Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science

Sakinah S. J. Alhadad [1]

Abstract
Understanding human judgement and decision making during visual inspection of data is of both practical and theoretical interest. While visualizing data is a commonly employed mechanism to support complex cognitive processes such as inference, judgement, and decision making, the process of supporting and scaffolding cognition through effective design is less well understood. Applying insights from cognitive psychology and visualization science, this paper critically discusses the role of human factors — visual attention, perception, judgement, and decision making — toward informing methodological choices when visualizing data. The value of visualizing data is discussed in two key domains: 1) visualizing data as a means of communication; and 2) visualizing data as research methodology. The first applies cognitive science principles and research evidence to inform data visualization design for communication. The second applies data- and cognitive-science to deepen our understanding of data, of its uncertainty, and of analysis when making inferences. The evidence for human capacity limitations — attention and cognition — are discussed in the context of data visualizations to support inference-making in both domains, and are followed by recommendations. Finally, how learning analytics can further research on understanding the role data visualizations can play in supporting complex cognition is proposed.

Notes for Practice

- For data visualizations to meet the intended purpose of supporting complex human cognition, design choices should be based on human factors research.

- This paper reviews some of these research insights from cognitive psychology and visualization science to inform human-centred design of data visualizations in order to support complex cognitive processes. This is achieved by demonstrating that, depending on the approaches to data visualization, attention and cognition may be facilitated or impaired.

- Evidence-informed guidelines are proposed for enhancing design of data visualization to support inference, judgment, and decision making for researchers, practitioners, and consumers of learning analytics.

Corresponding author [1]Email: s.alhadad@griffith.edu.au Address: Centre for Learning Futures, Griffith University, Nathan Campus, Brisbane, Queensland 4122.

1. Introduction

Visualizing data is an imperative component of scientific practice. Understanding real world phenomena such as learning is facilitated through scientific practice, as shaped through human perception and interpretation. The aspirational goal of scientific practice is for researchers and consumers of research alike to be able to make independent, objective assessments and inferences based on the research methods and data, rather than simply consuming the interpretations and inferences made by the authors themselves. The downside — human perception and cognition, even in scientific pursuit of objectivity and reason, are fallible. The upside — scientific practice *is* about iterative and incremental improvement. In this paper, the critical role of human factors — attention, perception, judgement, and decision making — in understanding and making methodological choices of data visualization, is discussed. In illustrating the ways in which human information processing

capacities can alter the physical representation of data to create meaning, the fallacy of *neutrality* or *fundamental objectivity of data* (Gelman & Hennig, 2017; Woodward, 2011) is also addressed.

Learning analytics research is inherently embedded in educational practice, and thus has a clear role in influencing educational practice and policy. The field, therefore, is uniquely positioned to meet the increasing demand for accessible and relevant evidence-based approaches to support emerging educational practice. Given that learning analytics research often focuses on unobtrusive measures such as log file, or physical traces to complement existing and new educational methods, the research ethos accommodates convergent measures for understanding learning in educational complexity. Hence, critical to the learning analytics field is not only visualization of potentially large datasets, but also of meaningful connections between datasets. Given the complexity of learning data, substantial resources are required to create effective data visualizations for comprehension and communication of the research. Navigating decision making for the optimal data visualization strategies to facilitate understanding of complex, interrelated datasets can be onerous, but is crucially important in enabling discovery of key patterns and connected understanding. The challenge for us is to effectively translate and integrate the existing research findings from related fields into learning analytics research.

As researchers, we tend to be appropriately cautious in the inferences we draw from our research. This caution in drawing conclusions is largely due to recognizing the difficulty in controlling for, or accounting for, all of the potentially important constructs/variables in a complex phenomenon of interest. How then, do researchers make decisions about ways to represent research data as visualization effectively and accessibly, while retaining scientific integrity? The aim of this paper is to provide evidence-based suggestions to inform decision making when considering data visualization methods. Drawing from cognitive psychology, data science, and the human–computer interaction literature, this paper highlights ways in which design decisions for *static* data visualizations can have downstream implications for judgement and decision making for two main purposes: 1) data visualization as a tool for communicating research, and 2) data visualization as a tool for understanding research. The first focuses on the evidence and practical recommendations on the basis of human visual perception and cognition in constructing data visualizations to communicate data as informed by evidence to support cognitive processes of inference. The second part focuses on the importance of data visualization for researchers in gaining a deeper understanding of their research data. Collectively, these can inform decisions for visualizing learning analytics for quality inference-making.

2. Data Visualization as Situated Information

2.1. A priori methodological decision making: The false dichotomy of objectivity–subjectivity of data in scientific reasoning

Data serves the purpose of enhancing our understanding about phenomena in the world, and is contextualized within research methodology. Accordingly, it stands to reason that the objective of "objectivity" in science can be in and of itself philosophically and epistemically misleading. Gelman and Hennig (2017) proposed that to move statistical discourse towards principles of good science, we ought to replace the false (and potentially misleading) dichotomy of "subjectivity" or "objectivity" in statistics discourse with broader collections of attributes. In shifting the discourse within these broader attributes such as *transparency and consensus* instead of "objectivity," and *awareness of multiple perspectives* and *context dependence* instead of "subjectivity," we make explicit and acknowledge the real value of subjectivity in ways that make clear the different ways of knowing in making sense of the world. This view is consistent with what cognitive science tells us about how humans make meaning from information, and recognizes that *data and statistics cannot exist without the situated context of research methodology* in inference making. *A priori* methodological decisions including sampling techniques, experimental (or non-experimental) design, or methods of dealing with missing data or outliers are similarly subjected to complex human cognitive processes. As Woodward (2011, p. 172) argues, "Data to phenomena reasoning, like inductive reasoning generally, is ampliative in the sense that the conclusion reached (a claim about phenomena) goes beyond or has additional content besides the evidence on which it is based (data)." Hence, for epistemological and methodological reasons, it is important for us to recognize in explicit terms that human information processing complexities are the basic conditions of scientific inquiry toward improving the transparency of scientific reasoning.

2.2. Toward human-centred data visualization: Understanding human perception and cognition

One of the key mechanisms for understanding and communicating scientific findings is through visualization of data. In learning analytics, data visualization is a core aspect of understanding and communicating research, and also part of the object of research and practice. Data visualizations are central in learning analytics work related to the provision of data to support learning and teaching to both learners and teachers (e.g., Clow, 2013; Gašević, Kovanović, & Joksimović, 2017; Verbert et al., 2014) in various methods (static, dynamic, table top, multimodal, dashboards). The informative power of data visualization in research and practice is as germane in learning analytics as in other domains of complex research phenomena. Hence,

understanding the mechanisms by which data visualization can support or hinder cognition is extremely valuable for learning analytics researchers and practitioners — there is much for the field to offer in this regard.

Data visualizations are visual representations of complex information, constructed in ways to enhance understanding (Ward, Grinstein, & Keim, 2015). In principle, the purpose and outcome of visual representations is that they can support cognition (Hegarty, 2011). There are many convincing arguments why visualization in research is important. Visualizing data should be part of the core communication strategy, and is as vital to this goal as are research methodology and statistical analyses. Data visualization enables communication of complex patterns that sometimes cannot be, or are cumbersome to convey in words (Ward et al., 2015). Visualizing data facilitates identification of patterns otherwise obscured in text or tables (Cleveland, 1994; Gelman, Pasarica, & Dodhia, 2002; Kastellec & Leoni, 2007; Tufte, 2001; Ward et al., 2015), particularly for multivariate relationships (Meyer, Shamo, & Gopher, 1999; Porat, Oron-Gilad, & Meyer, 2009). However, even with the higher order advantage of visualizing data to aid comprehension, there is evidence for interpretative bias (e.g., Braithwaite & Goldstone, 2013; Newman & Scholl, 2012; Shah, Mayer, & Hegarty, 1999). This necessitates careful and deliberate design and communication when visualizing data. How these external representations of data interact with people's internal cognitive processes for inference-making and knowledge construction is of primary importance in informing strategies for effective data visualizations.

In some ways, we have learnt a lot in the last twenty years or so; in other ways, some of the key challenges of the field — developing ways of systematically collaborating across disciplines, or challenging assumptions or misconceptions about what is effective design for visual representations (Scaife & Rogers, 1996; Wilkinson & Task Force on Statistical Inference, 1999) — are still as pertinent today. In moving forward, there is a substantial body of research that we can leverage to inform our decisions when visualizing data for inference-making. While comprehension of these data visualizations may vary across individuals (e.g., Conati & Maclaren, 2008; Toker, Conati, Steichen, & Carenini, 2013), the underlying mechanisms responsible for human cognition are generalizable.

3. Data Visualization as Communication

In this section, insights from cognitive psychology are synthesized to facilitate the bridging of human information processing and visualizing data as communication. This research forms the basis of evidence-based guidelines for constructing data visualizations to support cognition (see Table 1/Appendix A for a summary). Processing data visualizations, particularly in the context of education (specifically learning analytics), is not only cognitive. Emotions, for instance, play a complex, interactive role in processing information generally (see Yiend, 2010, and Barrett, 2017, for reviews), notwithstanding data visualization. The complex role that emotions play at various levels of engagement with data visualizations is beyond the scope of this paper (see Kennedy & Hill, 2017, for emerging research in this area).

3.1 Attention is limited and selective

3.1.1. Bottom-up versus top-down attentional mechanisms

One of the most robust findings in cognitive psychology is that the human capacity for processing information is limited. In the case of processing visual information, only a small amount of information available to the retina can be processed at any given time. Thus, visual attention and eye movements are often closely coupled, though attention can occur outside of the constraints of the visual field (see van Zoest, Van der Stigchel, & Donkl, 2017). Given complex environments of multiple possible inputs, the cognitive system selectively prioritizes some items for further processing, while others are effectively attenuated. According to the biased competition account of attentional selection (Desimone & Duncan, 1995), the biasing system comprises two attentional mechanisms widely believed to determine attentional priority; a bottom-up and a top-down attentional mechanism. Bottom-up attentional processing is said to be stimulus-driven. This is often driven exogenously by features that are visually salient (e.g., colour, luminance, shape, size, motion), and thought to occur involuntarily. Conversely, top-down attentional control is thought to be goal-directed — this is driven endogenously under volitional control, during which attention is guided by contextually relevant goals, intentions, and expectations. This simple heuristic of dual processes — bottom-up attention as automatic, and top-down attention as more deliberate (Yantis & Johnstone, 1990; Kahneman, 2003; Theeuwes, 2010) — is not as simple as is sometimes portrayed.

People's expectations, prior experiences, and prior knowledge can direct attention in top-down ways that are deliberate, or automatic. Goal-driven and salience-driven attention can be very rapidly modulated by value-driven attention, suggesting that previously learned reward associations can automatically bias information processing (Anderson & Yantis, 2013; Anderson, 2016). Similarly, previously biased prioritization for attentional selection of some features or items in a visual set over others can lead to a learning experience that shapes future attentional guidance and persists through bottom-up saliency manipulations

or changes in task goals (Kadel, Feldmann-Wüstefeld, & Schubö, 2017). Goal-driven attentional selection has also been shown to occur rapidly and automatically, an effect referred to as *contingent attentional capture* as dependent on task context and relational properties of the stimuli (Becker, Folk, & Remington, 2010; Becker, Lewis, & Axtens, 2017; Folk, Remington, & Johnston, 1992). While our understanding of the underlying neural mechanisms (and time course) of the interconnection between bottom-up and top-down modulation of visual attention is still being explored and debated (Ansorge, Horstmann, & Scharlau, 2010; Awh, Belopolsky, & Theeuwes, 2012; Pinto, van der Leij, Sligte, Lamme, & Scholte, 2013; Theeuwes, 2010), the view that both stimulus-driven and goal-driven factors ultimately interact to determine visual selection is fundamentally agreed upon (Connor, Egeth, & Yantis, 2004; van Zoest, Donk, & Theeuwes, 2004; Wolfe, 1994). That is, fundamentally, *salience of visual features of data visualization influences attentional capture, and prior experience and knowledge can guide visual attention,* and these occur in concert.

Why is this useful to know? Intuitions about good or desirable data visualization design practices may not always be consistent with the evidence from the research about human cognition. People (experts or non-experts) sometimes perceive certain visualization features to be effective design choices for visual representation of data, when evidence suggest instead that these could impair attention and cognition when processing graphs for meaning (Hegarty, Smallman, Stull, & Canham, 2009; Smallman & St. John, 2005; Zacks, Levy, Tversky, & Schiano, 1998). Prior learned graphical conventions in scientific practice can shape the pattern of visual inspection, sometimes referred to "visual routines" (Michal & Franconeri, 2017). Generally, viewers consistently orient their visual attention to text elements in visualizations, particularly during early stages of viewing (Matzen, Haass, Divis, & Stites, 2017), highlighting the importance of considerations of salience of the text features of visualizations in ways that complement and support the understanding of the data. Certain data visualization methods have also been shown to activate automatic retrieval of prior learning about graphical conventions and their relation to quantitative concepts (e.g., straight line = linear relationship; Shah, 1997). These selective patterns of visual inspection of graphs suggest that a general conceptualization of *graphical literacy* (Okan, Galesic, & Garcia-Retamero, 2016; Shah & Freedman, 2009) — the ability to read, construct, and interpret visual displays of data — may be shaped by disciplinary conventions. Graphical comprehension therefore appears to be determined by both bottom-up and top-down factors.

