Four types of ensemble coding in data visualizations

Danielle Albers Szafir
Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI, USA

Steve Haroz
Department of Psychology, Northwestern University, Evanston, IL, USA

Michael Gleicher
Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI, USA

Steven Franconeri
Department of Psychology, Northwestern University, Evanston, IL, USA

Ensemble coding supports rapid extraction of visual statistics about distributed visual information. Researchers typically study this ability with the goal of drawing conclusions about how such coding extracts information from natural scenes. Here we argue that a second domain can serve as another strong inspiration for understanding ensemble coding: graphs, maps, and other visual presentations of data. Data visualizations allow observers to leverage their ability to perform visual ensemble statistics on distributions of spatial or featural visual information to estimate actual statistics on data. We survey the types of visual statistical tasks that occur within data visualizations across everyday examples, such as scatterplots, and more specialized images, such as weather maps or depictions of patterns in text. We divide these tasks into four categories: *identification* of sets of values, *summarization* across those values, *segmentation* of collections, and *estimation* of structure. We point to unanswered questions for each category and give examples of such cross-pollination in the current literature. Increased collaboration between the data visualization and perceptual psychology research communities can inspire new solutions to challenges in visualization while simultaneously exposing unsolved problems in perception research.

Introduction

Some types of visual information must be extracted from small numbers of objects at a time, such as complex object identity (Wolfe, 1998) or spatial relationships (Franconeri, Scimeca, Roth, Helseth, & Kahn, 2012). Other types of information can be extracted and combined in parallel from large numbers of objects at once, such as the average object size (Ariely, 2001). A growing body of work seeks to understand such *ensemble coding* of spatially distributed visual information (for surveys, see Alvarez, 2011; Whitney, Haberman, & Sweeny, 2014). Researchers typically study this ability in order to draw conclusions about how ensemble coding helps extract information from natural scenes. For example, one might want to estimate the number of books on a shelf (Ross & Burr, 2010) or gauge the average emotional expression within a crowd of people (Haberman & Whitney, 2007).

Here we argue for another domain that should serve as an equally exciting inspiration for understanding ensemble coding: visual presentations of data (e.g., maps, charts, and graphs). Data visualizations are ubiquitous to students, scientists, and any broader audience that reads graphs, uses maps, or reads a newspaper. Visualizations communicate patterns in data by mapping data dimensions to visual features (for an overview, see Bertin, 1983; Heer, Bostock, & Ogievetsky, 2010). To illustrate, consider a scatterplot, which maps data values to spatial positions. For some types of inspection, such as mapping symbols to a legend or knowing whether a particular data value is lower or higher than another, we must serially inspect small numbers of data values at a time. But other types of information can be extracted in parallel, such as the approximate mean position, or size, of an entire cloud of points (Figure 1a) or the portion of a line graph with the highest variability (Figure 1b).

These judgments are ensemble judgments, and they merit more intense study both for their value as a case

Citation: Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of Vision*, *16*(5):11, 1–22, doi:10.1167/16.5.11.

doi: 10.1167/16.5.11 Received March 17, 2015; published March 16, 2016 ISSN 1534-7362

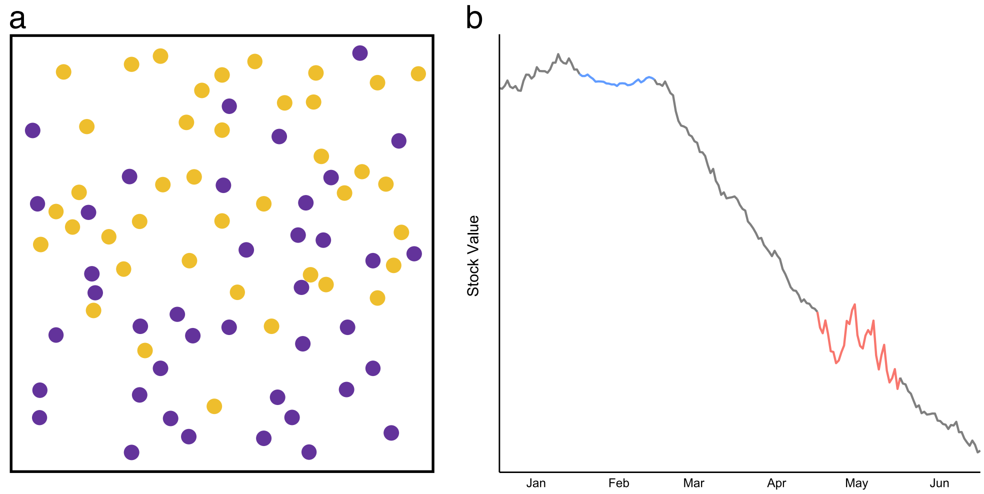


Figure 1. Even without explicitly shown statistics, a viewer can quickly and robustly observe that, on average, the orange dots have a higher *y* value than the purple dots, or that there is more variance in May (highlighted in red) than February (highlighted in green).

study in understanding how ensemble coding works in the visual system and for their practical importance within information visualization. Information visualization research and practice has been previously inspired by research in cognition and perception (for surveys, see Healey & Enns, 2012; Ware, 2008, 2013; for a framework for reasoning about perceptions of visualization designs, see Rensink, 2014). However, existing work focuses on how perception might inform visualization design. We instead aim to inspire a broader two-way conversation between vision science and visualization—understanding how viewers estimate properties of visualized data offers potential research directions for vision science, and this understanding can in turn inform more effective visualization designs.

Experiments in visualization have tried to quantify how effectively viewers perceive different properties encoded using various visual features. However, much of this work focuses on single-value tasks, such as finding and estimating values from individual data points (for examples, see Cleveland & McGill, 1984; Heer, Kong, & Agrawala, 2009; Javed, McDonnel, & Elmqvist, 2010). More recent work has begun to study ensemble coding in data visualization, such as the construction of averages within a scatterplot (Gleicher, Correll, Nothelfer, & Franconeri, 2013); variance, range, and outliers in line graphs and heat maps (Albers, Correll, & Gleicher, 2014); estimation and comparison of correlation (Harrison, Yang, Franconeri, & Chang, 2014; Rensink & Baldridge, 2010); and numerosity judgments in unit charts (Haroz, Kosara, & Franconeri, 2015). Our goal is to identify a broader set of such visualization tasks that benefit from ensemble coding and to increase research and discussion surrounding how these judgments work and how visual data displays can be better designed to support them.

In principle, the statistics that viewers perceive in a data visualization could be formally computed and shown directly to the viewer. However, visual extraction of statistics is often more attractive because formally computed statistics, which necessarily abstract over potentially critical patterns, are often insufficient to describe data. Try to imagine a scatterplot of a data set that exhibits the following statistics: the x and y variables both have a mean of 7.5 and variance of 5, and the correlation coefficient of x and y is 0.816. You are probably imagining that the underlying data look like the first plot in Figure 2. But any of the four data sets shown in Figure 2 would produce these statistics—all four have identical means across both x and y, variabilities across both x and y, correlation coefficients, and linear regression formulas ($y = 3 + 0.5x$; Anscombe, 1973). Yet each plot exhibits qualitatively different patterns. While there are increasingly complex statistics that could differentiate among these patterns, these statistics would not likely be run without the benefit of visual inspection to determine their necessity. Alternatively, attempting to provide statistical information explicitly, even through visual means, can quickly become cluttered and overwhelm the viewer, even for a small number of statistics (Figure 3). Visual estimation also provides a beneficial flexibility in terms of what data are being processed: Viewers have direct control over the different subsets of the data they choose to compute statistics for.

If allowing observers to extract statistics and patterns about data with their visual system offers an alternative to explicitly providing raw statistics, then understanding the effectiveness of this process is critical. The benefits of efficiently estimating visual statistics provide both an application of and challenges for research on the perception of these features. How

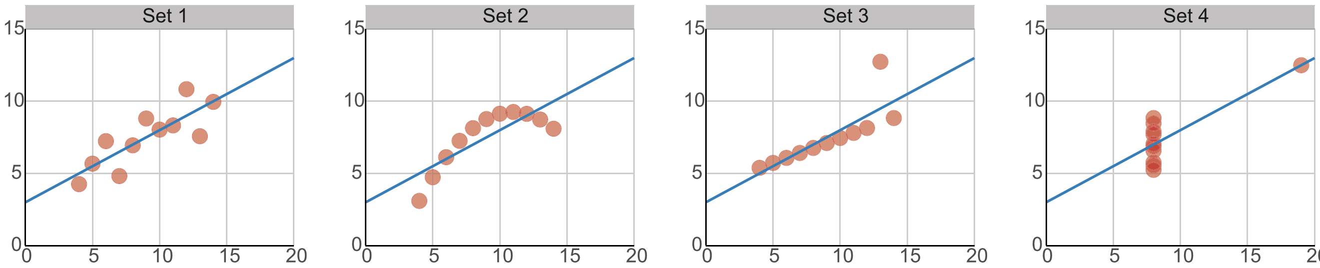


Figure 2. Common statistical abstractions may not capture potentially critical patterns in data. This example, from Anscombe (1973), shows four data sets that are identical across several common statistics (mean, variance, correlation, and linear regression) yet contain qualitatively different patterns. Visual inspection offers powerful and flexible processing of these differences, as well as rough approximations of statistics.

do the capabilities of the visual system match the needs of visual depictions of data? Conversely, the difficulties encountered in data visualization can challenge our understanding of perception. When perceptual psychologists cannot answer the questions asked by visualization designers, it shows the psychologists the gaps in their theory—what they did not realize that they did not know.

