

Why Shouldnt All Charts Be Scatter Plots? Beyond Precision-Driven Visualizations

Enrico Bertini
New York University

Michael Correll
Tableau Research

Steven Franconeri
Northwestern University

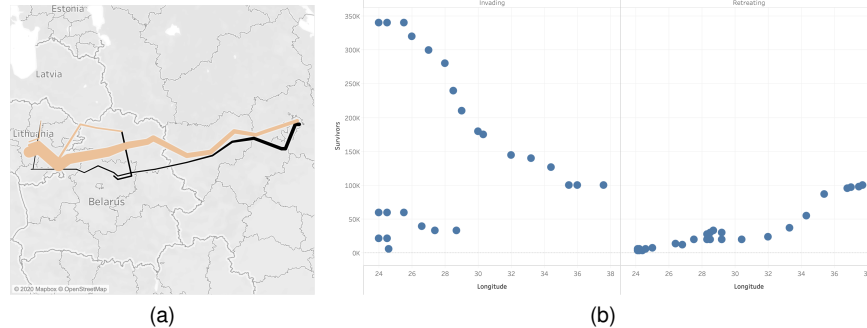


Figure 1: Two re-creations of Charles Minard’s map of the invasion of Russia in Tableau. In (1a) we hew as closely as possible to Minard’s design. In (1b) we attempt to encode all of the same data about the three groups of the invasion force using the most efficient channel: position on a common axis. In which circumstances might we prefer (1a) over (1b)?

ABSTRACT

A central tenet of information visualization research and practice is the notion of visual variable *effectiveness*, or the perceptual precision at which values are decoded given visual channels of encoding. Formative work from Cleveland & McGill has shown that position along a common axis is the most effective visual variable for comparing individual values. One natural conclusion is that *any* chart that is not a dot plot or scatterplot is deficient and should be avoided. In this paper we refute a caricature of this “scatterplots only” argument as a way to call for new perspectives on how information visualization is researched, taught, and evaluated.

Index Terms: Human-centered computing—Visualization—Visualization theory, concepts and paradigms

1 INTRODUCTION

Minard’s famous map of the Grande Armée’s invasion of Russia has been called one of the “best statistical drawings ever created” [43] and presents complex geographic, logistical, and weather data simultaneously. It is possible to recreate many aspects of Minard’s map in common visualization systems like Tableau (1a). However, many of the encodings used in the map (such as width of lines with arbitrary, non-aligned angles) are comparatively imprecise for the estimation of values. A naive recreation of the same data, based purely on the “efficiency” of visual variables, might look more like (1b). In (1b), the user can more precisely determine the size of individual groups of the Armée at specific time points. A quantitative evaluation of (1b)’s performance (in terms of response time and accuracy) might very well find it to be superior to (1a) on that basis. And yet, the conclusion that (1b) is a strictly superior visualization, or that it would be equally iconic and compelling as Minard’s map, seems unfounded. When might we prefer one version over the other, and what empirical evidence exists in the visualization literature to ground these preferences?

This example highlights an apparent contradiction at the heart of information visualization. On one hand, our exemplars of good visualizations can be diverse, complex, and reward contemplation [19]. On the other hand, our foundational empirical results and rules of thumb are often simple and minimalist. These rules are typically evaluated in terms of how quickly and accurately people extract specific information from charts, including formative psychophysical studies showing that viewers extract data values most precisely when they are encoded via position on shared axes [7].

Given these constraints, a natural conclusion is that *quantitative data should almost always be depicted in a dot plot or scatter plot*, perhaps breaking data into SPLOMS, small multiples, or employing brushing and linking when there are too many variables for one view. While this argument is a strawman, its premises lie at the heart of foundational visualizations books by authors like Bertin and Tufte, and embodied as charts of rankings of visualization effectiveness in text books that are at the heart of how we teach visualization to students [31, 47]. In our own teaching we have struggled with how to convey these perceptual and design principles without resorting to at least some form of this argument. Therefore, we find this strawman useful to knock down: the goal of this paper is to highlight additional set of constraints that compete with perceptual precision, both in the mind of the designer and the studies of the researcher. We argue for a more expansive view of visualization beyond the perceptually precise encoding and decoding of individual data values, and make the case for “inefficient” visualizations.