Thus, in informing our decisions about designing data visualizations to support cognition, it is important to note that some decisions that may be intuitive for accessibility or attention may actually be counterproductive, or even overshadowed by more complex, higher-order cognitive processes, such as goals and knowledge. These cognitive processes are cyclical, in that what drives attention early in visual processing is also then influenced by other factors that influence attention later. In order for visual stimuli to be acted upon, they must first be attended to, before higher-order cognitive processes such as working memory, judgement, and decision making can occur. The visual perception of the features of the visualized information is then represented in working memory. These representations are continually updated during graphical comprehension through a highly limited visual working memory capacity as the reader explores the visualization, thereby constraining the amount of information that gets deeper processing. To meet our purpose of using data visualization as a means to communicate research effectively, we need to understand these human information processing capacity limitations in order to inform design decisions of data visualization in ways that support cognition.

One advantage of understanding the human information processing capacity limitations that are relevant for data visualization is to inform approaches in managing cognitive load. That is, these strategies are aimed at reducing "information overload" (Shneiderman, 1996), or "cognitive load" (Ayres & Sweller, 2014; Sweller, van Merriënboer, & Paas, 1998) such that performance is not degraded. As with most research, cognitive-load-theory-based research has its strengths and weaknesses (Kirschner, Ayres, & Chandler, 2011), but certain methods of managing cognitive load for information processing and learning are supported by robust findings. Managing cognitive load based on capacity limitations means ensuring that the cognitive resources for processing data visualizations should best be allocated to essential processing for understanding the data and inference-making, rather than for figuring out basic features of the visual display. Critical to cognitive load theory is that extraneous cognitive load places the load on scarce working memory resources, thereby interfering with processing of essential material for learning and comprehension. When making decisions about reducing extraneous processing, the consideration of the interaction between bottom-up and top-down attentional mechanisms becomes practical — this is not a trivial task. Oversimplifying data representation would ease cognitive load, but can reduce scientific integrity or be misleading in many ways. Over-complexity in data representation could impair attention and cognition due to capacity limitations (Kosslyn, 2006; Tversky, 2005). At minimum, there are key identifying features of data visualization that are core to graph comprehension, and that could be said to be essential.

3.1.2. Dealing with visual clutter: Coherence, Chunks, Contiguity, Segmenting, Signalling

A key barrier to cognitive processing is visual clutter. Visual clutter can be characterized as an excess of information or visual properties in a visual display that result in chaotic, or high-density layout, and is a proxy for visual complexity (Rosenholtz,

Li, & Nakano, 2007). Generally, visual clutter should be avoided in data visualizations as it can severely impair cognition and task performance. Visual clutter has been shown to increase errors in detection and judgement, and additionally, also increases the confidence with which people make decisions — both outcomes are detrimental for data visualization (Baldassi, Megna, & Burr, 2006). Some properties of design that may contribute to visual clutter are: 1) having too much or excessive information, and 2) a lack of organized structure or schema for representation of data.

To address the first source of visual clutter — excessive information — a design approach is to only include information that is necessary for the intended purpose. The rationale for employing this strategy is to weed out irrelevant materials such as excessive details, visual embellishments, or dense layout, enabling the reader to allocate more cognitive capacity for essential processing. This, the *coherence principle*, is supported in 23 out of 23 experimental tests, with a healthy median effect size of 0.86 (Mayer & Fiorella, 2014). These decisions do depend on the type of visualization for comprehension. For example, labelling features on axes of line graphs, particularly for time-series visualizations, is essential for providing contextual information from which meaning can be constructed (Kubina, Kostewicz, Brennan, & King, 2017). For coherence to occur, one has to further consider the goal or task at hand — even a seemingly simple decision of including a comparative line in a line graph can hinder essential processing of key information for the task at hand (Alhadad, 2016). Thus, the consideration of the intended goal or outcome of a visualization is critical in identifying essential material such that the external representation of the visualization is more aligned with the internal representation in cognitive processing.
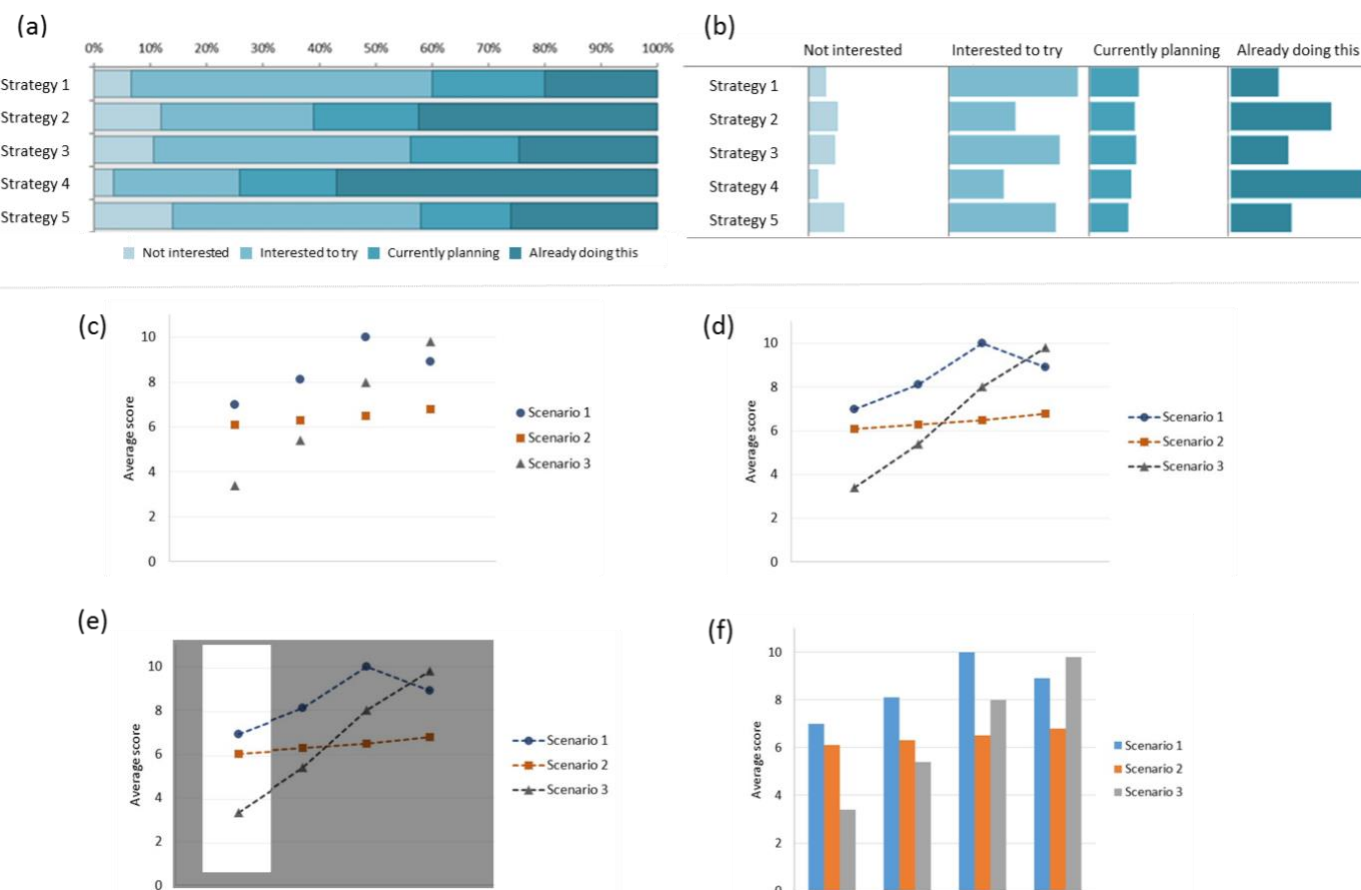

Figure 1. Illustration with fictitious data. Top panel: Comparing perceptual discriminability by choice of (a) presenting data using a stacked bar graph versus (b) a discrete horizontal bar graph to aid perceptual and cognitive discrimination of relative differences across and within categories. Bottom panel: Comparing visual features of a (c) dot plot with redundant encoding (shape/colour) across within categories, (d) as compared with dotted lines to aid tracking within each scenario, (e) with added box as a perceptual chunk and scaffolding to highlight relational differences across discrete scenarios, (f) a bar version of the same information for comparison between types of visualization.

Some decisions for visualizing data are simpler. Assessment for information loss is visually observable by sheer exploration. For instance, we know that humans are not very good at judging proportions if there are more than three cumulative

categories if presented in certain ways (e.g., stacked bars, pie charts; Tufte, 2001). Or if one requires comparison across multiple graphical representations, then the scale needs to be consistent for more accurate comparative analysis. Information about relative differences are often lost when visualized in ways that occlude perceptual discrimination of magnitude differences. As can be seen in Figure 1, perceptual discrimination of relative differences for each discrete category on the y-axis is really only perceptible when presented discretely (upper right panel), rather than as a continuous cumulative stacked bar (upper left panel). These examples highlight some strategies for visual features selection (e.g., colour, layout, type) that can improve directing of attention to important information as aligned to the purpose of the visualization (e.g., perceptual comparison of differences, trends over time).

To address the second source of visual clutter — lack of organization to structure, particularly for complex research designs and data — one approach is through *chunking* (Baddeley, 2003; Gobet et al., 2001; Gobet, 2005). Chunking refers to grouping elements into larger or broader units based on their meaning, learned associations, or cognitive set. That is, complex data visualizations may be segmented into more manageable, meaningful chunks of smaller units of information. This may be done perceptually (in this case, visually) or semantically. Perceptual chunking strategies include the use of common visual design parameters like colour, shape, or location to facilitate "chunking" of information into groups on the basis of similar cognitive sets. For example, applying the Gestalt principle of similarity through use of shared colour or shade allows the reader to associate the similarity as shared groupings to aid comprehension. Redundancy in combining these features (such as combining colour and shape; e.g., blue circles vs. orange squares) can strengthen the segmentation for more efficient processing (Nothelfer, Gleicher, & Franconeri, 2017). The Gestalt principle of connectedness suggests visualizing trends over time with lines instead of bars since lines perceptually connect otherwise discrete entities compared to bar graphs (Ali & Peebles, 2013; Zacks & Tversky, 1999). Put simply, the visual features are represented in a way that relates to the conceptual relations, which can be shaped by top-down goals, purpose, or graphical *schema*, to boost a reader's general knowledge about the graph (Pinker, 1990; Shah & Hoeffner, 2002). Thus, consideration about how the elements of a visual display can be grouped together into psychologically meaningful entities can support attention and cognition (Moore & Fitz, 1993; Pinker, 1990).

Perceptual chunking strategies can be supplemented semantically with text to strengthen retrieval cues for relevant knowledge structures to aid comprehension. Chunking in graphical comprehension occurs cyclically, where visual pattern recognition encodes a visual chunk, and then translated and related to other visual or cognitive referents before integration (Carpenter & Shah, 1998). However, this efficacy of chunking may not be similar across levels of relevant expertise. According to a model of chunking for expertise, novices tend to be less equipped to integrate complex visual information into meaningful chunks compared to experts (Gobet, 2005). This suggests that in order to ease comprehension of data visualization, chunking could be explicitly facilitated through integration with text in order to orient the reader's attention to important connections between visual components and concepts; where needed, chunks can be sequentially scaffolded to support comprehension. Indeed, incremental presentations of visualizations have been demonstrated to enhance readers' speed of exploring and comprehending large datasets (Fisher, Popov, Drucker, & Schraefel, 2012).

3.2 Data visualizations can be integrated with text to support inference-making

Data visualization in research often requires a figure caption to support understanding of the data. Hence, in accordance with the norms of the scientific practice of reporting data, text is inherently integrated with visualization to provide contextual information to aid cognition, while inferential text is provided in the body, or sometimes in the figure supporting text. Considerations of what to include or exclude in the figure caption are important for three key reasons. The first is a matter of pragmatism: simple space constraints in publications. Further, one would not want a figure caption as long as a paragraph as the start and end of the figure versus the body text would not be clear.

Secondly, considering what information should be spatially proximal to the data visualization has implications for attention and cognition. This spatial distancing of the essential information from the visualization can elicit the *spatial contiguity* effect (Mayer & Fiorella, 2014). Separation of the visualization and supporting text increases the likelihood of attention split between the materials essential for understanding the information across spatially disparate locations (Ayres & Sweller, 2014). This increases the cognitive load for information processing and reduces the reader's ability to understand the data visualization effectively. Placement of the key information in spatial proximity to the visualization eliminates the need to rely on working memory to integrate spatially disparate information.

Thirdly, consider the content of the textual information integrated with the visualization. The goal is to optimize coherence between the visualization and the supporting text in ways that aid understanding and recognize the interaction of top-down and bottom-up mechanisms of information processing. One approach is to *signal* the important aspects of the visualization. *Signalling*, in the context of managing cognitive load for graphical comprehension, refers to using text to direct attention to support cognition (Meyer, 1975; Schneider, Beege, Nebel, & Rey, 2018). This maybe done using headings, axis labels,

summaries, or directional language such as "first,… second,… third…" (Mautone & Mayer, 2007). Other semantic scaffolding includes understanding and making inferences about any uncertainty in the data (addressed more extensively in the next section), relationships across multiple visualization panels, or relationships between theory or important concepts and the visualized data.