We organize our exploration of this synergy between perception research and visualization around two key questions: What visual statistics can our perceptual system extract via ensemble coding, and what potential needs in visualization can these ensembles address? We can align visual statistics with visualization needs by understanding the different kinds of tasks viewers might want to accomplish. In visualization, tasks are, informally, the visual operations that people may want to perform with data, such as identifying points with high values or estimating the average of a set of values. A flurry of recent efforts in the data visualization community propose taxonomies and typologies of tasks (Amar, Eagan, & Stasko, 2005; Roth, 2012; Schulz, Nocke, Heitzler, & Schumann, 2013; Shneiderman, 1996), creating abstractions that seek to help knowledge gained in one environment transfer to visualizations with differing contexts and details (for an extensive survey and comparison of prior efforts, see Brehmer & Munzner, 2013). However, these taxonomies generally attempt to classify techniques used by designers rather than understand how properties of the data might be perceived in different designs.

In order to address the questions around ensemble coding that bridge perception and visualization, we need a categorization of visual tasks at the perceptual level: basic visual operations that serve as building blocks for more complex analyses. In this article, we introduce an organization of low-level tasks that require, or may require, ensemble coding into a framework of four categories: identification, summarization, segmentation, and structure estimation. Figure 4 depicts these categories, as well as examples of each, for four common ways of visually depicting data values (position, size, orientation, and color). Figure 5 demonstrates examples of these tasks applied to more complex visualization systems. Understanding which combinations of visual feature and task are most effective is a critical challenge. What statistics and patterns can we accurately extract, which are inaccurate, and which are systematically biased? How does the choice of feature used to represent the data affect our ability to extract absolute values, statistics, and patterns from data sets?

Both the perception and visualization research communities should explore and refine—or even completely reinvent—the grid of tasks and features in Figure 4. For perception research, it holds a diverse set of unsolved problems, not only for understanding ensemble coding across different features and statistics but also for revealing unsolved questions surrounding visual search, multifocal attention, and visual comparison (for a review of these topics, see Franconeri, 2013).

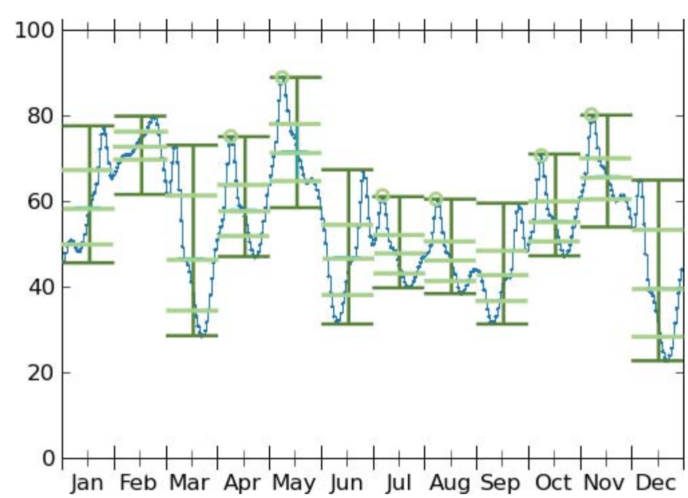


Figure 3. Providing explicit statistics (in this case, minimum, maximum, mean, variance, and outliers per month) can be overwhelming, even in a visual format.

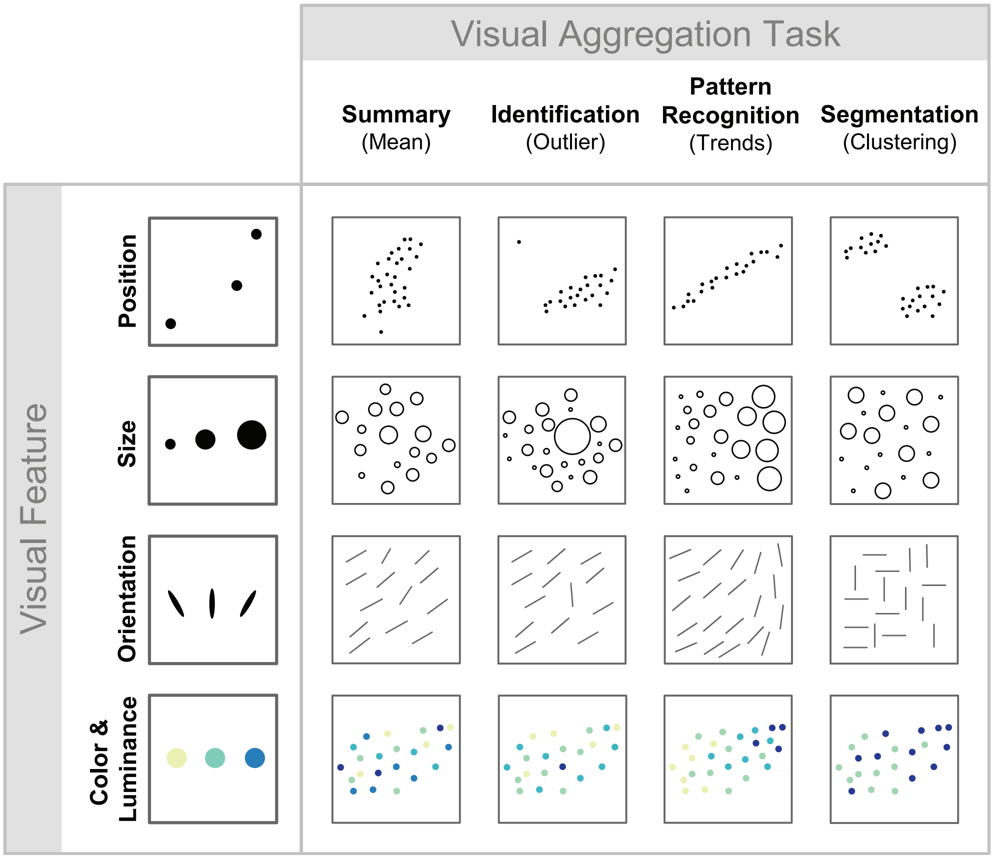
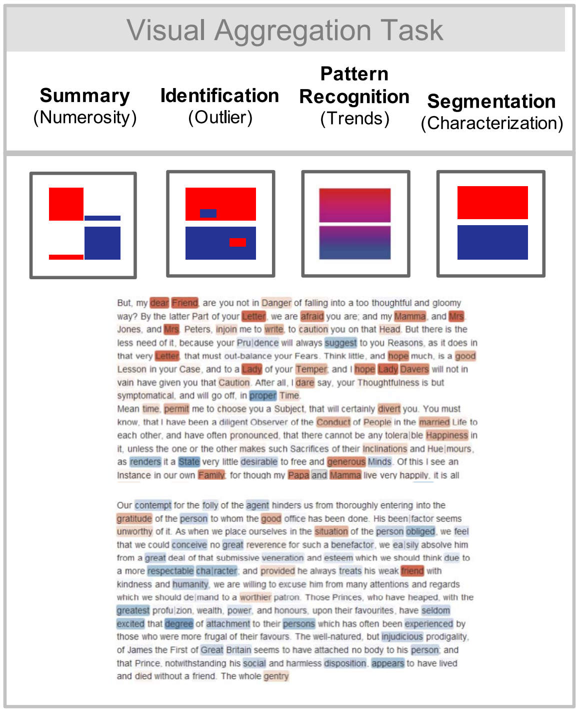


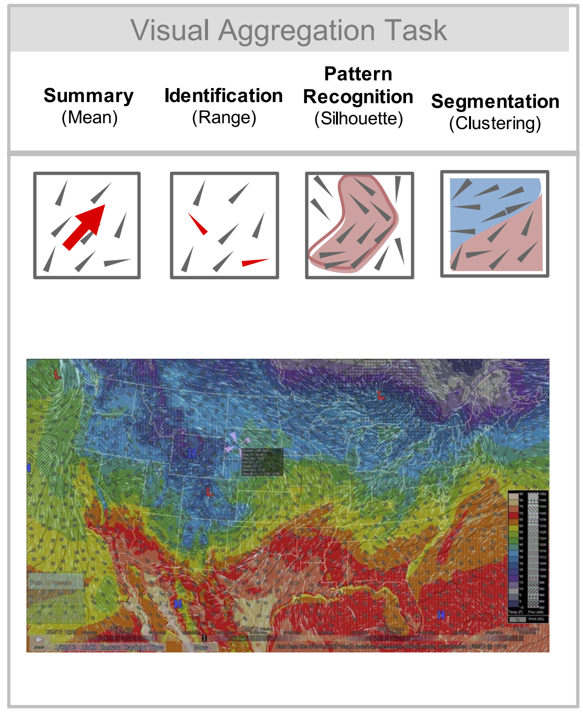
Figure 4. We identify four categories of visualization tasks (top) that require ensemble coding of information spread throughout the visual field. The tasks can be performed on multiple visual features, but not necessarily with equal speed or efficiency. In visualizations, choosing which visual feature is mapped to each dimension of the data set affects which tasks are most easily performed on which data dimensions (e.g., perhaps it is best to map size to the data dimension that will likely require a summary judgement, and position to the data dimension that will likely be segmented).