Our refutation of this argument focuses on attacking three hidden premises, critique of which reveals three classes of insufficiencies. The first premise is that accuracy in the extraction of data values is a sufficient measure of *perceptual precision*. The second premise is that perceptual precision is a sufficient measure of a chart’s utility in *communicating data*. The third premise is that utility in communicating data is sufficient to understand the larger *purpose and power* of visualizations. We address these three premises in turn, using the insufficiencies associated with each to motivate a more expansive view of visualization design and pedagogy, and to suggest directions for future research.

2 BEYOND INDIVIDUAL VALUES

Viewers are most efficient at computing the ratio between two visualized values when those values are encoded by their position on a common axis, as in a dot plot. The dominance of position encodings for this task is followed by a ranked list of other encodings including 2D area and orientation, and with intensity typically listed as the least precise encoding [7]. There is an implicit assumption that the ranking derived from this two-value ratio judgment represents an atomic unit for visualization, so that the additional precision conveyed by position should transfer to better perceptual performance in more complex tasks.

We challenge this assumption. While the precision of ratio judgments is one operationalization of perceptual performance, there are many others. For example, one fundamental analytics task is seeing the ‘big picture’ in a dataset. Figure 2, provides an example, showing a 6×12 grid of data values plotted as a dot plot (position), bar graph (position + length), bubble chart (area), line graphs (juxtaposed and superposed), and heatmap. If precision of value extraction is all that matters, then the dot plot should be the preferred design. In contrast, it is clear to our eyes that the position-encoded dot plot is the *least* effective visualization for seeing many potential ‘big pictures’ of the data. The bar graph to its right is far more useful for this task, likely because it adds a redundant encoding of length (or more likely, area [50]). The line graph below is useful because it adds an emergent encoding of the local deltas between points via the orientation of the lines. Our favorite ‘big picture view’ is actually the heatmap in the corner, despite its status as the bottom of the barrel for precision of extracting ratios between individual values.

Table 1 depicts a list of other likely perceptual tasks, inspired by work on low-level task taxonomies [3] & recent papers that examine visualization through the lens of perceptual psychology [1, 8, 21, 49]. We do not claim that this list is exhaustive, representative, or even correct. It is instead intended to show that ratio judgment tasks are only a small subset of likely perceptual tasks. The second column of Table 1 provides concrete examples of the abstract perceptual tasks within that concrete example.

In the first two rows of the table, we list two perceptual operations that can be computed over 2 points: metric relations between metric numbers (a ratio) and ordinal relations between pairs of metric values (is value A higher than B?), say the first two points of the first row of Figure 2. While data visualization research has focused almost exclusively on the former, the latter is arguably as important for real-world tasks. While we occasionally note that today is five degrees hotter than yesterday, we more typically note that it is hotter than yesterday. COVID-19 infection rates have increased. Profits are lower, and we are over budget.

The next set of rows list tasks that might unfold when a viewer is presented with many 2-point pairs of values, such as the first two rows of one of the visualizations in Figure 2. These tasks include metric comparisons, such as finding which pair has the largest ratio, estimating an average ratio, or clustering the sizes of ratios. They also include ordinal comparisons of metric value pairs, such as finding a unique relation ($A \leq B$ among $A \geq B$ pairs), or estimating which relations type is more frequent. We know of only two studies that have studied these important perceptual tasks [33, 39], but because both rely primarily on position encodings, they cannot confirm whether the Cleveland & McGill position ranking holds for these alternative tasks.

The next several rows show perceptual tasks that are not constrained over pairs of points, and instead could be computed over an entire set or subset of N values. These include identifying a single value with a given property (e.g., min, max, outlier), or summarizing a set of values by a single number (e.g., mean, variance, clusters). Recent work on the perception of these “aggregate” or “ensemble” tasks [1, 40] provides evidence that many encodings that are imprecise for individual values (such as color) have performance benefits

for these types of tasks over positional encodings like line charts.