Table 1. Summary of evidence-informed guidelines for visualizing data, with practical recommendations from cognitive and psychological research

Research evidence	Practical recommendation
Designing data visualizations for attention: Visual attention is limited and selective	
1. Salience of visual features influences attentional capture (Desimone & Duncan, 1995; Theeuwes, 2010)	• Ensure the visual features are designed in ways that do not detract from, but rather support the understanding of important information
2. A person's prior experience and knowledge can guide visual attention (Anderson & Yantis, 2013; Becker, Folk, & Remington, 2010; Desimone & Duncan, 1995; Theeuwes, 2010)	• Seek to understand the conventions of data visualizations in a particular research and practice domain; choose visualization design strategies to complement familiarity of the knowledge domain (see also recommendation 5) • Enhance this by scaffolding the narrative to support the understanding of the data visualization (see also point 7)
3. Chunking improves processing of complex information through minimizing cognitive load and working memory capacity limitations (Baddeley, 2003; Gobet, 2005)	• Apply visual design principles to group visualization features to foster generative processing or promote chunking (e.g., using colour or shape to cluster by related meaning [see also point 5]; or visualize complex relationships in small multiples [see also Table 2])
4. Avoid visual clutter as it can severely degrade attention and impair comprehension (Baldassi, Megna, & Burr, 2006; Rosenholtz, Li, & Nakano, 2007)	• Exclude irrelevant materials such as excessive details, visual embellishments, unnecessary information, or cluttered, dense layout; this enables the reader to allocate more cognitive capacity for essential processing (*coherence principle*)
Designing data visualizations for cognition: Use text with visualization to support cognition	
5. Prior experience and knowledge can guide visual attention (Anderson & Yantis, 2013; Becker, Folk, & Remington, 2010; Desimone & Duncan, 1995; Theeuwes, 2010)	• Organize information in ways that are aligned to conventional mental models to provide retrieval cues for knowledge structure to aid inference-making
6. When text and visualizations are spatially distant, attention is divided, and people process these less deeply (Mayer & Fiorella, 2014)	• Keep visualizations and supporting text close together ; this helps the reader build important connections between corresponding words and visualizations (*spatial contiguity principle*)
7. Signalling with text can shape and direct the reader's attention and thoughts about the visualizations (Mautone & Mayer, 2007; Mayer & Fiorella, 2014)	• Use text to help direct the reader's attention to key information in the data visualization to aid comprehension (*signalling + coherence principle*)

Signalling can also be done by visually highlighting certain aspects of the graph or text to direct attention. Ultimately, to guide the cognitive process of structuring the information into a *coherent* representation, visual and textual signalling may be combined. An example of coherent signalling includes narrative or storytelling aspects as a means to signal important aspects of the data to support comprehension and inference (examples in learning analytics: Echeverria et al., 2018; Rau, 2017). Similar principles of *spatial contiguity*, *coherence*, and *signalling* are applied in this paper (see captions for Figures 1–4).

While some decisions for visualizing data are simpler in isolation, holistic considerations are more challenging. Care will be needed to communicate research findings appropriately to audience and purpose while retaining scientific and data integrity. In some ways these strategies work by minimizing the effect of extraneous processing on cognitive load; in other ways they

work towards dealing with *germane* load (Cierniak, Scheiter, & Gerjets, 2009; Paas & Sweller, 2014), or *the working memory resources devoted to the internal components of the task*. Hence, the collective efficacy of the overall design on cognitive processes ultimately will need to be considered in concert. While these suggested ways of managing cognitive load are discussed in the context of static data visualizations, the principles are generalizable to dynamic or interactive presentations of data. The classic critique of tool versus design strategies à la "death-by-PowerPoint" applies here: while technology may be limited in some ways, there are often ways to design optimally for human information processing.[1]

4. Data Visualisation as Methodology

4.1 Visualizations can reveal important insights for research

Visualizing data in research is arguably one of the most critical steps in understanding the underlying properties of the data. Data visualization and statistical analyses are best performed as a unified process, as they complement each other to help us develop insight, understanding, and reasoning about the data. Executing one without the other, or if assessing each without reference to the context of its research methodology, would be a disservice to the scientific goal of understanding real-world phenomena. Indeed, visualizing the data is often recommended to evaluate the assumptions underlying univariate or multivariate data (Hair, Black, Babin, & Anderson, 2010; Gelman & Hennig, 2017) to understand the pattern of the data, as well as the nature of measurement error. Understanding the nature of variability or uncertainty in data is crucial to informing our inference-making in less biased ways.

In addition to helping one understand data variability, Pastore, Lionetti, and Altoe (2017) further suggest deliberately visualizing data in multiple ways as an inferential tool. Through visualizing the data in various methods or representations, one is better able to evaluate inferences made on the basis of the various ways in which uncertainty may be concealed — each method with its own limitations — and subsequently enables conferring plausibility to each inference and decision.

Generating a visual representation of the data in certain formats prior to statistical analysis offers a clear means of diagnostic examination of issues that may impact the decisions for appropriate statistical analysis or to deepen understanding of the statistical outcomes. Visually inspecting the data allows the researcher to identify issues such as violation of assumptions for particular *a priori* statistical techniques, data anomalies, among many other potential issues otherwise invisible through other data analytics methods. While it is often recommended that researchers should select methods of statistical analysis *a priori* (Wilkinson et al., 1999), suitability of the statistical techniques may ultimately be informed by the nature and shape of the data. Should assessment of data visualization for statistical assumptions reveal violation of any required properties for the statistical technique, the researcher will have to reassess the *a priori* statistical plans in recognition of the true underlying nature of the data collected.

Hence, data visualization can be key to making judgements about data integrity, quality, and uncertainty. Regardless of the methodological choice for representing data and its uncertainty — both statistically and visually — the consensus is clear on the importance of representing and communicating data through visualization as complementary to statistical analysis. The recommendation to do so has been advocated for as critical to understanding and communicating about data for over 20 years, possibly more (in psychology; Wilkinson et al., 1999), and still pertinent today. Of note are the following recommendations involving the role of visualizing data in research.

The first recommendation highlights the importance of visualizing data as preliminary inspection prior to statistical analysis, as data visualization can offer insights about the nature of the data that statistics are unable to:

> As soon as you have collected your data, before you compute any statistics, look at your data … if you assess hypotheses without examining your data, you risk publishing nonsense … Graphical inspection of data offers an excellent possibility for detecting serious compromises to data integrity. The reason is simple: Graphics broadcast; statistics narrowcast. (Wilkinson et al., 1999, p. 599)

The second recommendation highlights the value of adding representations of data uncertainty in visualizations (Wilkinson et al., 1999, pp. 606–607), further advocating that "Many of these procedures are found in modern statistical packages. It is

---

[1] One example of a data visualization debate occurred in 2005 when Edward Tufte critiqued NASA's use of PowerPoint, arguing that visual clutter obscured critical information (see https://www.inf.ed.ac.uk/teaching/courses/pi/2016_2017/phil/tufte-powerpoint.pdf). Jean-Luc Damont responded to the critique (see http://web.mit.edu/5.95/readings/doumont-responds-to-tufte.pdf). This line of debate continues (see critique by Paul Ralph: https://theconversation.com/why-universities-should-get-rid-of-powerpoint-and-why-they-wont-43323; and response by Jared Cooney Horvath and Jason Lodge: https://theconversation.com/its-not-powerpoints-fault-youre-just-using-it-wrong-43783).

time for authors to take advantage of them and for editors and reviewers to urge authors to do so." This of course, depends on how the data is visualized, hence the need to visualize data in ways that represent both precision and uncertainty. Graphing summary statistics only serves to facilitate interpretation that can already be gleaned from statistical summary values. The following section discusses some of these related key values of visualizing data for learning analytics research (see Table 2/Appendix B for a summary of evidence-informed recommendations for visualizing data to support understanding and communication of learning analytics research).

4.1.1 Value of visualizing data to understand underlying data properties

Recommendations to visualize data before, as well as during statistical tests to support cognition is central to scientific practice regardless of disciplinary background. The reason for this is clear — visualizing the data enables one to inspect it in ways that statistical tests can never reveal. Anscombe (1973) convincingly argued for the importance of doing so by illustrating that datasets with identical statistical properties may produce very different underlying patterns of dispersion and effects when visualized. More recently, Matejka and Fitzmaurice (2017a) illustrated this same point compellingly with a more sophisticated simulated annealing technique.[2] They further demonstrated that depending on the type of visualization, the representation of data dispersion and uncertainty may be inherently occluded. As illustrated in Figure 2, visualizations with similar statistical properties can look very different when visualizing the underlying data distributions.

Decisions for data visualization therefore are multifaceted. For instance, Tukey-style box plots may be more effective than bar graphs as they convey additional information about interquartile ranges, the median, and potential outliers (Krzywinski & Altman, 2014). They are, however limited in their capacity to represent uncertainty. Figure 2 demonstrates this: the depicted violin plots illustrate differences in data uncertainty visibly whereas the box plots conceal information about uncertainty. This has implications for making decisions about the type of data visualization, and understanding what information or insights each type is able to reveal (or occlude). As a consequence, the choice of visualizing the data may help or hinder the understanding of data and of subsequent inferences made. Choices that reveal the nature of data dispersion for inferences about precision or uncertainty is critical in understanding the data more deeply.
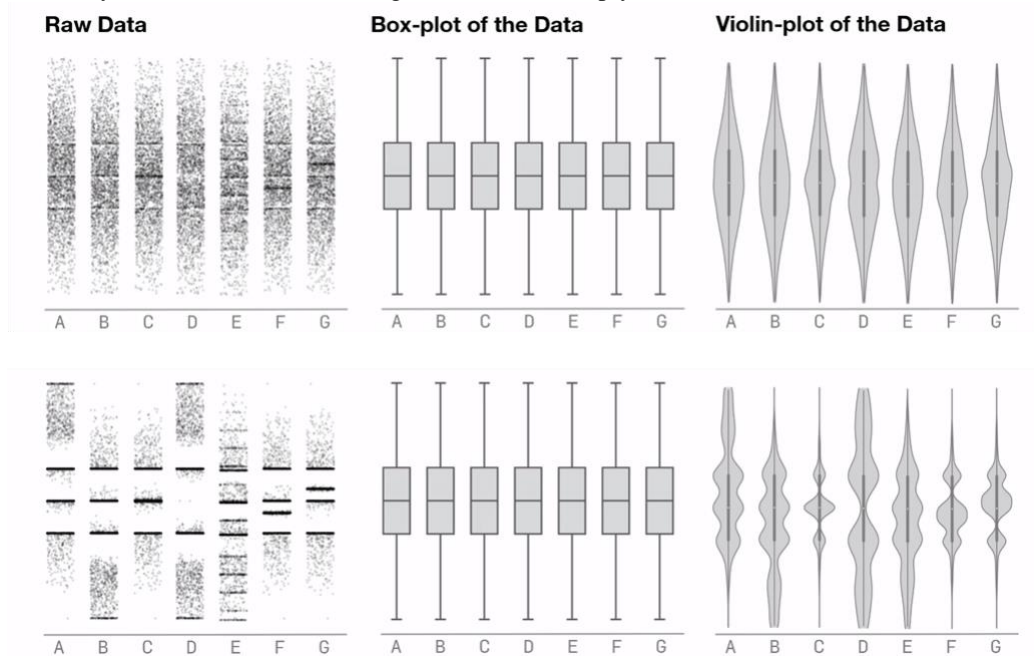

Figure 2. Illustration of the importance of decisions about methods of visualizing data uncertainty. These seven different data distributions (A–G) each have the same 1st quartile, median, and 3rd quartile values. Looking top to bottom across the two panels: Top panel — where the distributions are relatively similar, the differences between box plots and violin plots are minimal. But when data distributions are different (noting similar central tendencies properties), the different distributions of the data are visible in the raw data and through violin blots, but not through Tukey's box plots. Plots are adapted and reused with authorial permission (Matejka & Fitzmaurice, 2017b).

---

[2] For related open data, R and Python code, animated visualizations, and details on the simulated annealing technique (Matejka & Fitzmaurice, 2017a; 2017b), go to https://www.autodeskresearch.com/publications/samestats

Decisions about the type of visualization, method of representing uncertainty, and associated statistical and/or methodological information can still potentially lead to errors in interpretation and subsequent inference. The nature of misinterpretations may be shaped to some extent by graphical literacy (Scown, Bartlett, & McCarley, 2014; Stofer & Che, 2014; Tak, Toet, & van Erp, 2014), though some misinterpretations appear to occur regardless of prior knowledge (Peebles & Ali, 2015; Pentoney & Berger, 2016), or in some cases specific to types of visualization (Lem, Onghena, Verschaffel, & van Dooren, 2013; 2014). Visualizations with proportions and probabilities are particularly notorious for skewing interpretations and inferences in ways that do not address the true nature of the data (Gottlieb, Weiss, & Chapman, 2007). Some are even possibly deceptive (e.g., Gigerenzer & Edwards, 2003; Krämer & Gigerenzer, 2005). Even simple visualization strategies, such as ensuring the primary dependent measure axes are constant across graphs for comparison, make a difference. For example, presenting percentage change across groups without visualizing the dispersion of data, and the sample size for each cell, inhibits making informed interpretations. The percentages or proportions represented become non-meaningful for interpretation and thus bias understanding and inference-making (Goldstein & Gigerenzer, 2002; Gottlieb et al., 2007).