For visualization research, it has the potential to produce concrete guidelines for optimizing the mapping of visual features to data dimensions to support different tasks. In the following sections, we explore this grid, moving serially among its columns, providing samples of relevant research on the perceptual issues related to each task, the visualization applications that build on the task, and potential directions for future research. While by no means exhaustive, the sampling offers several potential research directions for perceptual psychologists that could also inspire more effective visualization design.

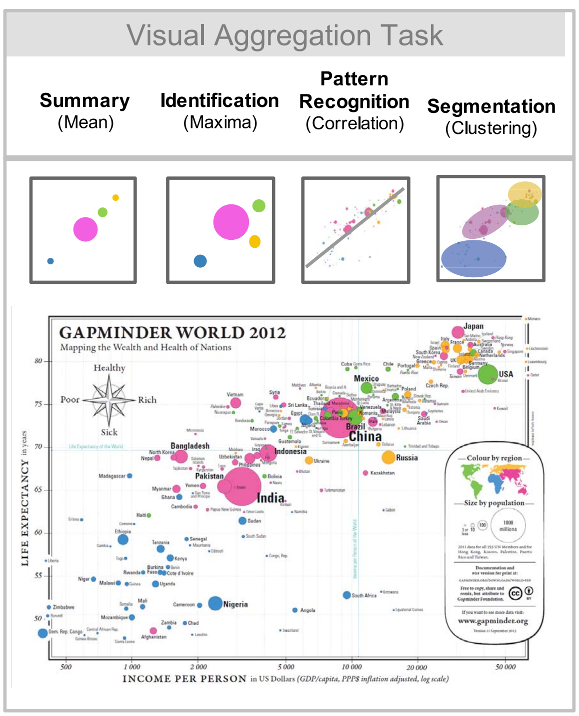
We preface our argument with some caveats. We do not intend this categorization to be a final answer, but instead the spark of a broader conversation. For example, we categorize outlier detection as an identification task, but one might also argue that it is a form of segmentation. Our discussion of work related to each task, from both perceptual psychology and data visualization, will not be exhaustive—instead, our goal is to provide a sampling of relevant work from each community. Appendix A provides a table of additional visualization references for the discussed tasks. Because of the need for brevity, some of the links that we draw may strike members of either community as problematic. For example, we mention findings in data visualization that appear inconsistent with work in perceptual psychology (see the discussion on mean position in scatterplots under Mean and variance estimation). A psychologist reading those sections might generate display constraints and confounding factors that could explain why the effect did not generalize, and might reflexively produce more precise guidelines that visualization designers could use to better predict when these effects will hold. We would be delighted by this response, as it highlights the importance of increased collaboration between visualization designers and perceptual psychologists. Finally, our review will focus on ensemble coding, but given the blurriness of its definition and the need for it to interact with other types of visual processing, we will include other related topics that are outside its strict definition, such as visual search, multifocal attention, feature-based attention, and shape recognition.



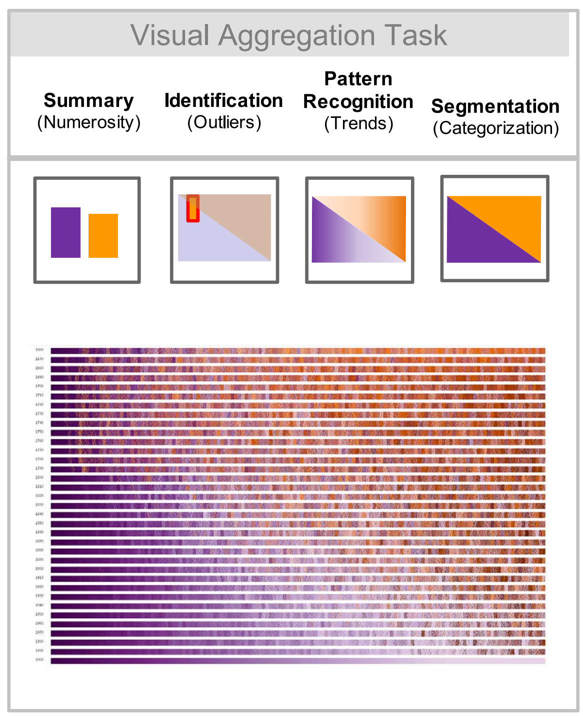
(a) Tagged text visualization uses color and position to allow users to identify linguistic patterns in a document (Alexander, Kohlman, Valenza, Witmore, & Gleicher, 2014).



(b) Weather maps use color and orientation to visualize information about wind speeds, temperatures, and other meteorological data (Ware & Plumlee, 2013).



(c) GapMinder uses size, position and color to reveal patterns in global demographics (Rosling, 2009).



(d) Inspired by work on ensemble encoding, Sequence Surveyor depicts changes in the use of 170,000 words across 34 decades, locally permuting color to help the visual system construct ensemble summaries of noisy information (Albers, Dewey, & Gleicher, 2011).

Figure 5. In these example visualization applications, data are mapped to multiple visual features, such as (a, d) color and position, (b) color and orientation, and (c) position and size, to support a variety of analysis tasks. Understanding how efficiently these features communicate different kinds of information can inspire effective visualization designs.

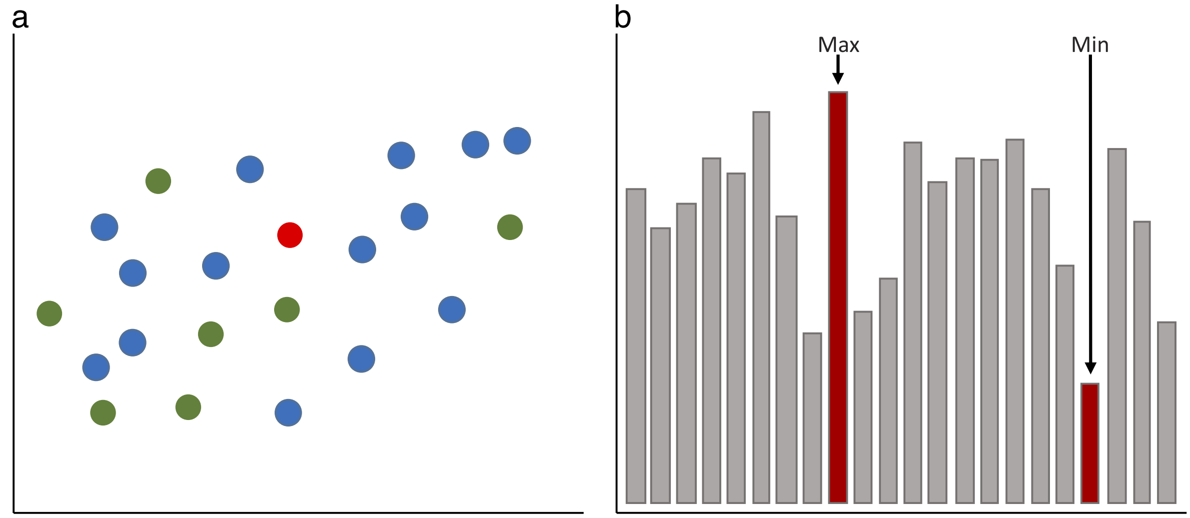


Figure 6. Identification tasks require a viewer to locate a specific set of data points, such as (a) the class of points in a scatterplot that are labeled as red or blue or (b) the minimum and maximum values that constitute the value range in a bar graph.

Identification tasks

An identification task requires a viewer to isolate a specified subset of data points (Figure 6). Often, viewers know which values they seek, in which case the task requires identifying points that match given values. In other cases, they may need to extract distributional information about a data set to identify values that are important relative to the distribution, such as locating the minima, maxima, or median. In a third kind of identification task, the viewer must actively search through the data set to detect outliers—values that are notably different from the rest of the data set.

Absolute-value identification

In absolute-value identification tasks, viewers isolate points that match a specified data value, such as the red points (e.g., Figure 6a), a circle of a given size, or all points in a certain spatial region of the display.

The constraints on a viewer's ability to locate a *single* target with a known feature—red, circular, or 2 units in diameter—have been long studied (Wolfe, 1998). As an example of one effect, finding a single target is more difficult when that target is perceptually more similar to the distractors around it, and when there is more diversity amongst the distractors (Duncan & Humphreys, 1989). This result is consistent with an established guideline for designing visualizations: Maximize the perceptual distance among the features that delineate data properties relevant to the identification task, but minimize differences between features that map properties irrelevant to the task (Wickens & Carswell, 1995).