The row labeled ‘Shape, trend’ refers to the need to holistically judge a single series, or to compare two series, in an open-ended manner. We suspect that this task, like stepping back to see the ‘big picture’ in the data, will not always be best supported by position encodings. The viewer might search for anything from basic patterns (rising, flat), to idiosyncratic motifs and shapes in the data [15]. Visual interfaces for times series search [24] have had to find ways for users to express shapes (and the properties of those shapes that they find important [11]) in fluid and dynamic ways, as the rigid definition of specific individual values may not capture the visual features of interest to the user.

The row labeled ‘filter’ refers to a visual subset operation based on the data values, e.g., pick out all of the high values. We do not have a full understanding of the filtering operation in visualization contexts, although existing work evaluates the detection of individual “oddball” outliers [16] or filtering across nominal categories (for instance, picking a particular class of points out of a scatterplot [14]). However, perceptually motivated designs for time series data have often used color as a form of perceptual “boosting” [34] to highlight anomalous items [2, 9].

Finally, recent work has begun to uncover the perceptual tasks that underlie more complex comparisons, such as judging the correlation between two sets of paired values [17, 36, 49]. This work suggests that, instead of judging high-level properties like correlation *per se*, a viewer relies on a more concrete proxy, such as the aspect ratio of the bounding box surrounding the points [49]. These hypothesized proxies may prove to be complex and may take many years to unpack – but when they are better understood, we cannot predict whether they would be best supported by position encodings.

3 METAPHORS AND CONGRUENCE

The choice of encoding channels is constrained by more than perceptual precision, with one major constraint being the congruency of its metaphors [27]. A channel must be consistent with (and convey) the concepts that they encode – for example, conveying quantity with area – serving as type of “affordance” for the data to show how it can be used (e.g., a push plate vs. a pull-bar for a door) [32].

One example of this conceptual congruence is a study that asked participants to describe simple bar and line graphs of the same two data points [51]. Bar charts lead to descriptions in terms of discrete comparisons whereas lines lead to descriptions of trends; indicating that different graphical solutions can suggest different type of interpretations. Interestingly, in these examples the channels used to convey quantitative information were identical (i.e., vertical position) and the only element that changes between the graphs is the affordance of the connected line implying continuous data, and the visual connection between points in the line graph.

Another example comes from one author’s experience with an exercise assigned in his information visualization class. The assignment asked the students to compare a set of countries in terms of amount of money donated or received, as recorded in the Aid Data data set ¹ (which records international aids disbursements globally). Students produced two main type of solutions: (1) a scatter plot with dots representing the countries and axes representing incoming and outgoing amounts; (2) a pair of aligned bar charts showing for each country the amount donated and the amount received. While both solutions employ position as the main channel to encode the donated/received amounts, the graphs invite the reader in making completely different sets of judgments. More precisely, while the scatter plot affords detection of correlations and groupings, the pair of aligned bar charts invites the reader to compare individual countries across two metrics (see Fig. 2 for a similar design comparison). These examples show that the ranking of visual channels (based

¹ <https://www.aiddata.org/>

Task	Example	Existing Work
1 pair of 2 points	<i>Between months 1 and 2 in 2014...</i>	
Metric: Retrieve value / two-point ratio	...what is their ratio?	[7, 18, 41]
Ordinal: What's the relation of this pair?	...which is bigger?	
N pairs of 2 points	<i>For months 1 and 2 across all years...</i>	
Metric: N pairs largest ratio	...which year had the largest ratio between values?	[39]
Metric: N pairs average ratio	...what is the average ratio between months across all years?	[33]
Metric: N pairs cluster into M relation types	...what types of ordinal relations are present?	
Ordinal: Are all ordinal relations present?	...do they all fall, or do any rise?	
Ordinal: Modal pair relation type	...is one ordinal relationship type most frequent?	[33]
N-point (patterns)	<i>Across all months and years...</i>	
Single point identification (min/max, outlier)	...which single data point is largest?	[40]
Summarization (mean, range)	...was 2011 or 2013 higher on average?	[40]
Cluster	...where would you place a threshold for low vs. high values?	
Extrapolation, imputation	...what is your guess for the value of the first month of 2016?	[38]
Shape, trend	...how does the trend visually differ between 2012 and 2013?	[11, 15, 24]
Filter	...which year has the most near-zero values?	
Statistical tasks	...which other year is most highly correlated with 2010?	[17, 21, 36, 49]

Table 1: A short list of potential value-related tasks in visualizations. Our recommendations for visual encodings and visualization designs are often related to only a narrow set of these tasks (often the first two, dealing with individual pairs of values). Empirical evidence about the precision of visual encodings for other tasks is often sparse or counter to existing recommendations.