4.1.2 Value of representing data in multiple ways

Multiple representations of the data can reveal insights that may be obscured with just a single representation. Of course, given a single research method and set of data, there may be multiple ways to visualize the data, each choice offering a slightly different view, and subsequently, perspective and insights into the underlying data properties and patterns. At minimum, we know that visualizing the data closest to representing the raw individual data points collectively (as per Figures 2 and 3) ultimately enables visual inspection of uncertainty that summarized visual representations (Figure 5) cannot offer; though complementing both types likely deepens the understanding of the phenomenon. Where temporality is critical to understanding the phenomena of interest, the data will need to be collected with this in mind, such that temporal visualization of the data is possible (Lund, Quignard, Shaffer, 2017; Riel, Lawless, & Brown, 2018; Thompson et al., 2013). Thus, *a priori* decisions about research methodology can influence the depth of insight that can be revealed through multiple representations of the data. In an example tutorial of epistemic network analyses, Shaffer, Collier, and Ruis (2016) illustrate clearly this affordance an *a priori* methodological choice can offer (see Figure 3). Based on these data visualization choices, one can see how certain aspects of information useful for understanding and making inferences from the data can be obscured or revealed. These multiple visual representations of the data may also be hierarchically or sequentially presented. Connecting back to Section 3 of this paper, these hierarchical methods of representing the data can be a method of managing cognitive load, both in terms of static and dynamic representations (see Elmqvist & Fekete [2010] for overview on hierarchical visualization and Victor [2011] for a discussion on scientific communication as sequential art). Collectively, the multiple representations reveal a deeper picture about the research, and in some ways can also be a strategy to drive storytelling when communicating research.

The use of bivariate relationships in educational and learning science research is something to be treated with more caution in application and interpretation. The immediate risk to the field is threefold: 1) the risk of generating and proliferating *spurious* correlations has potentially serious implications for our understanding of learning; 2) the risk of perpetuating misleading inferences such as drawing causal inferences from correlations generates problems for knowledge; 3) last, but certainly not least, the risk for educational practice is the extended inferential issues described in 1) and 2) when making decisions in educational practice on the premise that the research evidence is "good research." The third is arguably the worst potential unintended consequence of this practice, as it may immediately impact on student learning and well-being. Other than the problem of reducing complex (learning) relationships to a single predictor and an outcome, researching with, or visually representing data on a bivariate plane is problematic for many other reasons. Reporting of, or representation of potentially complex relationships in learning research as oversimplified bivariate relationships is likely to perpetuate myths about learning, unless explicitly addressed. Understanding the cognitive processes of judgement and decision making with correlations enables us to make informed decisions about designing and understanding representations of bivariate data.

The capacity of people to understand scatterplots and to infer relationships from them has been shown to be problematic in a number of ways. It is common for people to consistently underestimate the magnitude of correlation coefficients, or the strength of the relationship between two variables (see e.g., Cleveland, Diaconis, & McGill, 1982; Strahan & Hansen, 1978; Sher, Bemis, Liccardi, & Chen, 2017). People sometimes misjudge correlation strength from data cloud density alone, and have trouble judging the effect of scatterplot contaminations such as the presence of outliers on the correlation magnitude (Cleveland et al., 1982; Meyer et al., 1999). As such, the accompanying information about the population parameters — magnitude of the relationship, pattern of dispersion, and error — would serve to aid interpretation of the data.
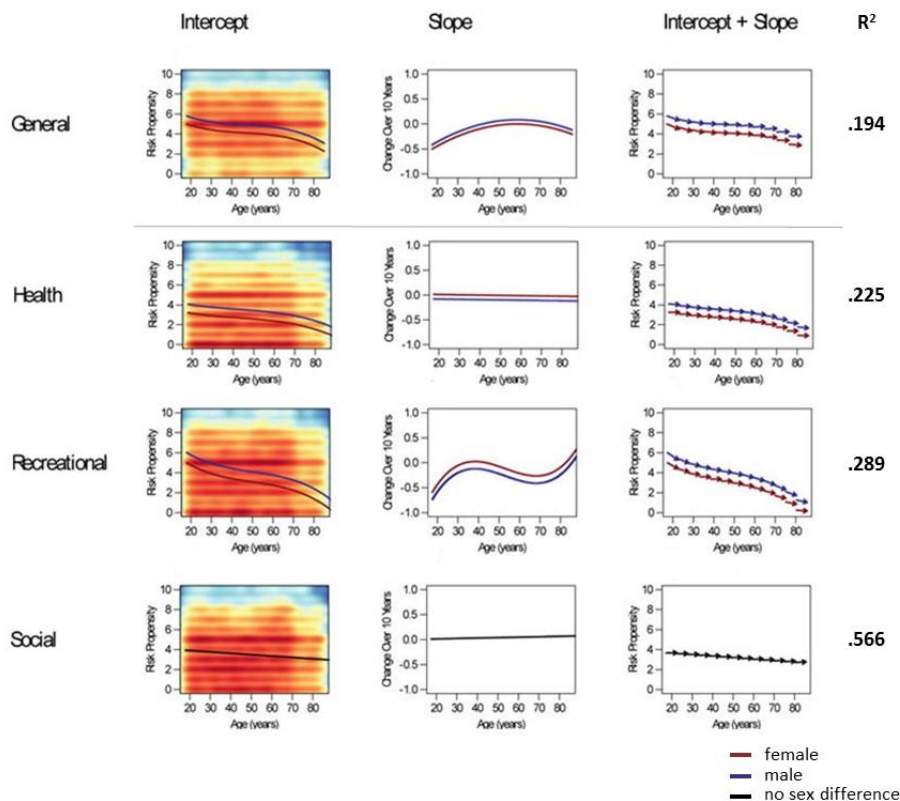
Figure 3. Extracted from Figures 2, 6, and 7 of Shaffer, Collier, & Ruis' (2016) tutorial. Top panel summarizes three student centroids on four dimensions, bottom panels reveals the respective individual patterns of connections for each discourse network across the four dimensions.

Another common misinterpretation is the inference of causality from correlation or regression analyses. Neither correlation nor regression allows causal inferences, however regression does allow inferences about a predictive relationship that cannot be made from simple correlations. While in some research circumstances, bivariate relationships may be logical first steps or the only possible analytical method, the reporting of these results could benefit from the understanding of these principles in understanding and communicating the research findings in ways that do not make over-claims, and allow appropriate non-causal inferences that acknowledge the methodological limitations.

Similarly, it is important to recognize that no matter how well this data is visualized, the methodological limitation of assessing bivariate relationships where, more often than not, these relationships are likely to be mediated or moderated by a "third factor" or more, drive equally limited interpretations. The human brain will create meanings from these paired, sometimes directional relationships, as limited as they are — that one thing drives another, as a superordinate mental model. This circularity is likely to occur as a function of the human cognition in meaning making, particularly in a restricted range of data and methodology. This highlights the unavoidable, unintended consequence of restricted approaches to restricted visualization to that of restricted inference. Thus, if one chooses to analyze or visualize data in restricted ways, this should be acknowledged and recognized when making inferences, as misleading inferences may still occur despite the best intentions. Particularly in the context of learning analytics, where assemblages of multiple sources and time points of data could be leveraged to understand learning processes, choosing to study the effects of a predictor on an outcome in isolation explicitly risks ignoring potentially important effects of other predictors on that relationship. Anscombe (1973, p. 21) makes this very point in his seminal paper: "The likelihood that we fool ourselves by only carrying out some ordinary regression calculations is much greater too. Usually when there are many 'independent' variables they are mutually related and we are interested in performing regression on subsets of them, possibly by a 'stepwise' procedure; so even the standard calculation is not so simple."

Figure 4. Illustration of the importance of data visualization in predictive relationships by way of small multiple visual representations. The R^2 values on the right are insufficient to convey the patterns of relationships, and the three data visualization panels — intercept, slope, and intercept + slope — illustrate how analysis of the bivariate relationship of interest can be context-, cohort-, and sex-dependent, allowing for a more nuanced understanding of the relationship between age and risk-taking propensity. Figure from Josef et al. (2016), adapted and reproduced with authorial permission.

It is clear that visualizing data, either bivariate or multivariate, can reveal insights about the data that numbers and statistics alone cannot offer. A quick search of the learning analytics literature across the main sources (Journal of Learning Analytics and the Proceedings of the Learning Analytics and Knowledge conferences) reveals that out of the 26 papers that used correlational or regression analyses, only nine visualized the data in graphical representation, while 20 others (including the nine papers with graphs) reported details in the form of a table (two had neither). This is certainly an avenue to push for complementary data visualizations to aid understanding of the research in its communication. Figure 4 illustrates two ways in which viewing visualization of relationships beyond two superordinate variables can influence methodological choices about measurement in research. First, note how R2 values (far right column in Figure 4) are insufficient to convey the patterns of relationships — the values indicate the amount or magnitude of variance attributed in the model, but do not illustrate the pattern of variance. Second, note how the visualizations in the top panel illustrate the generalized relationship between age and self-reported propensity for risk-taking indicate a curvilinear relationship over age groups with a general downward trend with increasing age. Third, comparing this general relationship of age and risk propensity to those measured in various domains or contexts, the visualizations illustrate the nuanced pattern of change over time and sex, as contextualized in each domain (Health, Recreational, and Social). This nuanced understanding would otherwise be occluded from a simple bivariate relationship, and from a lack of data visualization.

4.2 Importance of visualizing data uncertainty
If the goal of data visualization is to represent and communicate data effectively and accurately, then visualizing data uncertainty is essential. All too often, visualization of data uncertainty is omitted; doing so robs the reader of the chance to do an independent assessment of the data for inference. In the context of visualizing data uncertainty, the concept of uncertainty may be understood as a composition of various sources of uncertainty. Uncertainty in data can reflect imprecision of measurement or resolution, degree of subjective influence in the data, artefacts or background noise, or outliers or deviation

from a true value, among others. Advancements in visualization software have given researchers better access into the underlying nature of data uncertainty — error, variance, and patterns of uncertainty — thereby better enabling them to inspect, explore, and evaluate the data — small or large, highly complex — in more sophisticated ways. This advancement is particularly advantageous now that data of increasing size and complexity is becoming more common. Data uncertainty can arise from many sources, with various meanings and implications as a result. For instance, uncertainty in a dataset may arise as a function of sampling, or the way in which data is collected or generated. Data uncertainty may also simply represent the variability of the phenomenon investigated. Some data uncertainty may arise as a result of methodological choices related to "cleaning" the data, such as some methods of dealing with missing data, as well as considerations of visual representations of missing data (Eaton, Plaisant, & Drizd, 2005). An important aspect of understanding uncertainty is being able to identify random or systematic error (Loken & Gelman, 2017), which is complex, as it depends on factors beyond the visible data itself. The idea of being able to discern signal from noise is really just a useful analogy for reasoning, but does not convey the complexity of understanding data, nor the importance of considering research methodology factors in understanding data uncertainty. Data sampling methods and their associated sample size per cell of analysis, the measurement instruments, or the measurement model itself (see Bergner, 2017, for an in-depth discussion of courses of error in application of measurement models), for instance, all contextualize data uncertainty. The possibility of large linked data across systems in learning analytics (Thompson et al., 2013; Zouaq, Jovanović, Joksimović, & Gašević, 2017) means that complexity in dealing with and understanding data uncertainty is more pronounced, and that considerations of methodological choices, including dealing with data uncertainty for inference-making, become even more critical (Lodge, Alhadad, Lewis, & Gašević, 2017).

The benefits of including estimates of uncertainty when reporting statistical results visually and in text, irrespective of Bayesian or frequentist leanings, is undisputed and widely encouraged (Bonneau et al., 2014; Cumming, 2014; Loken & Gelman, 2017; Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). By including visualization of uncertainty when visualizing data, we provide a more accurate depiction of the data, and subsequently, potentially enhance judgement and decision making. Data uncertainty is part and parcel of research — and the degree of uncertainty depends on many factors, including methodological choices, construct and complex relationship investigated, and the alignment of the two. By and large, no measure is completely precise. As such, understanding uncertainty is arguably part of constructing epistemic notions of our way of knowing about the world. Making decisions about the optimal ways to visualize this when the evidence for impact on graph comprehension and inference is less clear will be challenging, but necessary. Our desire for unambiguous conventions to guide decisions for graphing uncertainty is, at the current stage, unlikely to be fulfilled. As with the degree of uncertainty inherent in any dataset, we require a degree of flexibility in understanding, representing, and communicating data uncertainty in ways to support appropriate inference-making. Evaluating uncertainty may also depend on the epistemological stance that a researcher adopts; their knowledge about the research and statistical methodology that contextualizes the data will strongly influence the cognitive lens with which interpretation occurs.