There are also rich links between the visualization task of localizing the set of targets matching a known feature—*all* of the red, circular, or 2° objects—and feature-based attention, which allows a viewer to select multiple objects with a visual "feature filter" (e.g., Saenz, Buraĉas, & Boynton, 2003). This feature filter is known to extend broadly across a visual display, allowing the selection of large numbers of objects (or data points) at the same time (Levinthal & Franconeri, 2011), and rules have been proposed governing how multiple features can, and cannot, be logically combined (e.g., red AND square; Huang & Pashler, 2007; Nakayama, Silverman, et al., 1986). Understanding how these filters operate across multiple features could inform visualizations that support identification tasks acting across different dimensions of a data set simultaneously. For example, in Figure 5b, understanding how well viewers can combine color and orientation can help determine how effectively viewers can identify, for example, high-temperature regions where the wind blows toward the west.

When a set of points cannot be easily selected by visual features, it must instead be selected by the points' locations. Perhaps a viewer needs to select Points 4, 16, and 18 within a scatterplot because the text of their labels makes them currently relevant. Here, research on attentional selection of multiple objects (Scimeca & Franconeri, 2015) explains the ability to perform this task in visualization. As an example, there are limits to the number of locations that can be selected (up to seven or eight in total), but this limit is closer to three or four in typical displays, where objects become more tightly packed (Scimeca & Franconeri, 2014). A recent collaboration between perception and visualization researchers has shown that these selection constraints generalize to conditions similar to following points in a scatterplot through an animated transition from one plot to another (Chevalier, Dragicevic, & Franconeri, 2014).

Relative-value identification

While some absolute-value identification tasks may not require ensemble coding, relative-value identification tasks rely on distributional information about a data set in order to identify data points with a prespecified position within that distribution. Because extraction of the distribution is needed to find relative values, even traditional visual search tasks for relative values would seem to require an initial ensemble coding pass of a display before defining the target to search for. Examples of relative-value tasks include extracting the minimum or maximum value for the entire set of data (e.g., the lowest data value) or within a subset of the data (e.g., the lowest red). In Figure 5b, for example, an analyst might search for the range of wind directions in California. In Figure 5c, which two countries have the largest populations and which have the smallest? In a bar graph, the range of the data distribution might be revealed by simultaneous visual selection of both the minimum and maximum values (Figure 6b). The strategy for locating minima and maxima is unclear, though it may require ensemble coding, as both minima and maxima are defined with respect to all of the points in a collection.

There are other relative-value identification tasks that beg for study by perceptual psychologists. In Figure 6b, how well can you estimate the median value in the bar graph, and what perceptual process allows that judgment? One heuristic could be to find the range, determine the imaginary horizontal line that hovers in the midpoint of that range, and search for bars with tops near that area. That strategy works for nonskewed distributions, but fails when the data are skewed. What perceptual strategies would be more robust, what downsides would they have, and how could they be taught to graph readers? What graph designs would permit other strategies for finding the median—for example, how could your abilities change for data plotted as color values instead of the positional and length values in Figure 6b? What happens when you ask all of these questions for the modal (most frequent) value instead of the median?

Outlier identification

Outlier detection tasks are defined by the need to identify targets that are different than others in the collection. They are similar to relative-value identification tasks, except that the position of the target data points within the distribution is not specified a priori—viewers discover them while foraging through a data set, allowing saliently different data points to "pop out" (Neisser, 1964; Prinzmetal & Taylor, 2006). This set of tasks reflects one of the strongest advantages of using the visual system to compute statistics in visualization: cases where critical statistics are difficult to know (and therefore cannot be mathematically computed) a priori. As a result, outlier identification provides a number of opportunities for research in both perception and visualization.

When position is used to represent data, it is unclear what perceptual strategies allow viewers to determine when a point in a scatterplot or a bar in a bar graph might be seen as an outlier. Studies of perceptual segmentation, as discussed in Segmentation tasks, may offer insight into how a positional outlier may be identified and the role ensembles might play in detecting these values. When data are instead plotted in a featural space, such as when values are encoded with color in a heatmap, outliers might be modeled by their perceptual salience (Itti, 2005; Itti & Koch, 2001). But what process might allow a viewer to detect outliers that are not prototypical extrema, such as outliers in the middle of a widely spaced bimodal color distribution? Identifying such outliers may rely more strongly on the power of ensemble coding to extract information about the overall distribution of values.

Goals, context, tasks demands, and experience account for much of the variability in salience for natural scenes, but whether this is still true in relatively simpler data displays remains to be tested. Work on attentional control (Serences et al., 2005) and priming of features by recent experience (Chetverikov & Kristjansson, 2014) may contain important insights for visualization designers, and the context of data-visualization tasks could inspire perceptual psychologists with new questions.

All three of the tasks described in the foregoing can be performed either within an entire data set or within a specific subset of the data. This subset can be spatially defined: What is the minimum value in the left half of Figure 6b? What is the pop-out color in the first paragraph of the example shown in Figure 5a, or the upper left corner of Figure 5d? The subset can also be featurally defined: Among the red circles in Figure 5c, are there any positional outliers at the bottom of the display? In displays that simulate data visualizations, the spatially local level surprisingly does not appear to play a stronger role in computing the salience of a potential outlier. Detecting the presence of an outlier depends on an item's global uniqueness rather than local uniqueness (Haroz & Whitney, 2012), implying the use of scene-wide variance rather than only local contrast.

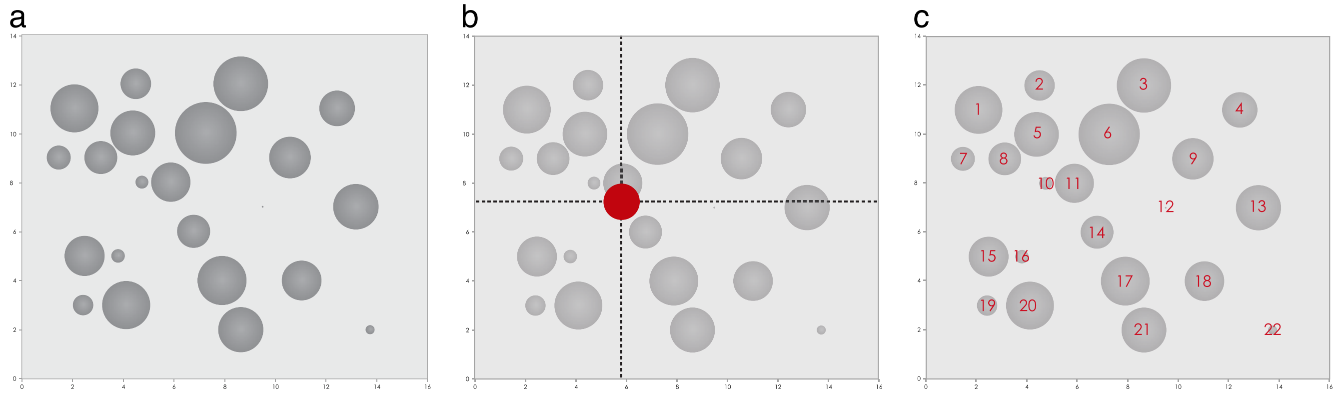


Figure 7. Summary tasks require viewers to estimate a value that summarizes a collection, such as its (b) mean and (c) numerosity.

Summary tasks

Summary tasks require the viewer to extract properties that describe the collection in aggregate. In contrast to identification tasks, which extract subsets of objects, summary tasks create representative values, such as descriptive statistical measures. For example, a viewer may estimate the average height of a bar in a bar chart or the average position of points in a scatterplot (Figure 7b). While some summaries might overlap with value extraction, as in extracting a median, most summaries are not values from the set.

Mean and variance estimation

Estimating mean and variance are common summary tasks in visualization. Figure 8a depicts monthly stock prices as individual line graphs, with data from one company colored red and data from the other colored blue. An analyst could estimate the mean orientation of the red and blue lines to compare how monthly stock prices change on average between the two companies, and the orientation variance to compare the stability of stock values. Means can be efficiently computed for several visual features, including size (Ariely, 2001; Chong & Treisman, 2003, 2005a, 2005b; Fouriezos, Rubenfeld, & Capstick, 2008), orientation (Alvarez & Oliva, 2008; Bulakowski, Bressler, & Whitney, 2007; Choo, Levinthal, & Franconeri, 2012; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), motion speed (Watamaniuk & Duchon, 1992) and direction (Watamaniuk, Sekuler, & Williams, 1989), brightness (Bauer, 2010), color (Webster, Kay, & Webster, 2014), and position (Hess & Holliday, 1992; Melcher & Kowler, 1999; Morgan & Glennerster, 1991; Whitaker, McGraw, Pacey, & Barrett, 1996). Variance among values can be efficiently computed for orientation (Morgan, Chubb, & Solomon, 2008). Our ability to compute the mean of a collection is surprisingly robust in the face of other types of variability across collections, for irrelevant dimensions like spatial frequency (Oliva & Torralba, 2006), density (Chong & Treisman, 2005b; Dakin, 2001), numerosity (Chong & Treisman, 2005b; Dakin, 2001), temporal sequence (Chong & Treisman, 2005a), and distributional variance (Dakin, 2001).