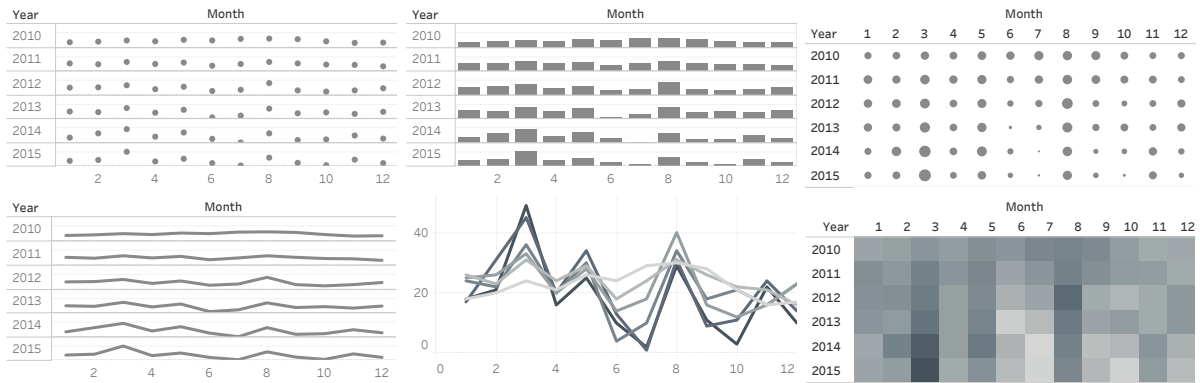


Figure 2: Sample data for the tasks laid out in Table 1. Consider which tasks are subjectively easier or harder across these different designs.

on precision) is not *sufficient* in the visualization design space. In other words, knowing that a channel affords more precision does not provide sufficient guidance for visualization design.

The role of metaphors and expressiveness can be observed at multiple levels of granularity. At the level of individual channels there are several examples of how channels may express or fail to express certain types of information. For example, color hue cant express ordinal or quantitative information because the human eye does not assign an order to colors that vary exclusively in hue. Similarly, colors have strong semantic associations, therefore appropriate associations between concepts and colors may improve readability and comprehension [25]. Area size is a moderately precise channel when conveying quantity, but it cannot easily show negative values because larger sizes are firmly associated with larger (positive) quantities.

Different semantic associations can also be created by using different symbols or graphical marks. A classic example is the representation of part-to-whole relationships and the question of whether pie charts should be considered effective solutions for the representation of such data [37]. When in a visualization the designer wants to explicitly convey information about the fact that a given value is part of a whole, specific metaphors work better than others. For example, in comparing a pie chart, a stacked bar and a group of bars, it is evident that only the pie chart and the stacked bar explic-

itly convey the part-to-whole metaphor. Following the reasoning behind the ranking of visual variables, the solution with separate bars (position encoding) should be preferred over the stacked bar (length encoding) or pie chart (angle and area encoding) because it provides a more precise representation. Other similar examples of this kind exist. For instance, bars on maps are rarely used, whereas circles are often preferred in their place. Line charts are preferred over bars when the goal is to convey a temporal trend. Icon arrays are preferred over aggregate values in risk estimation. All of these examples demonstrate that there is something more than ranking of visual channels and that reasoning about visualization at the level of individual channels can be limited and potentially misleading.

Even a combination of accuracy and efficiency cannot fully characterize the effectiveness of data visualizations. One must also measure how easy it is to extract information out of it.