Evidence-based strategies for visual representation of these margins of error to *support cognition* are scant (see Zuk & Carpendale, 2007). Anyone who has tried teaching someone about statistical variability and ways of measuring or quantifying this knows how sticky the concept of error variance really can be. Indeed, vast amount of research suggest that people tend to be probabilistically challenged, whether novice or expert. People have trouble understanding probabilities described in many forms, and tend to adopt heuristics maladaptively in making judgements and inferences (Gottlieb et al., 2007; see Gigerenzer & Brighton, 2009, and Kahneman, 2003 for discussions of potential underlying mechanisms). This effect occurs even for domain-specific types of errors in the field. As in the case of psychology, Belia, Fidler, Williams, and Cumming (2005) show that even experts make judgement errors about error bars in relation to statistical significance, and sometimes even misunderstand different types of error measurements such as confidence intervals. Further, when compared with student novices who had not received any prior education on statistical inference, researchers equally tended to misinterpret confidence intervals (Hoekstra, Morey, Rouder, & Wagenmakers, 2014). This problem is more acute when the findings are to be presented to disciplines other than their own, with different methodological conventions, and to people with various levels of understanding of research, data, and statistics. Further, certain types of graphical representations are less effective in facilitating the understanding of data dispersion or uncertainty, even when one has the knowledge and experience to do so. For example, statistics students appear to consistently misinterpret histograms and box plots despite demonstrating the required knowledge and having the time to interpret them accurately (Lem et al., 2013).

The depiction of data uncertainty through visual representation features such as error bars are intended to help people reason out the distribution of values that the measured variable could take and factor that into their interpretations. While adding visualization of uncertainty helps instill some caution in interpreting findings, these benefits do not appear to universally transfer to understanding the implications of the patterns of error. Making visible uncertainty in principle, should allow people to make judgements about data precision, distribution, quality, or representativeness of the phenomenon investigated.

However, there are still gaps in research that could help us make better-informed decisions about methods for visualizing data uncertainty to support cognition.

One, it is unclear that people use uncertainty visualizations to understand the statistical uncertainty as opposed to methodological or unmeasured uncertainty (the problem of understanding measured versus extraneous, unmeasured uncertainty). Often when dealing with visualizing uncertainty we refer to representations of statistical uncertainty — some of which may arise from various methodological parameters, such as method of sampling, sample size, measurement instruments, repeated measures, experimental and individual variability, etc. Other sources of uncertainty may arise from limitations of the device of measurement or the instrument itself (e.g., Matejka, Glueck, Grossman, & Fitzmaurice, 2016), sometimes called bounded uncertainty (Olston & Mackinlay, 2002). Research in understanding how uncertainty is understood and expressed will help inform strategies for data visualization design (Skeels, Lee, Smith, & Robertson, 2010).

Two, there is currently no consensus on *how* visualizations of uncertainty aid understanding of data in ways that modelling uncertainty is understood in statistics. This second issue is particularly important for us to understand and develop expertise to reason and make inferences appropriately on the basis of data uncertainty. Novices derive little information from, or ignore error estimates in data visualizations (Scown et al., 2014; Tak et al., 2014). When plotted in bar graphs, error bars or confidence intervals are often missed or bias the interpretation of the uncertainty estimation (Newman & Scholl, 2012; Pentoney & Berger, 2016). These perceptual errors have downstream implications for inference making.

Three, in relation to visualization methods for inference making, we need to better understand the conditions under which visualizations of uncertainty are used or disregarded. Despite these gaps in research, evidence suggests that the visibility of estimates of uncertainty, particularly with visualizations, tend to improve some aspects of judgement with data compared to instances when uncertainty is invisible (Nadav-Greenberg & Joslyn, 2009; Burton, Pomeroy, Radenovic, & McCarley, 2017). Arguably, being able to understand uncertainty is critical for reasoning about the ways of knowing and in research for making inferences about complex learning phenomena.

So what does this mean for us in making evidence-informed decisions when constructing data visualizations that are effective and accurate representations of the data, and that represent uncertainty in ways that support (and not hinder) cognition? According to Cleveland (1994), when making decisions about methods of representing uncertainty, emphasis should be placed on the impact on inference making. This recommendation is critical, though difficult to implement and understand. Even with the focus away from statistical significance to that of practical significance, in order to make these judgements, one still needs to understand the basis of and for error, in the context of the research and its methodology. Behaviourally, in the task of assessing visualization of uncertainty, people do not tend to *naturally* use information about uncertainty in making inferences (Kramer, Telfer, & Towler, 2017). This suggests that the design of visualizations should incorporate ways to orient attention to important features of uncertainty and why that might be. Some suggest the presentation of key visual and textual information to aid understanding of important aspects of reasoning and interpretation of data uncertainty (Louis & Zeger, 2009; Morey et al., 2016). Figure 5 is an attempt to illustrate a possible case of visualizing data that prioritizes the consideration of factors for inference-making. The top panel shows a common method of visualizing comparison of two groups across multiple domains using a bar graph, with the representation of data uncertainty with error bars. Based on the discussion above, evidence suggests that this method biases the estimate of uncertainty within-the-bar, and the use of the bars hinders perceptual discrimination for comparison across groups, particularly where there are multiple domains of measurement. The lower panel is an attempt to minimize the extraneous and germane cognitive load by prioritizing the need to: 1) evaluate differences effectively between groups across multiple domains, and 2) judge the practical importance (as opposed to statistical significance) of the differences using confidence intervals of the differences.

All of these considered, it is worth noting that the representation of error margins in between-subject designs for understanding and reasoning about data are comparatively more straightforward relative to those in within-subject designs (Jarmasz & Hollands, 2009; O'Brien & Cousineau, 2014). Thoughtful discussions and recommendations for representing and reporting of data uncertainty in research designs involving repeated measures are covered elsewhere (see Cumming, Fidler, & Vaux, 2007; Fuchs, Pölz, & Bathke, 2017; Nicholls, 2016). Taken together, this section provides evidence-informed *considerations* for visualizing data, and its implications for inference-making. Visualizing data, and particularly data uncertainty, facilitates deeper and more nuanced understanding of making inferences with data. As a field, this is an imperative discipline to apply to learning analytics as we move into increasingly high stakes settings of data- and algorithmic-informed decision making. Omission of this could lead to the fallacy of the average (Aguilar, 2017), proliferation of spurious correlations, and mislead understanding about the nature of the data and of its representativeness. This may influence any inferential inductions about likelihoods of the phenomenon occurring in the population, and critically for educational practice, to inferences made at the individual level ("translation to N=1" issue). While representing data and its uncertainty visually are not inferential panacea, they, at the very least, allow us the opportunity to test and challenge our inferential assumptions.
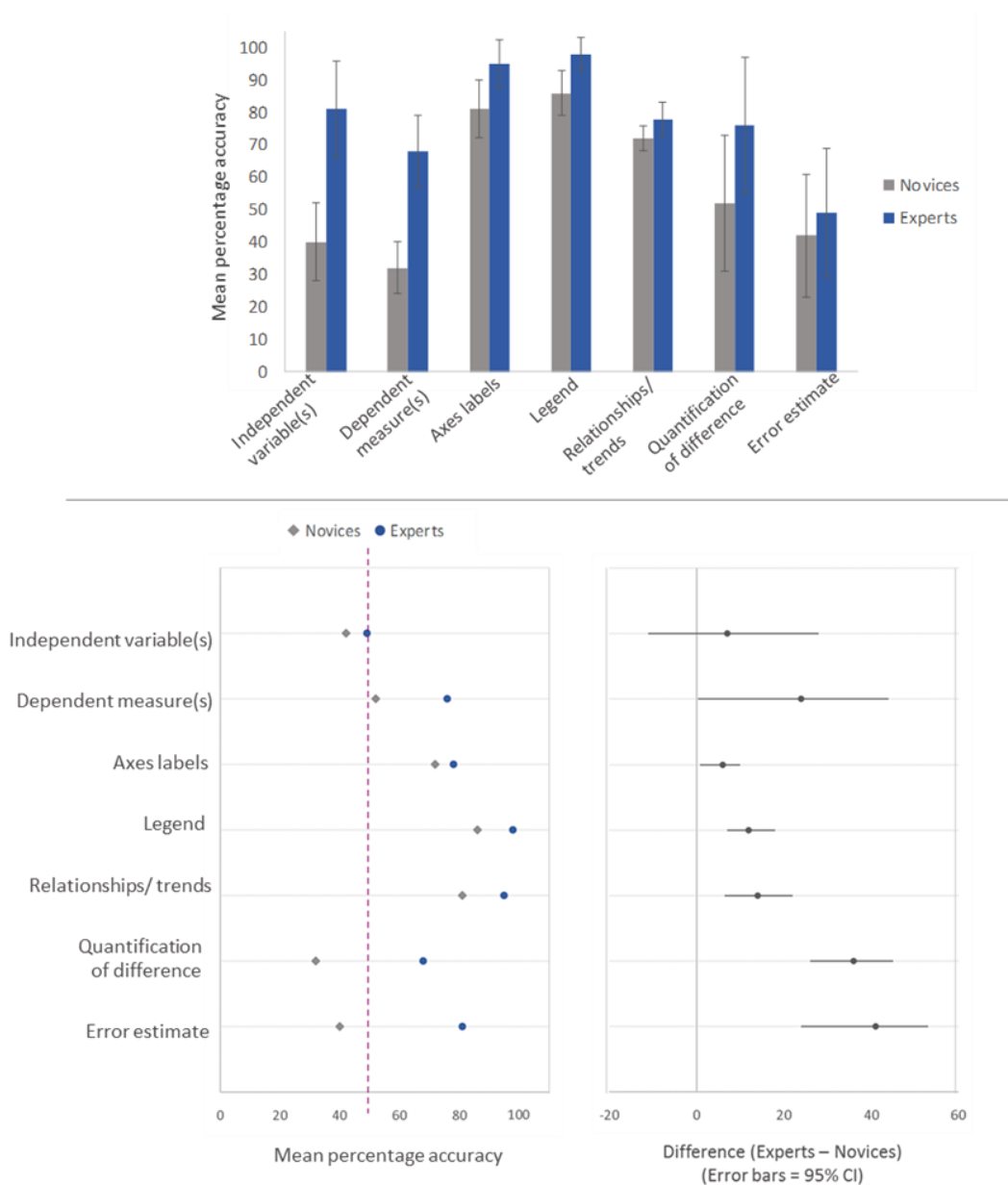
Figure 5. An example of representing uncertainty (only in the context of providing summary data, in this case, mean scores, rather than raw data) in research communication to support cognition. In this fictitious case, the percentage accuracy in explaining the graphical or statistical concept is visualized in two ways. The top panel (bar graph) is less effective as an assessment of differences between groups and of how the associated data uncertainty is perceptually hindered and biased. The bottom panel illustrates an alternative method using a dot plot (bottom left panel), with an indicator of performance accuracy at a chance level reference point (indicated by the dotted pink line). This corresponds with (bottom right panel) the 95% confidence intervals of the differences (as indicated by the thin line) in the same visual field for coherent assessment of differences and practice significance.

Table 2. Summary of evidence-informed guidelines for visualizing data as methodology to support complex cognition in understanding and communicating research

Research evidence	Practical recommendation
Visualize data as an inferential tool to reveal important insights about underlying data properties	
1. Visual representations can support cognition (Hegarty, 2011)	• Visualize data as a key means of understanding and communicating information
2. Visualizing data facilitates identification of patterns that are otherwise obscured in text or tables (Gelman, Pasarica, & Dodhia, 2002; Kastellec & Leoni, 2007)	• Before computing any statistics, visualize your data to evaluate assumptions of underlying univariate or multivariate data, and to detect any serious compromises to data integrity
3. Particularly for complex, multivariate relationships, people are able to detect statistical effects more accurately and more quickly with visualizations than with tables (Meyer, Shamo, & Gopher, 1999; Porat, Oron-Gilad, & Meyer, 2009)	• For understanding research, visualize the data in various/ multiple methods or representations to be able to better evaluate the ways in which underlying properties may be concealed by a single method • For communicating research, consider evidence-informed principles to make methodological choices when constructing data visualizations to support attention and cognition (see Table 1)
4. Some types of visualizations inherently obscure the representation of the data properties (including uncertainty, see points 5–7; Anscombe, 1973; Krzywinski & Altman, 2014; Matejka & Fitzmaurice, 2017a)	• Explore multiple ways of visualizing the data to make critical judgements about how the visualization decisions may support or potentially lead to errors in interpretation and subsequent inference
Visualize uncertainty to support inference-making	
5. Visualizing uncertainty is a useful aid for researchers to make statistical and inferential decisions (Kramer, Telfer, & Towler, 2017; Burton, Pomeroy, Radenovic, & McCarley, 2017)	• Visualize data in ways that enable you to see and understand the nature of the data dispersion and to test assumptions of the data
6. Visualizing uncertainty appears to aid readers' understanding of the criticality of data uncertainty than without its visualization; BUT people do not appear to naturally use uncertainty information to aid inference-making (Hoekstra et al., 2014; Kramer et al., 2017; Newman & Scholl, 2012; Morey et al., 2016)	• Use text to orient attention important aspects of uncertainty visualized, and why they matter
7. Understanding data uncertainty is difficult for both novices and experts (Newman & Scholl, 2012; Pentoney & Berger, 2016; Scown, Bartlett, & McCarley, 2014; Tak, Toet, & van Erp, 2014)	• Emphasize the impact of data uncertainty in communicating the implications for research to practice translation; use language to scaffold data uncertainty to support visualization of uncertainty, or use language to explain the implication of uncertainty in your dataset on inference-making explicitly

5. Next Steps: Towards Human-Centred Learning Analytics Research and Practice

5.1 Towards a human-centred dashboard (and other data visualizations) design

The scope of this paper was limited to that of static data visualizations. Nonetheless, some of the empirical research and theoretical principles covered here are generalizable to dynamic or interactive visualizations, to an extent. The situatedness of processing data visualizations does include the mechanism through which the data is visualized and presented to users or consumers. The specific ways in which this research evidence might be applied or translated could be an avenue for future work. Further, learning analytics as a field provides the unique opportunity to empirically investigate the human information processing mechanisms of data visualizations in the various contexts of educational practice, and of human factors, generally.