Figure 7a provides a sample visualization from which mean size and position can be rapidly extracted (see the red circle in Figure 7b). Figure 5c depicts a more complex visualization, in which mean size

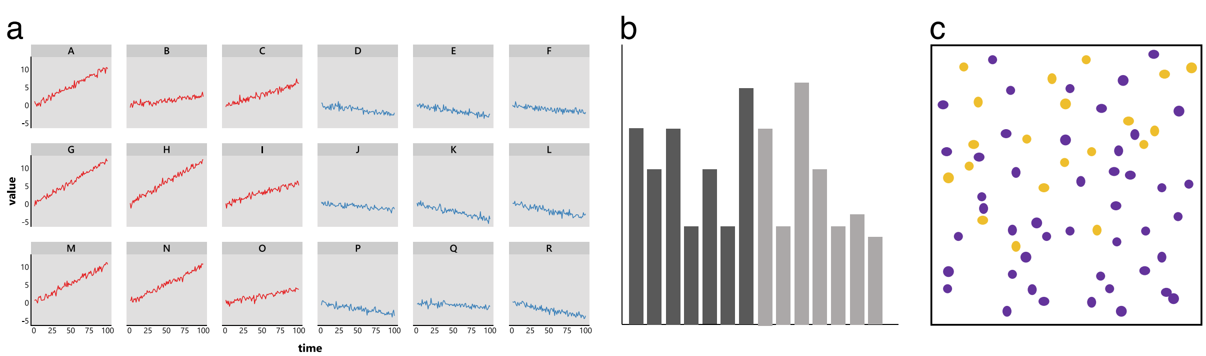


Figure 8. Estimating mean, skew, and numerosity are three types of summary tasks common to data visualization.

provides insight into demographic data. To compare average population size across different geographic regions, for example, you could identify circles of different colors and average the size of the resulting set. You could use a similar process to identify the average population size of low-income countries by spatially grouping the objects within the left third of the x-axis and computing the average size of the resulting groups, allowing you to note that they tend to be small on average, with low size variability.

Recent work has tested the ability of viewers to estimate the mean value of collections within data visualizations. One study tested how well viewers could compare the mean position of two groups of points in a scatterplot (Gleicher et al., 2013), focusing on how the colors and shapes used to mark different data sets affected viewers' ability to compare their mean heights. Some results were consistent with intuitions from perceptual psychology. For example, using color to distinguish the two groups (making points orange vs. purple) led to higher accuracy in mean judgments compared to using shape (making points circles vs. triangles). But adding more diversity among the distractor classes (e.g., adding green objects to the orange and purple display) did not impair performance for comparing mean position between the two groups, as would be expected from previous work on visual search (Duncan & Humphreys, 1989). Adding more perceptual spacing between classes by combining cues (orange and circular vs. purple and triangular) surprisingly did not improve performance, contrary to findings from previous work on visual search (Duncan & Humphreys, 1989). Note that the underlying mechanism for such mean position judgments may or may not be an ensemble one, depending on your definition of *ensemble*. If the horizontal and vertical positions are truly averaged in the same manner as other dimensions such as size or luminance, then the definition fits. But if the center is computed by shape-recognition heuristics that focus on a low spatial frequency envelope (Harrison et al., 2014), then whether that counts as ensemble coding depends on your definition.

Another set of studies tested how well viewers can estimate the mean and variance from visualizations of time series data (Albers et al., 2014). These studies showed trade-offs between how accurately viewers can estimate mean and variance (summary tasks) versus range and extrema (identification tasks) from data visualized using either color or position. While each statistic could be extracted from both visual features, there was a salient difference between the types of tasks best supported by each: Mean and variance were more accurately extracted from data encoded using color, whereas extrema and range were more accurately extracted from positional visualizations. These results suggest different processing abilities for color and position—color may facilitate summation of values at low spatial frequencies into a representation similar to a color histogram, while position may better represent shape-boundary properties. At the same time, the results show a trade-off for visualization design—color better supports summary tasks, while position better supports identification tasks.

People can also estimate the mean of a set of orientations (Parkes et al., 2001). Not all types of orientation are averaged with the same precision: The average orientation of the boundary contours of a set of objects can be more precisely extracted than the average orientation of their internal textures (Choo et al., 2012). In visualizations like the weather map shown in Figure 5b, this predicts an improved ability to summarize wind directions in maps that use oriented glyphs over maps that use oriented textures. Here, extracting a mean orientation across local regions has clear utility for understanding how the general wind direction in cold regions (purple) differs from the wind direction in warm regions (red). Ensemble coding of orientation is also useful in the stock market visualization in Figure 8a. How accurately could a financial analyst determine whether the red company shows more variable performance than the other? While the Weber fraction for variance—the point where differences in variance become indistinguishable—has been studied for orientation (Morgan et al., 2008), there are far fewer studies of variance perception, compared to studies of perceptions of average value.

Distribution statistics

We can extract mean and variance from a collection of data points, but what about other aspects of a data set's distribution? For example, visualizations may reveal the skew and kurtosis of a data set to a viewer. The bar chart depicted in Figure 8b contains two data sets that differ in skew. In more complex visualizations of oceanographic data, skew and kurtosis of sea-surface height are important for making predictions about the movement and position of eddies for applications in oil exploration, where eddies can damage offshore drilling equipment (Hollt et al., 2014). Little is understood about the accuracy or biases of our perception of these high-order statistics. One possibility is that the visual system does not encode skew per se but may approximate it after extracting a collection's centroid (Dakin & Watt, 1997). Better insight into how these statistics can be inferred by the visual system can help in designing visualizations to support a broader variety of statistical analyses on raw data.

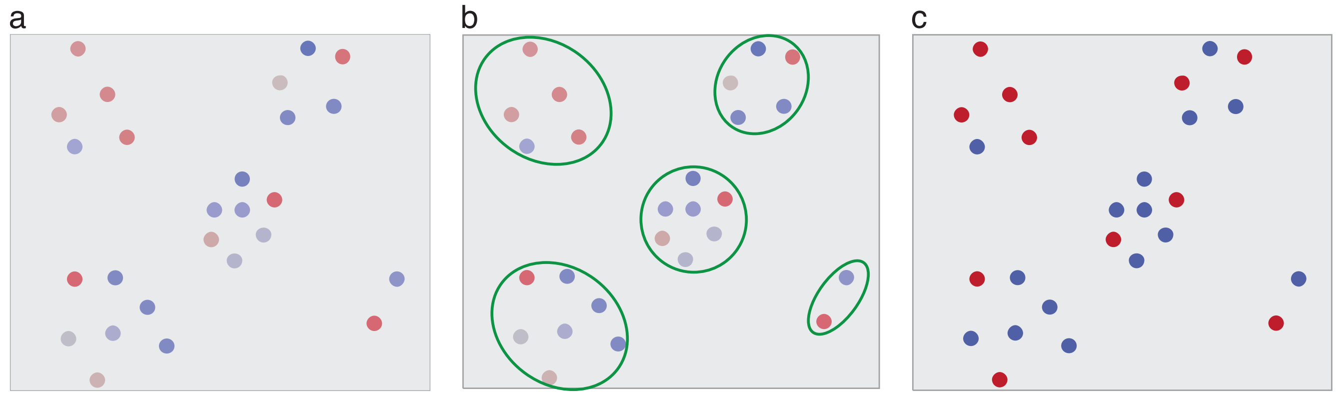


Figure 9. Segmentation tasks, such as (b) spatially or (c) featurally clustering data elements, require the viewer to visually segment the data set into discrete clusters.

Numerosity

Visualizations regularly require viewers to judge the numerosity of a set of data points (Figure 7c). This judgment might be an estimate of an absolute number of data points—how many bubbles are in Figure 7a?—or a comparison between two or more values—are there more purple or orange points in the scatterplot in Figure 8c? Numerosity estimation is surprisingly robust in the face of variability in other dimensions, such as contrast, orientation, and density (Burr & Ross, 2008). These findings align with results from visualization showing that estimation of relative numerosity is robust across color and orientation (Healey, Booth, & Enns, 1996).

However, changes in some featural dimensions, such as luminance, can bias numerosity estimates. Darker collections can appear more numerous (Ross & Burr, 2010), so a visualization designer should be careful when using luminance to differentiate collections of data whose relative numerosity is relevant to the viewer. Grouping objects using visual connection can cause viewers to underestimate the number of original parts (Franconeri, Bemis, & Alvarez, 2009). A visualization designer working with network data (typically visualized as points connected by lines) should be wary of the effect of these connections on number perception. The number of possible simultaneous numerosity estimates may also be limited (Halberda, Sires, & Feigenson, 2006), which implies that the number of visually distinct categories in a visualization should be limited if simultaneous numerosity estimation is critical.

