultaneously or over time.
- When communicating risk to low-numeracy audiences, rely on absolute instead of relative rates, convey probabilities with frequencies (e.g., 3 out of 10) instead of percentages (e.g., 30%), and use well-constructed icon arrays with the same denominator.
- Supporting comprehension and understanding is especially important when the intended audience may have low domain knowledge, knowledge about graphing conventions, numeracy, or working memory capacity.

## Acknowledgments

We thank Evan Anderson and Madison Tyrcha for their invaluable assistance in preparing this manuscript.

## Footnotes

Declaration of Conflicting Interests The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

---

Funding S. L. Franconeri was supported by National Science Foundation Grant IIS-CHS-1901485 and IIS-HCC-2107490. L. M. Padilla was supported by National Science Foundation Grants 2122174 and 2028374. P. Shah was supported by National Science Foundation Grants 2030059 and 2027822 and Institute of Education Sciences Grant R305A170489-F046846. J. M. Zacks was supported by National Institute on Aging Grant R01-AG062438, Office of Naval Research Grant N00014-17-1-2961, and a James S. McDonnell Foundation Opportunity Award. J. Hullman was supported by National Science Foundation Grant 1907941 and a Microsoft Research Faculty Fellowship.