Provision or presentations of data to various stakeholder groups in learning analytics is a core part of the field's aim to optimize understanding and improvement of wicked educational problems. As such, understanding the effects of design features of these data visualizations in the various nuanced educational practice contexts on the intended cognitive (and socioemotional) processes would benefit the field as a whole.

For instance, where learning analytics dashboards are concerned, it has been argued that their design (or any learning analytics data visualizations provided to stakeholders in any other forms) should be linked to learning theory and learning sciences (e.g., Kelly, Thompson, & Yeoman, 2015; Sedrakyan, Malmberg, Verbert, Järvelä, & Kirschner, 2018). In understanding the adoption, use, and integration of learning analytics visualizations in educational practice (whether for learners, educators, or policymakers), the questions often missed, but worth asking are whether the data visualizations, at the first level, are designed in ways that are 1) supportive (not limiting) of attention and cognitive processes imperative for judgement and decision making with the data; 2) aligned to the methodology with which the data is collected; 3) scaffolded where necessary with important contextual and complementary information to support meaning-making? There is, of course, a lot of good work in the literature in elucidating our understanding of how contextualized data provision may support educational practice across various stakeholders (e.g., van Leeuwen, 2015; Jivet, Scheffel, Specht, & Drachsler, 2018). More recently, Sedrakyan and colleagues (2018) suggest practical ways that learning analytics can be linked with learning science concepts when developing dashboards to support learning regulation. The importance of situating the visualized data as aligned to the educational purpose cannot be understated. However, for these efforts to optimize the likelihood of supporting the intended cognitive and learning processes, these efforts should also, at minimum, be designed in ways that support basic attention and cognitive processes as described in this paper. When designing data visualizations for learning analytics research or practice, other practical guidelines in the literature would be worth being synthesized with the present paper in guiding practice. See Klerx, Verbet, and Duval (2017) for practical guidelines on how to get started on developing data visualizations for this purpose; and Hillaire, Rappolt-Schlichtmann, and Ducharme (2016) for prototyping guidelines.

The study of designing for representations of data and scenario uncertainty to support cognition is one that would benefit from more research, both in terms of understanding the differential features of visualizing uncertainty in supporting or hindering attention and cognition, as well as understanding how narration or data storytelling might support this process. Specifically, narrative may serve to communicate not only data uncertainty, but also scenario uncertainty, hence providing a strategy to support understanding of uncertainty in data visualizations. Understanding whether these strategies in designing data visualizations support building tolerance for uncertainty in making meaning, judgements, and inferences would be of empirical and theoretical interest. This would build on emerging research (at least in learning analytics; Echeverria et al., 2018; and in human–computer interaction, e.g., Hullman & Diakopoulos, 2011) in deepening our understanding of this necessary, but inherently complex process.

Further, this paper did not address the complexity of the social, cultural, and emotional aspects of processing data visualizations. These processes are inextricably linked, and are important to consider when attempting to understand human factors in processing data visualization for insight. While these processes are critical in the sensemaking process, there is a relative dearth in the research focusing on the roles these factors play (Kennedy & Hill, 2017). Further, in learning analytics, this may be beyond visualizing just data. These mechanisms are inherent in our capacity to construct visual representations of complex phenomena or systems (e.g., learning design representations, or patterns; see Muñoz-Cristóbal and colleagues, 2018; and systems dynamics visual representations; see Howard & Thompson, 2016). To support capacity limits of the human mind in ways that leverage research evidence on human information processing and complex (higher-order) cognition and emotion, these are domains that learning analytics researchers are well placed to gain a deeper understanding in contributing to the wider field and bridging across disciplines.

5.2 Towards holistic understanding of "data literacy" in learning analytics

Often, *data literacy* is mentioned as a key component of being able to integrate data into one's educational practice (e.g., Alhadad, Thompson, Knight, Lewis, & Lodge, 2018; Kitto, Buckingham Shum, & Gibson, 2018; Wolff, Gooch, Cavero Montaner, Rashid, & Kortuem, 2016). The concept of data literacy is contextual, as is data- or evidence-informed practice. While this paper has reviewed some studies that investigated human processing capacity limitations as well as individual differences in comprehending data visualizations, this review was not intended to provide a comprehensive view of this related research. Nonetheless, it is fair to conclude that the ability to understand various graphical formats has a clear impact on meaning making and subsequent inferences. Further research could focus on elucidating the various aspects of data literacy that may be necessary in *complex cognition* (Knauff & Wolf, 2010) that one would expect when integrating learning analytics into practice. Gibson and Martinez-Maldonado (2017) developed a nice conceptual model toward understanding embodied cognition in the meaning making process — the challenge is to synthesize this with other aspects of complex cognition and

emotion when dealing with data in learning analytics practice. Gaining a deeper understanding of how graphical literacy might be linked to other aspects of understanding data within the umbrella of *data literacy* will not only benefit the field in understanding this for learning analytics design and implementation, but also for scaffolding, professional learning, and, where teams are concerned, understanding the complexity in inter- and transdisciplinary efforts where data visualizations are concerned.

6. Conclusions

In this paper, human capacity limitations in the context of cognitive processing of data visualization for understanding research and making inferences are discussed as supported by research evidence. The bottom-up and top-down factors identified provide the basis for making evidence-informed visualization design decisions. The unique importance of decisions about methods of visualization (visual features, type of visualizations) in incorporating key information is particularly important for multifaceted, multivariate models in learning analytics research. This paper discusses some of the key elements of human cognition to support evidence-informed decision making about visualizing data. Of course, other factors that have not been discussed here are also critical for understanding the influence of human perception and cognition in making these decisions, and for understanding the impact of these decisions on inference-making. As part of the imperative discipline to progress learning analytics research and practice in ethical and evidence-informed ways, the next step to enhance our intellectual journey here may be to better understand how people apply the research evidence to educational practice, as dependent on methodological properties such as visualizing data as information. A vast literature focusing on heuristics and cognitive biases influences how people make sense of information, which is an important part of this puzzle. This would constitute a positive change in how learning data can be graphically depicted in research and practice (e.g., through learning analytics dashboards), as it helps us better inform our designs for less biased, more accurate inference making, and would inform the development of data visualization tools to better support human cognition.

References

Aguilar, S. J. (2017). Learning analytics: At the nexus of big data, digital innovation, and social justice in education. *TechTrends, 62*(1), 37–45. http://dx.doi.org/10.1007/s11528-017-0226-9

Alhadad, S. S. J. (2016). Attentional and cognitive processing of analytics visualizations: Can design features affect interpretations and decisions about learning and teaching? In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning: Proceedings of the 33rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (ASCILITE 2016), 28–30 November 2016, Wellington, New Zealand (pp. 20–32). Adelaide: Australasian Society for Computers in Learning in Tertiary Education. http://2016conference.ascilite.org/wp-content/uploads/ascilite2016_alhadad_full.pdf

Alhadad, S. S. J., Thompson, K., Knight, S., Lewis, M., & Lodge, J. M. (2018). Analytics-enabled teaching as design: Reconceptualisation and call for research. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, NSW, Australia (pp. 427–435). New York: ACM. http://dx.doi.org/10.1145/3170358.3170390

Ali, N., & Peebles, D. (2013). The effect of Gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. *Human Factors, 55*(1), 183–203. http://dx.doi.org/10.1177/0018720812452592

Anderson, B. A. (2016). The attention habit: How reward learning shapes attentional selection. *Annals of the New York Academy of Sciences, 1369,* 24–39. http://dx.doi.org/10.1111/nyas.12957

Anderson, B. A., & Yantis. S. (2013). Persistence of value-driven attentional capture. *Journal of Experimental Psychology:*

Human Perception & Performance, 39, 6–9.

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician, 27*(1), 17–21.

Ansorge, U., Horstmann, G., & Scharlau, I. (2010). Top-down contingent attentional capture during feed-forward visual processing. *Acta Psychologica, 135,* 123–126.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Science, 16*(8), 437–443. http://dx.doi.org/10.1016/j.tics.2012.06.010

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 206–226). New York: Cambridge University Press.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839. http://dx.doi.org/0.1038/nrn120

Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual clutter causes high-magnitude errors. *PLOS: Biology, 4*(3), 387–394. http://dx.doi.org/10.1371/journal.pbio.0040056

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. http://dx.doi.org/10.1093/scan/nsw154

Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The role of relational information in contingent capture. *Journal of Experimental Psychology: Human Perception and Performance, 36*(6), 1460–1476. http://dx.doi.org/10.1037/a0020370

Becker, S. I., Lewis, A. J., & Axtens, J. E. (2017). Top-down knowledge modulates onset capture in a feedforward manner. *Psychological Bulletin & Review, 24,* 436–446. http://dx.doi.org/10.3758/s13423-016-1134-2

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods, 10*(4), 389–396. http://dx.doi.org/10.1037/1082-989X.10.4.389

Bergner, Y. (2017). Measurement and its uses in learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *The handbook of learning analytics* (pp. 34–48). Alberta, Canada: Society for Learning Analytics Research (SoLAR). http://dx.doi.org/10.18608/hla17.003

Bonneau, G. P., Hege, H., Johnson, C. R., Oliveira, M. M., Potter, K., Rhenigans, P., & Schultz, T. (2014). Overview and state-of-the-art of uncertainty visualization. In C. Hansen, M. Chen, C. Johnson, A. Kaufman, & H. Hagen (Eds.), *Scientific visualization: Mathematics and visualization* (pp. 3–27). London: Springer. http://dx.doi.org/10.1007/978-1-4471-6497-5_1

Braithwaite, D. W., & Goldstone, R. L. (2013). Flexibility in data interpretation: Effects of representational format. *Frontiers in Psychology, 4*, 980. http://dx.doi.org/10.3389/fpsyg.2013.00980

Burton, O., Pomeroy, D., Radenovic, V., & McCarley, J. S. (2017). Visualization of uncertainty aids spatial judgements but fails to improve metacognitive efficiency. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting, 61*(1), 1390–1393. http://dx.doi.org/10.1177/1541931213601831

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75–100. http://dx.doi.org/10.1037/1076-898X.4.2.75

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 225*(2), 315–324. http://dx.doi.org/10.1016/j.chb.2008.12.020

Cleveland, W. S. (1994). *The Elements of Graphing Data*. Murray Hill, NJ: Hobart Press.

Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science, 216*(4550), 1138–1141. http://dx.doi.org/10.1126/science.216.4550.1138

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 6,* 683–695. http://dx.doi.org/10.1080/13562517.2013.827653

Conati, C., & Maclaren, H. (2008). Exploring the role of individual differences in information visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces* (AVI '08), 28–30 May 2008, Napoli, Italy (pp. 199–206). New York: ACM. http://dx.doi.org/10.1145/1385569.1385602

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology, 14*(19), R850–R852. http://dx.doi.org/10.1016/j.cub.2004.09.041.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29. http://dx.doi.org/10.1177/0956797613504966

Cumming, G., Fidler, F., & Vaux, D. L. (2007). Error bars in experimental biology. *The Journal of Cell Biology, 177*(1), 7–11. http://dx.doi.org/10.1083/jcb.200611141

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222. http://dx.doi.org/10.1146/annurev.ne.18.030195.001205

Eaton, C., Plaisant, C., & Drizd, T. (2005). Visualizing missing data: Graph interpretation user study. In M. F. Costabile & F. Paternò (Eds.), *Human–Computer Interaction: INTERACT 2005. Lecture Notes in Computer Science*, *3585* (pp. 861–872). Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/11555261_68

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Buckingham Shum, S. (2018). Driving data storytelling from learning design. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, NSW, Australia (pp. 131–140). New York: ACM. http://dx.doi.org/10.1145/3170358.3170380

Elmqvist, N., & Fekete, J. (2010). Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization & Computer Graphics, 16*(3), 1077–2626. http://dx.doi.org/10.1109/TVCG.2009.84

Fisher, D., Popov, I., Drucker, S. M., & Schraefel, M. (2012). Trust me, I'm partially right: Incremental visualization lets analysts explore large datasets faster. *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '12), 5–10 May 2012, Austin, TX, USA (pp. 1673–1682). New York: ACM.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance, 18*(4), 1030–1044. http://dx.doi.org/10.1037/0096-1523.18.4.1030