Understanding how the size of an item influences perceived numerosity could help reduce bias when size and quantity visualize independent dimensions. Correll, Alexander, and Gleicher (2013) found that in comparing the quantities of red and blue words in displays like Figure 5a, longer words biased viewers toward perceiving a higher quantity. These results led to a new visualization approach that helps account for this bias: Increasing the spacing between letters in short words increases the overall length of the colored word, and improves numerosity estimation in text displays.

Segmentation tasks

Segmentation tasks require viewers to organize data points into subsets (Figure 9). Unlike identification tasks that isolate data points that adhere to specific constraints, these subsets are formed based on their similarity within some visual dimension, typically either space (position) or a feature dimension (e.g., color or orientation). Ensemble processes might guide these segmentation tasks by providing distributional information that allows detection of salient spatial or featural clusters. For example, if luminance values formed a bimodal distribution, with one light mode and one dark mode, it could signal two corresponding clusters of points.

Segmentation by spatial position

Segmentation is perhaps most intuitive when data are mapped to spatial position. Viewing Figure 9a, it is apparent that there are five primary spatial groups (Figure 9b). Spatial segmentation helps viewers quickly form meaningful subsets of related items within a data set. For example, in Figure 5c we see multiple spatial clusters that identify countries with similar demographic properties: a tight cluster of countries that share both high GDP per capita and high life expectancy in the upper right corner of the visualization, a looser cluster in the center with intermediate

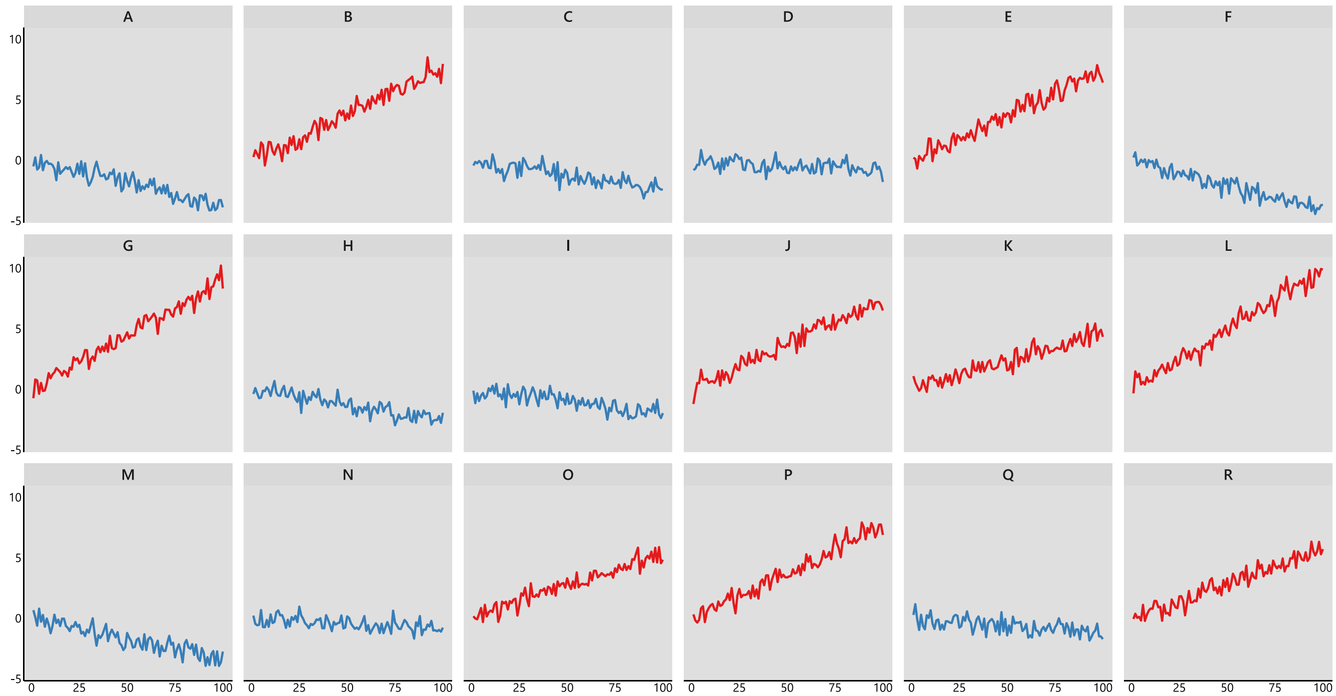


Figure 10. Segmentation tasks, such as clustering, can be accomplished not only with positional mappings but also by featural mappings like color or orientation.

GDP per capita and life expectancy, and a scattered group of countries in the lower left with low GDP per capita and life expectancy. The importance of this segmentation process demands several explanations from perceptual psychologists, such as what counts as a cluster in the visual field, how many clusters can be created, and what might bias this segmentation process.

Understanding visual grouping may be particularly useful in answering some of these questions. For example, spatial clustering should be largely based on the gestalt grouping cue of proximity, and studies of proximity grouping suggest that it is a parallel and mandatory cue (Rock & Palmer, 1990). It also tends to dominate over other grouping cues, such as color (Oyama, 1961). While several clusters can be constructed simultaneously across a display, performing additional operations on these clusters, such as extracting the shape of each collection, can be a slow, or even strictly serial, operation (Trick & Enns, 1997).

An understanding of segmentation in data visualizations will also require an understanding of visual crowding. Items that are identifiable on their own can become indistinguishable or crowded when surrounding objects are too close (Whitney & Levi, 2011), and this problem worsens in the periphery (Pelli, Palomares, & Majaj, 2004). Crowding limits could contribute to the limit on the number of clusters that can be created in a visual display (Franconeri, Alvarez, & Cavanagh, 2013).

Segmentation by features

A viewer can also cluster data points using featural similarity. Figure 9c depicts an alternative segmentation of the display in Figure 9a, using color value instead of spatial position. Figure 10 depicts an example where data can be clustered by orientation and color, ignoring their locations. Figure 5 depicts clustering by color and orientation in more complex displays: Data in all four visualizations can be clustered according to color values, data in Figure 5b can be clustered by orientation, and data in Figure 5c can be clustered by size. These types of segmentation operations might rely heavily on feature distributions generated by ensemble coding. If a distribution is bimodal or multimodal, that information could underlie the viewer's understanding that a display contains two or more dominant values (Utochkin, 2015)—which the viewer might inspect sequentially by attending to each value in turn.

In some cases, using multiple visual features within a single visualization may make feature clusters harder to see in either feature dimension alone. For example, a visualization might use color hue to encode one property of a data point and luminance to encode a second property. Luminance variation across these points might inhibit viewers' abilities to segment points that have similar hues. It may likewise be difficult to segment different points of different luminance levels if their hues are vastly different (Callaghan, 1984).

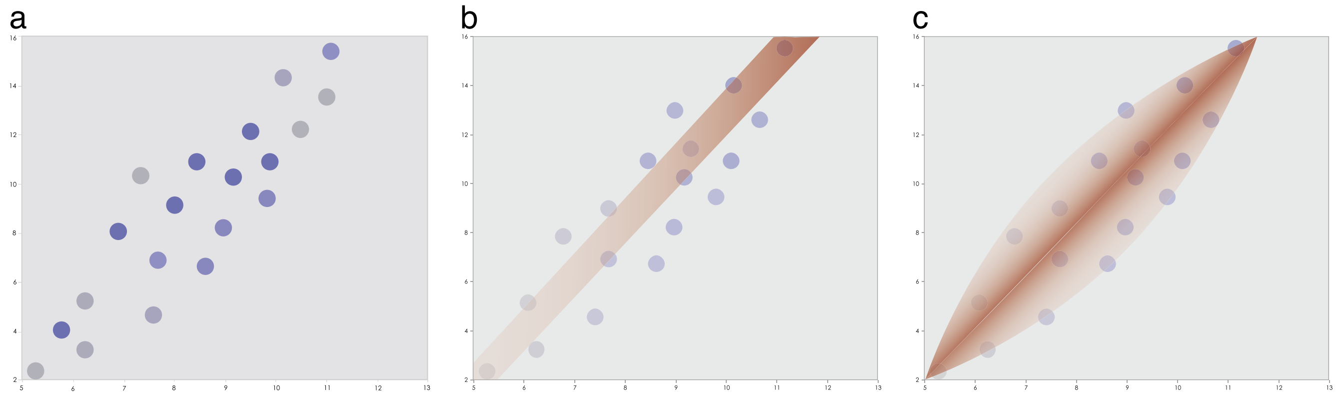


Figure 11. Structure estimation tasks extract patterns from collection of values, such as (a) trends and (b) correlation. In these scatterplots, these tasks can be computed across both position and color.

However, some features may support more robust segmentation. For example, if both color hue and shape are used to encode different properties of a data set, segmenting points based on shape is likely to be more challenging for shapes of different hues, whereas viewers can segment points of different colors regardless of shape (Callaghan, 1989). Understanding the role of ensemble coding in segmentation may offer guidance for which features (or combinations thereof) can help viewers better identify divisions in visualized data.