Two concepts developed in the literature on cognitive science seem to be of pertinence here. The first one is the “congruence principle” suggested by Tversky et al. [44]. The principle states that “the content and format of the graphic should correspond to the content and format of the concepts to be conveyed” and it seems to apply perfectly to the type of concerns we discussed above. The second concept is “cognitive fit,” developed by Vessey. In the words of Vessey [45]: “... performance on a task will be enhanced when there is a cognitive fit (match) between the information emphasized

in the representation type and that required by the task type.” While the theory of cognitive fit has been developed originally to explain the difference between symbolic and graphical representations (i.e., tables vs graphs), there is no reason to believe the same logic can’t be used to describe differences between alternate graphical representations. A good matching between the “information emphasized in the representation type” and the information a reader is expected to extract seems to be a good guiding principle for data visualization.

4 RHETORIC, PERSUASION, AND MEMORY

A last category of objections to a world of only scatterplots is that many visualizations are unconcerned with accurate extraction of individual (or even aggregate) values. Charts are often designed to persuade, educate, and motivate. Designing for serendipitous discovery, educational impact, hedonic response, or changes in behavior is in some cases only tangentially connected with the precision of a particular visualization. Wang et al. [46] call for us to “revis[e] the way we value visualizations” on this basis, and Correll & Gleicher [10] point to a whole class of designs and design guidelines that seem counter-productive in terms of precision but that nonetheless result in benefits in terms of higher-level cognitive goals. In this section we briefly discuss some of these mismatches.

Hullman et al. [20] point to possible benefits for “visual difficulties in charts: that is, by making the viewer do more work to decode the values, there is potentially an impact on the retention of those values. Rather than designing charts to be as precise as possible, for longer-term or higher-level tasks we may wish to slow the viewer down. Lupis [26] Data Humanism manifesto calls for visualizations that encourage viewers to “spend time with the data, with examples of dense, multidimensional glyph-based visualizations that do not afford quick and precise extraction of values. Similarly, Bradley et al. [6] call for a “slow analytics movement that encourages ownership and retention of analytical tasks rather than precision.

Part of the pedagogical utility of charts is not merely conveying the information, but ensuring that the information is retained. One immediate downside to a world of only scatter and dot plots is that our charts would all look similar, and so unlikely to be differentiated much in memory. Borkin et al. [4, 5] find that charts with pictorial elements and other visual features of interest are more memorable than plain and otherwise unadorned charts. Kostelnick [23] recommends occasional deviations from minimalist design in the service of “clarity” which can include such factors as engaging the reader’s attention. Many of the most impactful charts in visualization have had non-standard or otherwise less than precise forms (e.g., Fig 1a).

There may be benefits for imprecise visualizations for analysts as well, not just for passive viewers or learners. Often when designing a system we may have no idea of the form or category of insights present in our data. The serendipitous discovery of important features of the data-set may not be well-covered by existing design principles that are designed for the precision at intended or standard analytical tasks: lucky, chance, and stochastic exploration may be more important than reliably picking out values. Thudt et al. [42] discuss the challenges of designing for serendipitous information discovery, and suggest that standard designs may be ill-suited to the unconstrained and stochastic sort of exploration that can be necessary for making discoveries, whereas Dörk et al. [12] point to the challenges of designing for the wandering “information flneur.”

There are also potentially costs to overly-precise visualizations. Kennedy et al. [22] claim that the “clean layouts” of minimalist visualizations can grant an imprimatur of authority and objectivity to data that may not match that standard. Likewise, Drucker [13] points to the “seductive rhetorical force” of visualizations to convince viewers that the data they contain is not merely a potentially flawed, biased, and uncertain view, but an objective truth about the world. This unwillingness to question charts due to a perception of their objectivity can override even strong political conventions

or skepticism [35]. “Messier” designs (such as sketchy [48] or uncertainty-conveying [38] renderings) can introduce a willingness to critique or a greater appreciation for uncertainty not present in more precision-driven visualizations.

5 CONCLUSION

Much of the empirical and theoretical basis for visualization work comes from studies examining the efficiency of visual channels in extracting information, and using these results to generate a ranking of these channels [7]. These rankings power many of our design guidelines and constraints [30], are ubiquitous in our textbooks [31, 47], and are instantiated in the logic of many of our automated or semi-automated visualization design tools [27, 28]. And yet, these rankings do not seem to capture important components of how people use, interpret, and learn from visualizations. We should be expansive in how we analyze, conceptualize, and teach visualization. Otherwise, we risk a situation where academia focuses on the narrow, scatterplot-like section of the vast, more interesting world of visualization as a whole.