Fuchs, N., Pölz, W., & Bathke, A. C. (2017). Confidence intervals for population means of partially paired observations. *Statistical Papers, 58*(1), 35–51. http://dx.doi.org/10.1007/s00362-015-0686-y

Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the learning analytics puzzle: A consolidated model of a field of research and practice. *Learning: Research & Practice, 17*(1), 63–78. http://dx.doi.org/10.1080/23735082.2017.1286142

Gelman, A., & Hennig, C. (2017). Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Statistics in Society Series A, 180,* 1–31. http://dx.doi.org/10.1111/rssa.12276

Gelman, A., Pasarica, C., & Dodhia, R. (2002). Let's practice what we preach: Turning tables into graphs. *The American Statistician, 56*(2), 121–130. http://dx.doi.org/10.1198/000313002317572790

Gibson, A., & Martinez-Maldonado, R. (2017). That dashboard looks nice, but what does it mean? Towards making meaning explicit in learning analytics design. *Proceedings of the 29th Australian Computer–Human Interaction Conference* (OzCHI 2017) 28 November–1 December 2017, Brisbane, QLD, Australia (pp. 528–532). http://dx.doi.org/10.1145/3152771.3156171

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1,* 107–143. http://dx.doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *British Medical Journal, 327*(7417), 741–744. http://dx.doi.org/10.1136/bmj.327.7417.741

Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology. 19,* 183–204. http://dx.doi.org/10.1002/acp.1110

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236–243. http://dx.doi.org/10.1016/S1364-6613(00)01662-4

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*(1), 75–90. http://dx.doi.org/10.1037/0033-295X.109.1.75

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18*(3), 240–246. http://dx.doi.org/10.1111/j.1467-9280.2007.01883.x

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*, 446–474. http://dx.doi.org/10.1111/j.1756-8765.2011.01150.x

Hegarty, M., Smallman, H. S., Stull, A. T., & Canham, M. (2009). Naïve cartography: How intuitions about display configuration can hurt performance. *Cartographica, 44,* 171–186. http://dx.doi.org/10.3138/carto.44.3.171

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin, 21*(5), 1157–1164. http://dx.doi.org/10.3758/s13423-013-0572-3

Howard, S. K., & Thompson, K. (2016). Seeing the system: Dynamics and complexity of technology integration in secondary schools. *Education & Information Technologies, 21*(6), 1877–1894. http://dx.doi.org/10.1007/s10639-015-9424-2

Hillaire, G., Rappolt-Schlichtmann, G., & Ducharme, K. (2016). Prototyping visual learning analytics guided by an educational theory informed goal. *Journal of Learning Analytics, 3*(3), 115–142. http://dx.doi.org/10.18608/jla.2016.33.7

Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2231–2240. http://dx.doi.org/10.1109/TVCG.2011.255

Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology, 63*(2), 124–138. http://dx.doi.org/10.1037/a0014164

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for

educational practice. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, NSW, Australia (pp. 32–40). New York: ACM. http://dx.doi.org/10.1145/3170358.3170421

Josef, A. K., Richter, D., Samanez-Larkin, G. R., Wagner, G. G., Hertwig, R., & Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology, 111*(3), 430–450. http://dx.doi.org/10.1037/pspp0000090

Kadel, H., Feldmann-Wüstefeld, T., & Schubö, A. (2017). Selection history alters attentional filter settings persistently beyond top-down control. *Psychophysiology, 54*(5), 736–754. http://dx.doi.org/10.1111/psyp.12830

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review, 93*(5), 1449–1475. http://dx.doi.org/10.1257/000282803322655392

Kastellec, J. P., & Leoni, E. L. (2007). Using graphs instead of tables in political science. *Perspectives on Politics, 5*(4), 755–771. http://dx.doi.org/10.1017/S1537592707072209

Kelly, N., Thompson, K., & Yeoman, P. (2015). Theory-led design of instruments and representations in learning analytics: Developing a novel tool for orchestration of online collaborative learning. *Journal of Learning Analytics, 2*(2), 14–43. http://dx.doi.org/10.18608/jla.2015.22.3

Kennedy, H., & Hill, R. L. (2017). The feeling of numbers: Emotions in everyday engagements with data and their visualization. *Sociology.* Advanced online publication. http://dx.doi.org/10.1177/0038038516674675

Kirschner, P. A., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory research: The good, the bad, and the ugly. *Computers in Human Behavior, 27*(1), 99–105. http://dx.doi.org/10.1016/j.chb.2010.06.025

Kitto, K., Buckingham Shum, S., & Gibson, A. (2018). Embracing imperfection in learning analytics. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, NSW, Australia (pp. 451–460). New York: ACM. http://dx.doi.org/10.1145/3170358.3170413

Klerx, J., Verbert, K., & Duval, E. (2017). Learning analytics dashboards. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *The handbook of learning analytics* (pp. 143–150). Alberta, Canada: Society for Learning Analytics Research (SoLAR). http://dx.doi.org/10.18608/hla17.012

Knauff, M., & Wolf, A. G. (2010). Complex cognition: The science of human reasoning, problem-solving, and decision-making. *Cognitive Processing, 11*(2), 99–102. http://dx.doi.org/10.1007/s10339-010-0362-z

Kosslyn, S. M. (2006). *Graph design for the eye and mind.* New York: Oxford University Press.

Kramer, R. S. S., Telfer, C. G. R., & Towler, A. (2017). Visual comparison of two data sets: Do people use the means and the variability? *Journal of Numerical Cognition, 3*(1), 97–111. http://dx.doi.org/10.5964/jnc.v3i1.100

Krämer, W., & Gigerenzer, G. (2005). How to confuse with statistics or: The use and misuse of conditional probabilities. *Statistical Science, 20*(3), 223–230. http://dx.doi.org/10.1214/088342305000000296

Krzywinski, M., & Altman, N. (2014). Points of significance: Visualizing samples with box plots. *Nature Methods, 11*. http://dx.doi.org/119-120. 1038/nmeth.2813

Kubina, R. M., Jr., Kostewicz, D. E., Brennan, K. M., & King, S. A. (2017). A critical review of line graphs in behavior analytics journals. *Educational Psychology Review, 29,* 583–598. http://dx.doi.org/10.1007/s10648-015-9339-x

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). On the misinterpretation of histograms and box plots. *Educational Psychology, 33*(2), 155–174. http://dx.doi.org/10.1080/01443410.2012.674006

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2014). Interpreting histograms. As easy as it seems? *European Journal of Psychology of Education, 29*(4), 557–575. http://dx.doi.org/10.1007/s10212-014-0213-x

Lodge, J. M., Alhadad, S. S. J., Lewis, M. J., & Gašević, D. (2017). Inferring learning from big data: The importance of a transdisciplinary and multidimensional approach. *Technology, Knowledge & Learning, 22*(3), 385–400. http://dx.doi.org/10.1007/s10758-017-9330-3

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science, 355*(6325), 584–585. http://dx.doi.org/10.1126/science.aal3618

Louis, T. A., & Zeger, S. L. (2009). Effective communication of standard errors and confidence intervals. *Biostatistics, 10*(1), 1–2. http://dx.doi.org/10.1093/biostatistics/kxn014

Lund, K., Quignard, M., & Shaffer, D. W. (2017). Gaining insight by transforming between temporal representations of human interaction. *Journal of Learning Analytics, 4*(3), 102–122. http://dx.doi.org/10.18608/jla.2017.43.6

Matejka, J., & Fitzmaurice, G. (2017a). Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '17), 6–11 May 2017, Denver, Colorado, USA (pp. 1290–1294). New York: ACM. http://dx.doi.org/10.1145/3025453.3025912

Matejka, J., & Fitzmaurice, G. (2017b). The datasaurus dozen — same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. https://www.autodeskresearch.com/publications/samestats

Matejka, J., Glueck, M., Grossman, T., & Fitzmaurice, G. (2016). The effect of visual appearance on the performance of

continuous sliders and visual analogue scales. *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '16), 7–12 May 2016, San Jose, CA, USA (pp. 5421–5432). New York: ACM. http://dx.doi.org/10.1145/2858036.2858063

Matzen, L. E., Haass, M. J., Divis, K. M., & Stites, M. C. (2017). Patterns of attention: How data visualizations are read. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10284,* 176–191. http://dx.doi.org/10.1007/978-3-319-58628-1_15

Mautone, P. D., & Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology, 99*(3), 640–652. http://dx.doi.org/10.1037/0022-0663.99.3.640

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.015

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory.* New York: Elsevier.

Meyer, J., Shamo, M. K., & Gopher, D. (1999). Information structure and the relative efficacy of tables and graphs. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41,* 570–587. http://dx.doi.org/10.1518/001872099779656707

Michal, A. L., & Franconeri, S. L. (2017). Visual routines are associated with specific graph interpretations. *Cognitive Research: Principles & Implications, 2*(20). http://dx.doi.org/10.1186/s41235-017-0059-2

Moore, P., & Fitz, C. (1993). Gestalt theory and instructional design. *Journal of Technical Writing & Communication, 23*(2), 137–157. http://dx.doi.org/10.2190/G748-BY68-L83T-X025

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. (2016). The fallacy of placing confidence in confidence intervals. *Psychological Bulletin Review, 23*(1), 103–123. http://dx.doi.org/10.3758/s13423-015-0947-8

Muñoz-Cristóbal, J., Hernández-Leo, D., Carvalho, L., Martinez-Maldonado, R., Thompson, K., Wardak, D., & Goodyear, P. (2018). 4fFAD: A framework for mapping the evolution of artefacts in the learning design process. *Australasian Journal of Educational Technology, 34*(2), 16–34. http://dx.doi.org/10.14742/ajet.3706

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering & Decision Making, 3*(3), 209–227. http://dx.doi.org/10.1518/155534309X474460

Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review, 19*(4), 601–607. http://dx.doi.org/10.3758/s13423-012-0247-5

Nicholls, A. (2016). Confidence limits, error bars and method comparison in molecular modeling. Part 2: Comparing methods. *Journal of Computer-Aided Molecular Design, 30*(2), 103–126. http://dx.doi.org/10.1007/s10822-016-9904-5

Nothelfer, C., Gleicher, M., & Franconeri, S. (2017). Redundant encoding strengthens segmentation and grouping in visual displays of data. *Journal of Experimental Psychology: Human Perception & Performance, 43*(9), 1667–1676. http://dx.doi.org/10.1037/xhp0000314

O'Brien, F., & Cousineau, D. (2014). Representing error bars in within-subjects designs in typical software packages. *The Quantitative Methods for Psychology, 10*(1), 56–67. http://dx.doi.org/10.20982/tqmp.10.1.p056

Okan, Y., Galesic, M., Garcia-Retamero, R. (2016). How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *Journal of Behavioral Decision Making, 29,* 271–294. http://dx.doi.org/10.1002/bdm.1891

Olston, C., & Mackinlay, J. D. (2002). Visualizing data with bounded uncertainty. *Proceedings of the IEEE Symposium on Information Visualization* (INFOVIS 2002), 28–29 October 2002, Boston, MA, USA (pp. 37–40). Los Alamitos, CA: IEEE Computer Society. http://dx.doi.org/10.1109/INFVIS.2002.1173145

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 27–42). New York: Cambridge University Press.

Pastore, M., Lionetti, F., & Altoe, G. (2017). When one shape does not fit all: A commentary essay on the use of graphs in psychological research. *Frontiers in Psychology, 8*(1666). http://dx.doi.org/10.3389/fpsyg.2017.01666

Peebles, D., & Ali, N. (2015). Expert interpretation of bar and line graphs: The role of graphicacy in reducing the effect of graph format. *Frontiers in Psychology, 6,* 1673. http://dx.doi.org/10.3389/fpsyg.2015.01673

Pentoney, C. S., & Berger, D. E. (2016). Confidence intervals and within-the-bar bias. *The American Statistician, 70*(2), 215–220. http://dx.doi.org/10.1080/00031305.2016.1141706

Pinker, S. (1990). A theory of graph comprehension. I. L. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hiltdale, NJ: Lawrence Erlbaum Associates.

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A., & Scholte, S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision, 13*(3), 1–14. http://dx.doi.org/10.1167/13.3.16

Porat, T., Oron-Gilad, T., & Meyer, J. (2009). Task-dependent processing of tables and graphs. *Behaviour & Information Technology, 28*(3), 293–307. http://dx.doi.org/10.1080/01449290701803516

Rau, M. A. (2017). How do students learn to see concepts in visualizations? Social learning mechanisms with physical and

virtual representations. *Journal of Learning Analytics, 4*(2), 240–263. http://dx.doi.org/10.18608/jla.2017.42.16

Riel, J., Lawless, K. A., & Brown, S. W. (2018). Timing matters: Approaches for measuring and visualizing behaviours of timing and spacing of work in self-paced online teacher professional development courses. *Journal of Learning Analytics, 5*(1), 25–40. http://dx.doi.org/10.18608/jla.2018.51.3

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 17–22. http://dx.doi.org/10.1167/7.2.17

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human–Computer Studies, 45,* 185–213. http://dx.doi.org/10.1006/ijhc.1996.0048

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signalling affects learning with media. *Educational Research Review, 23,* 1–24. http://dx.doi.org/10.1016/j.edurev.2017.11.001

Scown, H., Bartlett, M., & McCarley, J. S. (2014). Statistically lay decision makers ignore error bars in two-point comparisons. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting, 58*(1), 1746–1750. http://dx.doi.org/10.1177/1541931214581364

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2018). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior.* Advance online publication. http://dx.doi.org/10.1016/j.chb.2018.05.004

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9–45. http://dx.doi.org/10.18608/jla.2016.33.3

Shah, P. (1997). A model of the cognitive and perceptual processes in graphical display comprehension. In M. Anderson (Ed.), *Reasoning with diagrammatic representations* (pp. 94–101). Menlo Park, CA: AAAI Press.