Structure estimation tasks

Structure estimation tasks require viewers to extract patterns from sets of data points that are not always intuitively captured by single statistics (Figure 11). Anscombe's quartet (Figure 2) illustrates the importance of structure estimation tasks: The four data sets are identical across several statistics, yet have qualitatively different patterns. These patterns often require visualizations for a viewer to understand them.

Trend detection

Detecting trends—the qualitative relationship between two variables—is perhaps the most ubiquitous form of structure estimation. The reader is likely most accustomed to trends between two variables mapped to position on a Cartesian grid, as in a scatterplot (Figure 11b). The trend that as x increases, y increases is immediately apparent and appears linear, as opposed to curved or U-shaped. The visual system is adept at comparing the relative strength of linear correlations (Figure 11c; Harrison et al., 2014; Rensink & Baldridge, 2010). While the most common (and likely most powerful) visual mappings for representing the trend between two variables pair two spatial dimensions, feature-based depictions are also common when both spatial dimensions have already been mapped to other aspects of the data. For example, the bubble chart in Figure 12a may not contain an x-y trend, but size clearly increases with the x value. Figure 12b depicts a trend that relies on neither spatial axis, but it is clear that as size increases, luminance decreases.

More complex examples of such trends between a positional dimension and a featural dimension are depicted in Figure 5a and d—certain colors occupy certain spatial positions in each of these displays. In Figure 5a, this trend reflects that words common in novels (red) are more frequent at the beginning of the passage, whereas words associated with philosophical discussions (blue) are more frequent at the end (Alexander, Kohlmann, Valenza, Witmore, & Gleicher, 2014). Figure 5d visualizes the 2,000 most popular works per decade over the last 350 years, with each decade represented by a row and word popularity mapped to the x-axis. The figure shows that words that are popular in modern writing (purple) have slowly replaced those that were popular in earlier texts (orange; Albers, Dewey, & Gleicher, 2011).

When trends are depicted across two spatial axes, we have some idea of how they might be detected. For example, simple shape-recognition networks might classify whether a cloud of points matches an oval, a curve, or a U-shaped object (e.g., Field, Hayes, & Hess, 1993; Uttal & Tucker, 1977), though some data suggest that such simple tricks may not be sufficient to explain performance, at least for certain kinds of trends (Rensink, 2014). The ability to identify these shapes within visualized data may help reveal complex relationships between variables. For example, mapping two variables to position will form a line if they are highly correlated or a parabola if there is a quadratic relationship between them. But when at least one feature dimension is involved, our understanding of how such patterns are recognized or encoded shrinks

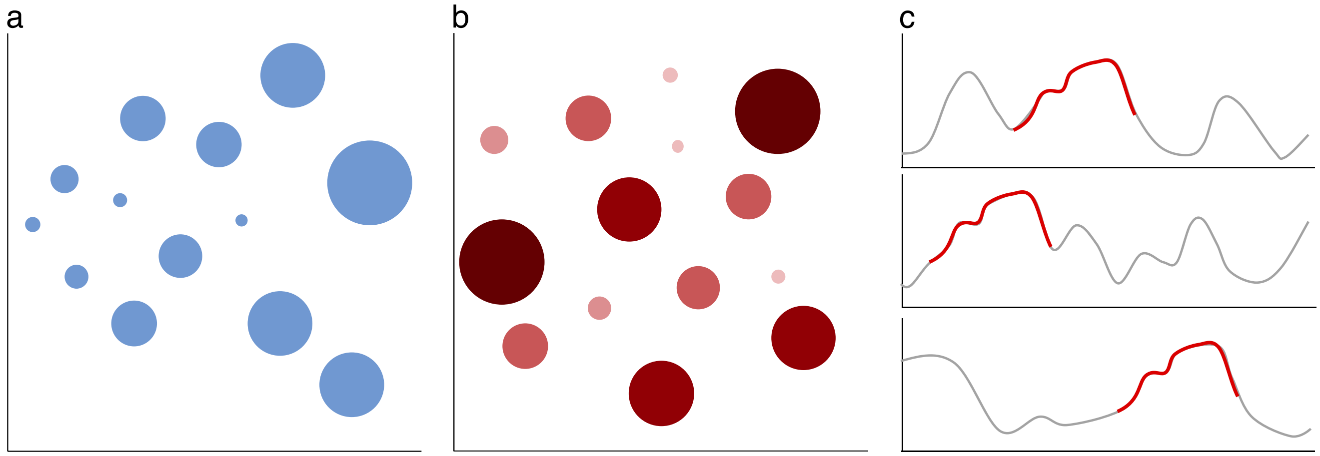


Figure 12. Trend and motif detection are two examples of structure-recognition tasks in visualization.

drastically. There is evidence that the visual system is capable of detecting featural correlation, sometimes as effectively as with positional correlation (Rensink, 2014), and that correlation may be estimated with similar accuracy across visualization designs using several different kinds of features (Harrison et al., 2014). However, it is not clear whether this ability generalizes to other forms of trend detection, such as characterizing nonlinear relationships between features.

One possible strategy for characterizing featural trends is that a form of ensemble coding extracts one or two feature distributions from the display and they are compared. While there is some evidence that this cross-feature pattern detection may occur for orientation and size (Oliva & Torralba, 2006), the mechanism for this detection is unclear, as is whether it works for other feature combinations. We see this problem as a fertile one for perceptual psychologists to explore.

An innovative set of proposals suggest a more mechanistically precise alternative for detecting featural trends: that trends involving at least one feature dimension are processed by serial selection of certain feature values at a time (Huang & Pashler, 2002, 2007). For example, extracting a trend between luminance and size in a bubble chart might involve selecting dark and then light items, approximating the mean size at each of these luminance levels, and storing each mean size in memory for later comparison. Note that this method of structure estimation would be more similar to the segmentation operations in the previous section than the other operations in this section. There is evidence for this type of serial processing in some kinds of visual comparison (Huang & Pashler, 2002) and visual grouping (Huang & Pashler, 2007; Levinthal & Franconeri, 2011), but there is a need for empirical work demonstrating that such a model could explain trend detection among, for example, the graphs shown in Figure 12.

Similarity detection

Another aspect of structure estimation is determining similarity across different regions in a data set, such as inferring how similar wind currents are across different geographic or temperature regions (Figure 5b). Similarity detection occurs at different scales: from the more holistic task of estimating the similarity of the shape of two line graphs to the more local task of detecting small repeated patterns across two line graphs (Figure 12c). Detecting repeated structures, commonly called *motifs*, across a data set is important in applications such as biology, where these patterns often indicate blocks of genetic material with important biological functions that are conserved across different organisms (Albers et al., 2011; Meyer, Munzner, & Pfister, 2009), or in time-series data, where they often represent related events (Lin, Keogh, Lonardi, Lankford, & Nystrom, 2004). Important motifs may appear among noise or other distortions within a data set or may be inverted in order. For example, in visualizing energy usage over time, an event (e.g., turning on a device) may cause a drastic increase in energy usage. This motif indicates the event occurrence, and the motif's inversion may occur when the event ends (e.g., the device is turned off). How might the visual system detect similarity between different collections in a visualization? How might it find small-scale repeated patterns that form motifs, and how do noise and inversion influence our ability to identify these patterns? How might the efficiency with which we determine similarity change for different visual features?

Ensemble coding may be an important part of computing similarity between visualized collections. For colocated objects, the visual system might compute variance in a region (Morgan et al., 2008), and regions of low variance indicate high local similarity. The visual