Of course we are not proposing to throw the baby out with the bathwater. The ranking of visual variables has had enormous impact on visualization research and practice, informing design decisions for tool development and providing pedagogical value in numerous guidelines, textbooks, and courses. Our intent is to raise awareness about the ways an excessive and narrow focus on channel ranking may be acting as a detrimental limitation to our field in terms of: (1) understanding actual data visualization practice; (2) development of data visualization tools and techniques; (3) methodologies for data visualization design and evaluation and (4) pedagogy of data visualization. The question is: how can we rectify and expand the theory behind the ranking of visual variables? When does it work? When does it not work? And, maybe even more importantly, what else to do we need in its place or in addition to it? From this initial analysis of the various insufficiencies we have identified it seems clear there is much to do in this area. It is our hope that this work sparks interesting conversations and potentially lead other practitioners and designers to develop alternative (or more refined) practices, conceptualizations, and epistemologies [29] for visualization.

ACKNOWLEDGMENTS

This work was supported by NSF awards IIS-1901485 and IIS-1900941.

REFERENCES

- [1] D. Albers, M. Correll, and M. Gleicher. Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 551–560, 2014.
- [2] D. Albers, C. Dewey, and M. Gleicher. Sequence surveyor: Leveraging overview for scalable genomic alignment visualization. *IEEE transactions on visualization and computer graphics*, 17(12):2392–2401, 2011.
- [3] R. Amar and J. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *IEEE Symposium on Information Visualization*, pp. 143–150. IEEE, 2004.
- [4] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics*, 22(1):519–528, 2015.
- [5] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013.
- [6] A. J. Bradley, V. Sawal, and C. Collins. Approaching humanities questions using slow visual search interfaces. In *IEEE 4th Workshop for Visualization and the Digital Humanities*, 2019.