Shah, P., & Freedman, E. G. (2009). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science, 3*(3), 1–19. http://dx.doi.org/10.1111/j.1756-8765.2009.01066.x

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review, 14*(1), 47–69. http://dx.doi.org/10.1023/A:1013180410169

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91*(4), 690–702. http://dx.doi.org/10.1037/0022-0663.91.4.690

Sher, V., Bemis, K. G., Liccardi, I., & Chen, M. (2017). An empirical study on the reliability of perceiving correlation indices using scatterplots. *Computer Graphics Forum, 36,* 61–72. http://dx.doi.org/10.1111/cgf.13168

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 3–6 September 1996, Boulder, CO, USA (pp. 336–343). http://dx.doi.org/10.1109/VL.1996.545307

Skeels, M., Lee, B., Smith, G., & Robertson, G. G. (2010). Revealing uncertainty for information visualization. *Information Visualization, 9,* 70–81. http://dx.doi.org/10.1057/ivs.2009.1

Smallman, H. S., & St. John, M. (2005). Naïve realism: Misplaced faith in realistic displays. *Ergonomics in Design, 13,* 14–19.

Stofer, K., & Che, X. (2014). Comparing experts and novices on scaffolded data visualizations using eye-tracking. *Journal of Eye Movement Research, 7*(5), 1–15. http://dx.doi.org/10.16910/jemr.7.5.2

Strahan, R. F., & Hansen, C. J. (1978). Underestimating correlation from scatterplots. *Applied Psychological Measurement, 2,* 543–550. http://dx.doi.org/10.1177/106480460501300303

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. http://dx.doi.org/10.1023/A:1022193728205

Tak, S., Toet, A., & van Erp, J. (2014). The perception of visual uncertainty representation by non-experts. *IEEE Transactions on Visualization & Computer Graphics, 20.* http://dx.doi.org/935-943. 10.1109/TVCG.2013.247

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77–99. http://dx.doi.org/10.1016/j.actpsy.2010.02.006

Thompson, K., Ashe, D., Carvalho, L., Goodyear, P., Kelly, N., & Parisio, M. (2013). Processing and visualizing data in complex learning environments. *American Behavioral Scientist, 57*(10), 1401–1420. http://dx.doi.org/10.1177/0002764213479368

Toker, D., Conati, C., Steichen, B., & Carenini, G. (2013). Individual user characteristics and information visualization: Connecting the dots through eye tracking. *Proceedings of the Conference on Human Factors in Computing Systems* (CHI '13), 27 April–2 May 2013, Paris, France (pp. 295–304). New York: ACM. http://dx.doi.org/10.1145/2470654.2470696

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Tversky, B. (2005). Functional significance of visuospatial representations. In P. Shah & A. Miyake (Eds.), *Handbook of*

higher-level visuospatial thinking (pp. 1–70). Cambridge, UK: Cambridge University Press.

van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of Learning Analytics*, *2*(2), 138–162. http://dx.doi.org/10.18608/jla.2015.22.11

van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 746–759. http://dx.doi.org/10.1037/0096-1523.30.4.749

van Zoest, W., Van der Stigchel, S., & Donk. M. (2017). Conditional control in visual selection. *Attention, Perception, & Psychophysics, 79,* 1555–1572. http://dx.doi.org/10.3758/s13414-017-1352-3

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., et al. (2014). Learning dashboards: An overview and future research opportunities *Personal & Ubiquitous Computing, 18*(6), 1499–1514. http://dx.doi.org/10.1007/s00779-013-0751-2

Victor, B. (2011). Scientific communication as sequential art. http://worrydream.com/ScientificCommunicationAsSequentialArt/

Ward, M. O., Grinstein, G., & Keim, D. (2015). *Interactive data visualization: Foundations, techniques, and applications*, 2nd ed. Boca Raton, FL: CRC Press.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594–604. http://dx.doi.org/10.1037/0003-066X.54.8.594

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2)*,* 202–238. http://dx.doi.org/10.3758/BF03200774

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *Journal of Community Informatics, 12*(3), 9–26. http://ci-journal.org/index.php/ciej/article/view/1286

Woodward, J. F. (2011). Data and phenomena: A restatement and defense. *Synthese, 182*(1), 165–179. http://dx.doi.org/10.1007/s11229-009-9618-5

Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human, Perception, & Performance, 16*(1), 135–149. http://dx.doi.org/10.1037/0096-1523.16.1.135

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition & Emotion, 24*(1), 3–47. http://dx.doi.org/10.1080/02699930903205698

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition, 27*(6), 1073–1079. http://dx.doi.org/10.3758/BF0320123

Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied, 4*(2), 119–138. http://dx.doi.org/10.1037/1076-898X.4.2.119

Zouaq, A., Jovanović, J., Joksimović, S., & Gašević, D. (2017). Linked data for learning analytics: Potentials and challenges. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *The handbook of learning analytics* (pp. 347–355). Alberta, Canada: Society for Learning Analytics Research (SoLAR). http://dx.doi.org/10.18608/hla17.030

Zuk, T., & Carpendale, S. (2007). Visualization of uncertainty and reasoning. In A. Butz, B. Fisher, A. Krüger, P. Olivier, & S. Owada (Eds.), S*mart Graphics: Lecture Notes in Computer Science, 456*9 (pp. 164–177). Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-540-73214-3_15
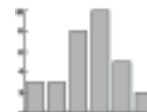
Appendix A

GUIDELINES FOR VISUALIZING DATA

Designing data visualizations for attention: Visual attention is limited and selective

1. SALIENCE

Salience of visual features influences attentional capture[5,10]

Ensure the visual features are designed in ways that do not detract from, but rather support, the understanding of information.

2. PRIOR EXPERIENCE & KNOWLEDGE

A person's **prior experience and knowledge** can guide visual attention[1,4,5,10]

Seek to understand the conventions of data visualizations in a particular research and practice domain; choose visualization design strategies to complement familiarity of the knowledge domain (see also recommendation 5). Enhance this by scaffolding the narrative to support the understanding of the data visualization (see also point 7).

3. CHUNKING

Chunking improves processing of complex information through minimising cognitive load and working memory capacity limitations [2,6]

Apply visual design principles to group visualization features to foster generative processing or promote chunking (e.g., using colour or shape to cluster by related meaning [see also point 5]; or visualize complex relationships in small multiples.

4. AVOID VISUAL CLUTTER

Avoid visual clutter as it can severely degrade attention and impair comprehension[3,9]

Exclude irrelevant materials such as excessive details, visual embellishments, unnecessary information, or cluttered, dense layout; this enables the reader to allocate more cognitive capacity for essential processing (coherence principle).

Designing data visualizations for cognition: Use text with visualization to support cognition

5. PRIOR EXPERIENCE & KNOWLEDGE

Prior experience and knowledge can guide visual attention[1,4,5,10]

Organize information in ways that are aligned to conventional mental models to provide retrieval cues for knowledge structure to aid inference-making.

6. SPATIAL DISTANCE

When text and visualizations are **spatially distant**, attention is divided, and people process these less deeply[8]

Keep visualizations and supporting text close together; this helps the reader build important connections between corresponding words and visualizations (spatial contiguity principle).

7. SIGNALLING

Signalling with text can shape and direct the reader's attention and thoughts about the visualizations[7,8,11]

Use text to help direct the reader's attention to key information in the data visualization to aid comprehension (signalling + coherence principle).

Author: Sakinah S. J. Alhadad
Visual Layout: Dale Hansen
Icons:
Icons 1, 3, 4, 5, 6, and 7 made by Freepik from www.flaticon.com
Icon 2 made by Those Icons from www.flaticon.com

References:
1. Anderson & Yantis, 2013
2. Baddeley, 2003
3. Baldassi, Megna, & Burr, 2006
4. Becker, Folk, & Remington, 2010
5. Desimone & Duncan, 1995
6. Gobet, 2005
7. Mautone & Mayer, 2007
8. Mayer & Fiorella, 2014
9. Rosenholtz, Li, & Nakano, 2007
10. Theeuwes, 2010

Appendix B

VISUALIZING DATA AS AN INFERENTIAL TOOL

Visualize data as an inferential tool to reveal important insights about underlying data properties

1. Visual representations can support cognition[4]
Visualize data as a key means of understanding and communicating information.

2. Visualizing data facilitates identification of patterns that are otherwise obscured in text or tables[3,6]

Before computing any statistics, visualize your data to evaluate assumptions of underlying univariate or multivariate data, and to detect any serious compromises to data integrity.

3. Particularly for complex, multivariate relationships, people are able to detect statistical effects with more speed and accuracy with visualizations than with tables[10,14]
For understanding research, visualize the data in various/multiple methods or representations to be able to better evaluate the ways in which underlying properties may be concealed by a single method.

For communicating research, consider evidence-informed principles to make methodological choices when constructing data visualizations to support attention and cognition.

4. Some types of visualizations inherently obscure the representation of data properties[1,8,9] (including uncertainty, see points 5–7)
Explore *multiple ways* of visualizing the data to make critical judgements about how visualization decisions may support or potentially lead to errors in interpretation and subsequent inference.
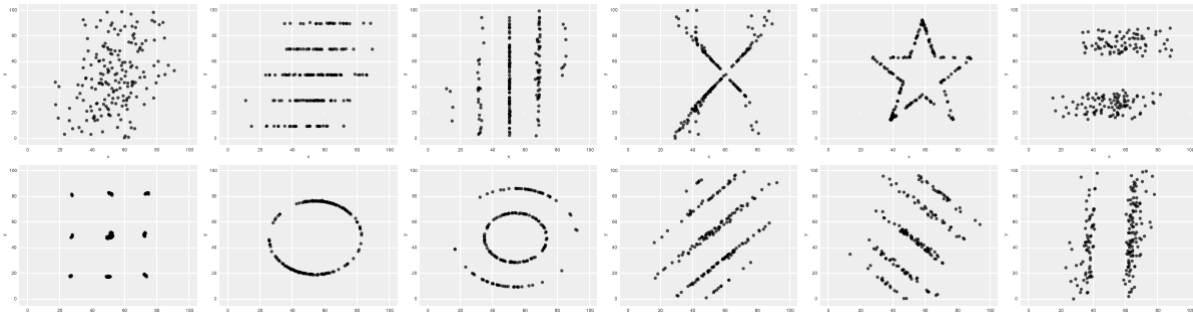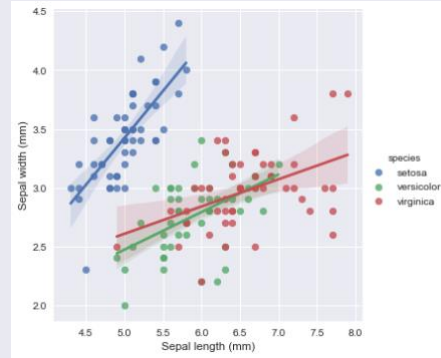
Visualize uncertainty to support inference-making

5. Visualizing uncertainty is a useful aid for researchers to make statistical and inferential decisions[2,7]
Visualize data in ways that enable you to see and understand the nature of the data dispersion and to test assumptions of the data.

6. Visualizing uncertainty appears to aid readers' understanding of criticality of data uncertainty than without its visualization; BUT people do not appear to naturally use uncertainty information to aid inference-making[5,7,11,12]
Emphasize the impact of data uncertainty in communicating the implications for research to practice translation. Use language to scaffold data uncertainty to support visualization of uncertainty, or use language to explain the implication of uncertainty in your dataset on inference-making explicitly.

7. Understanding data uncertainty is difficult for novices and experts[12,13,15,16]
Use text to orient attention to important aspects of uncertainty visualized, and why they matter.

Violin-plot
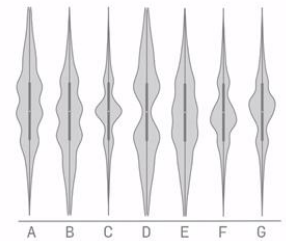
Author: Sakinah S. J. Alhadad

Visual Layout: Dale Hansen

Image Sources:
Multiple Linear Regression scatter plot courtesy of Michael Waskom.

Scatterplots and violin plots are from the "Datasaurus Dozen", courtesy of Justin Matejka and George Fitzmaurice.

Violin Plot courtesy of Justin Matejka.

References:
1. Anscombe, 1973
2. Burton, Pomeroy, Radenovic, & McCarley, 2017
3. Gelman, Pasarica, & Dodhia, 2002
4. Hegarty, 2011
5. Hoekstra et al., 2014
6. Kastellec & Leoni, 2007
7. Kramer, Telfer, & Towler, 2017
8. Krzywinski & Altman, 2014
9. Matejka & Fitzmaurice, 2017a
10. Meyer, Shamo, & Gopher, 1999
11. Morey et al., 2016