- [7] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.
- [8] M. Correll, D. Albers, S. Franconeri, and M. Gleicher. Comparing averages in time series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1095–1104, 2012.
- [9] M. Correll, A. L. Bailey, A. Sarikaya, D. H. OConnor, and M. Gleicher. Layercake: a tool for the visual comparison of viral deep sequencing data. *Bioinformatics*, 31(21):3522–3528, 2015.
- [10] M. Correll and M. Gleicher. Bad for data, good for the brain: Knowledge-first axioms for visualization design. In *DECISive: Workshop on Dealing with Cognitive Biases in Visualisations*, 2014.
- [11] M. Correll and M. Gleicher. The semantics of sketch: Flexibility in visual query systems for time series data. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 131–140. IEEE, 2016.
- [12] M. Dörk, S. Carpendale, and C. Williamson. The information flaneur: A fresh look at information seeking. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1215–1224, 2011.
- [13] J. Drucker. Humanistic theory and digital scholarship. *Debates in the digital humanities*, pp. 85–95, 2012.
- [14] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *IEEE transactions on visualization and computer graphics*, 19(12):2316–2325, 2013.
- [15] A. Gogolou, T. Tsandilas, T. Palpanas, and A. Bezerianos. Comparing similarity perception in time series visualizations. *IEEE transactions on visualization and computer graphics*, 25(1):523–533, 2018.
- [16] S. Haroz and D. Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2402–2410, 2012.
- [17] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber’s law. *IEEE transactions on visualization and computer graphics*, 20(12):1943–1952, 2014.
- [18] J. Heer and M. Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 203–212, 2010.
- [19] J. Heer, M. Bostock, and V. Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, 53(6):59–67, 2010.
- [20] J. Hullman, E. Adar, and P. Shah. Benefitting infovis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2213–2222, 2011.
- [21] N. Jardine, B. D. Ondov, N. Elmqvist, and S. Franconeri. The perceptual proxies of visual comparison. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1012–1021, 2019.
- [22] H. Kennedy, R. L. Hill, G. Aiello, and W. Allen. The work that visualisation conventions do. *Information, Communication & Society*, 19(6):715–735, 2016.
- [23] C. Kostelnick. The visual rhetoric of data displays: The conundrum of clarity. *IEEE Transactions on Professional Communication*, 51(1):116–130, 2008.
- [24] D. J.-L. Lee, J. Lee, T. Siddiqui, J. Kim, K. Karahalios, and A. Parameswaran. You can’t always sketch what you want: Understanding sensemaking in visual query systems. *IEEE transactions on visualization and computer graphics*, 26(1):1267–1277, 2019.
- [25] S. Lin, J. Fortuna, C. Kulkarni, M. Stone, and J. Heer. Selecting semantically-resonant colors for data visualization. *Computer Graphics Forum*, 32(3pt4):401–410, 2013.
- [26] G. Lupi. Data humanism: the revolutionary future of data visualization. *Print Magazine*, 30, 2017.
- [27] J. Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141, 1986.
- [28] J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics*, 13(6):1137–1144, 2007.
- [29] M. Meyer and J. Dykes. Criteria for rigor in visualization design study. *IEEE transactions on visualization and computer graphics*, 26(1):87–97, 2019.
- [30] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE transactions on visualization and computer graphics*, 25(1):438–448, 2018.
- [31] T. Munzner. *Visualization analysis and design*. CRC press, 2014.
- [32] D. Norman. *Things that make us smart: Defending human attributes in the age of the machine*. Diversion Books, 2014.
- [33] C. Nothelfer and S. Franconeri. Measures of the benefit of direct encoding of data deltas for data pair relation perception. *IEEE transactions on visualization and computer graphics*, 26(1):311–320, 2019.
- [34] D. Oelke, H. Janetzko, S. Simon, K. Neuhaus, and D. A. Keim. Visual boosting in pixel-based visualizations. In *Computer Graphics Forum*, vol. 30, pp. 871–880. Wiley Online Library, 2011.
- [35] E. M. Peck, S. E. Ayuso, and O. El-Etr. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.
- [36] R. A. Rensink and G. Baldrige. The perception of correlation in scatterplots. *Computer Graphics Forum*, 29(3):1203–1210, 2010.
- [37] D. Skau and R. Kosara. Arcs, angles, or areas: Individual data encodings in pie and donut charts. *Computer Graphics Forum*, 35(3):121–130, 2016.
- [38] H. Song and D. A. Szafr. Where’s my data? evaluating visualizations with missing data. *IEEE transactions on visualization and computer graphics*, 25(1):914–924, 2018.
- [39] A. Srinivasan, M. Brehmer, B. Lee, and S. M. Drucker. What’s the difference? evaluating variations of multi-series bar charts for visual comparison tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.
- [40] D. A. Szafr, S. Haroz, M. Gleicher, and S. Franconeri. Four types of ensemble coding in data visualizations. *Journal of vision*, 16(5):11–11, 2016.
- [41] J. Talbot, V. Setlur, and A. Anand. Four experiments on the perception of bar charts. *IEEE transactions on visualization and computer graphics*, 20(12):2152–2160, 2014.
- [42] A. Thudt, U. Hinrichs, and S. Carpendale. The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1461–1470, 2012.
- [43] E. R. Tufte. *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT, 2001.
- [44] B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.
- [45] I. Vessey. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision sciences*, 22(2):219–240, 1991.
- [46] Y. Wang, A. Segal, R. Klatzky, D. F. Keefe, P. Isenberg, J. Hurtienne, E. Hornecker, T. Dwyer, and S. Barrass. An emotional response to the value of visualization. *IEEE computer graphics and applications*, 39(5):8–17, 2019.
- [47] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.
- [48] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2749–2758, 2012.
- [49] F. Yang, L. T. Harrison, R. A. Rensink, S. L. Franconeri, and R. Chang. Correlation judgment and visualization features: A comparative study. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1474–1488, 2018.
- [50] L. Yuan, S. Haroz, and S. Franconeri. Perceptual proxies for extracting averages in data visualizations. *Psychonomic bulletin & review*, 26(2):669–676, 2019.
- [51] J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory & cognition*, 27(6):1073–1079, 1999.