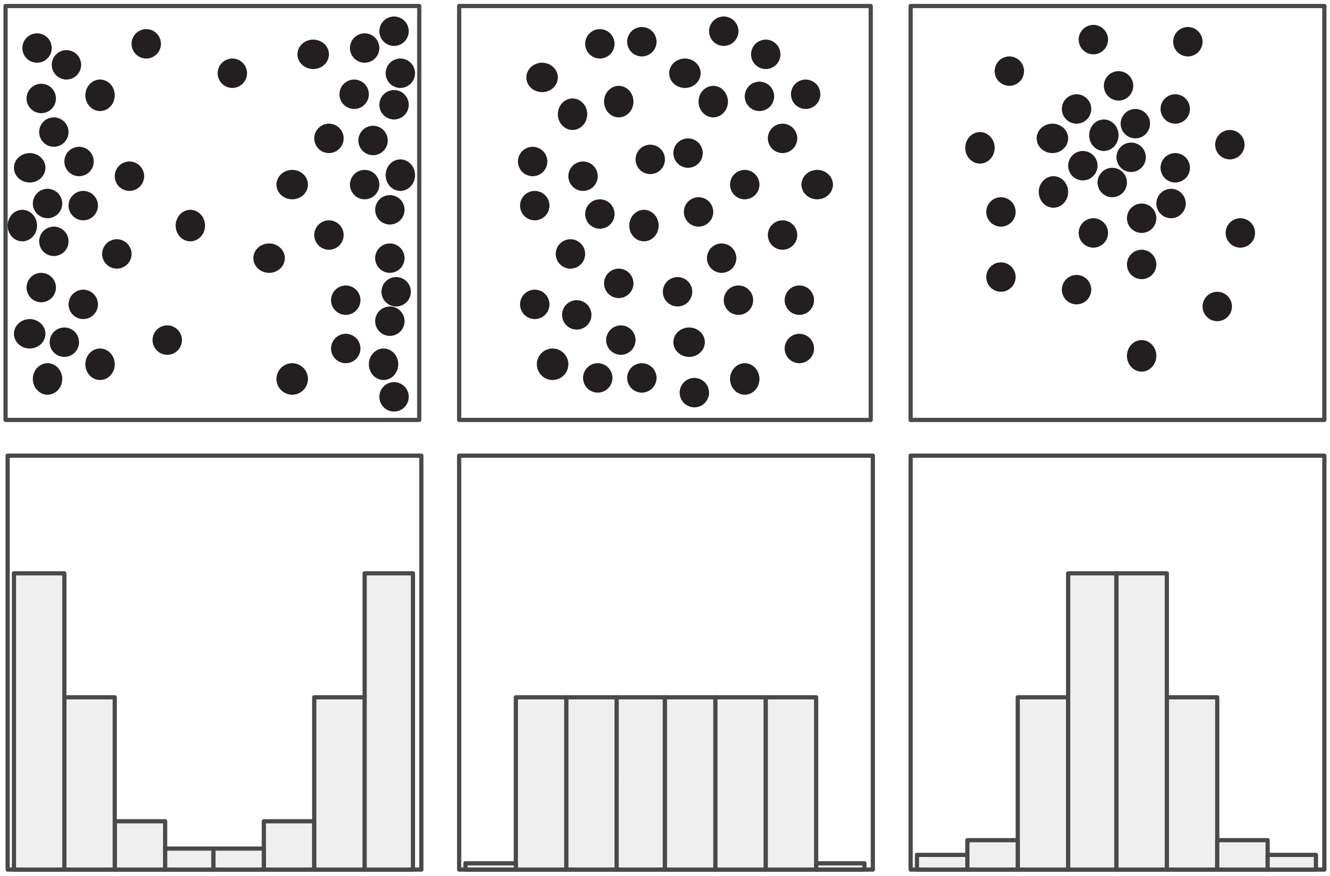


Figure 13. Visualizations may communicate important information about the distribution of values within a data set, beyond simply mean and variance.

system might identify motifs by detecting small regions with similar ensemble statistics as a viewer scans a display. A potential strategy for estimating similarity across different clusters, such as red and blue points in a scatterplot, might involve computing ensembles within spatial or featural clusters (Corbett & Melcher, 2014) and then comparing those statistics between clusters (Dakin, 2014). This strategy relies on comparing ensembles as opposed to detailed patterns to estimate similarity across different subsets of data and correlates well with how viewers perceive similarity between pieces of artwork, another type of complex visual scene—here, perceived similarity correlates with comparisons between mean luminances of corresponding spatial regions in a painting (Graham, Friedenberg, McCandless, & Rockmore, 2010).

Distribution shape

Understanding the shape of a data distribution is important for a number of statistical inference tasks. In understanding demographics data, a viewer may wish to characterize the distribution of wealth across a population. Alternatively, the viewer may wish to compare different attributes of a population, such as the distribution of age versus that of income.

Prior work provides evidence of an interaction between ensemble coding processes and properties of featural distributions, such as smoothness and range of variance (Utochkin & Tiurina, 2014). However, little is known about how well the visual system might infer whether a distribution is uniform, Gaussian, multimodal, or some other classification. For positional representations, ensemble coding seems to allow a viewer to readily perceive the mean and variability of a collection, but other aspects of the distribution might be available as well (Figure 13). How well—if at all—can the visual system perceive distribution shape? And how might encoding data with different visual features, such as color or luminance, affect that ability?

Conclusions

Data visualizations allow us to explore and analyze data using our visual system by mapping data values to spatial positions and visual features. Because viewers can use ensemble coding to efficiently extract statistical information from a data set, a better understanding of these ensemble mechanisms could provide design guidelines for data displays that maximize a viewer's ability to process data visually. The fact that many of these guidelines have yet to be firmly established—what types of visual features and displays facilitate what kinds of visual statistical decisions—reveals unsolved questions for perceptual psychologists. These factors make the study of ensemble coding of data visualizations a fertile territory for collaboration between the perception and visualization communities.

To help organize the discussion at this interface, we have introduced a task categorization and surveyed both past work and open problems for each category, across perception and visualization. The four categories of tasks are ubiquitous in data visualization: identification, summarization, segmentation, and structure estimation. A single example can clearly illustrate the importance of each of these tasks: tagged text visualization (Figure 5a). You can quickly *identify* outlier text tags in blue in the first paragraph. You can *summarize* that there are two to three dozen red tags in total. You can *segment* the major division between red tags in the top paragraph and blue in the bottom. And you *estimate structure* in the data to detect a red-to-blue trend that is systematically related to vertical position in the text.

In addition to ensemble coding, research on the control of visual attention is relevant to all four task categories. In the examples that we have explored, we assume that viewers have perfect control over which subset of data they operate on—the data points on top, the red objects, the triangles, and so on. In reality, visually selecting relevant data points can be difficult or noisy. For example, data displays often contain animations, motion, or transients that can distract the viewer (Bartram, Ware, & Calvert, 2003; Hollingworth, Simons, & Franconeri, 2010) and may impair selective ensemble coding of only the relevant visual features. Attentional control can be especially difficult when multiple dimensions of data are depicted simultaneously. For example, a bar graph might map sales to color, profits to height, and time to horizontal order. This visualization would present multiple dimensions of information via multiple visual features simultaneously. Work on attentional control shows that when there is simultaneous variability in multiple feature dimensions, the "wrong" dimension can distract the viewer (Lustig & Beck, 2012). Some existing work explores attentional control in the context of data visualization (for a survey, see Healey & Enns, 2012), but many questions of interest to both communities remain. Both perceptual psychology and data visualization would benefit from a better understanding of whether our current conclusions about attentional control, which draw from a set of laboratory tasks, apply to the more complex displays found in data visualization. Collaboration with data visualization researchers brings this benefit to perceptual psychologists more generally, as a way of testing whether knowledge gained from simplified displays and tasks is robust across new contexts.

Our categorization of tasks and links to relevant work in both communities are by no means intended to be exhaustive—and will not be the last word. Instead, our goal is to foster conversation between these communities around ensemble phenomena. We find three themes particularly exciting.

First, we assume that the collection of visual processing abilities that we call ensemble coding—processing of information that can be extracted and combined in parallel from large numbers of objects at once—evolved and developed to compute statistics in the natural world. Those statistics are likely to be based on heuristics and other "good enough" strategies that suffice for the natural world, but we know that some ensemble judgments introduce biases in statistical inferences from data displays that would not be present in formal computed statistics (Sweeny, Haroz, & Whitney, 2012). How common are those biases, which are potentially problematic among different visual features, and how can data displays be designed to avoid them?

Second, there is research in data visualization that explores which dimensions allow the most precise extraction of individual values (e.g., Cleveland & McGill, 1984). These studies have found that spatial position is precise, object length is not quite as good, angular extent is bad, and color saturation is among the worst methods for precisely representing individual data values. But these rankings change for ensemble coding. For example, color depictions can beat spatial-position depictions when a viewer needs to analyze average values from a subset of raw data (Albers et al., 2014). While inconsistent at first glance, we believe that this contrast may inspire a new requirement for dimensions that lead to efficient ensemble coding—they may actually have to be *imprecisely* coded, so that their distributions tend to overlap, leading to better representation of distributions as a whole. In contrast, dimensions that are precisely coded may be tougher to combine, because representations of values tend to remain individuated (Franconeri et al., 2013).

Third, we believe that collaborative work between these communities will help perceptual-psychology researchers define the set of operations that are possible via ensemble coding. Currently, judgment of average value is the dominant task given to participants in these studies. We hope that we have shown that other judgments, such as range, median, skew, modality, and correlation (e.g., Rensink & Baldridge, 2010), would provide excellent testing grounds for exploring whether these values are extracted via ensemble coding or by combining ensemble codings with other visual strategies.

Keywords: ensemble encoding, data visualization, visual search

Acknowledgments

We thank Todd Horowitz for helpful advice. We also thank Liqiang Huang and Ron Rensink for their thoughtful comments. This work was funded by NSF awards IIS-1162037 and IIS-1162067 and NSF BCS-1056730 CAREER.

Commercial relationships: none.
Corresponding author: Danielle Albers Szafir.
Email: danielle.szafir@colorado.edu.
Address: Department of Information Science, University of Colorado Boulder, Boulder, CO, USA.

References

Albers, D., Correll, M., & Gleicher, M. (2014). Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 551–560). New York: ACM.

Albers, D., Dewey, C., & Gleicher, M. (2011). Sequence surveyor: Leveraging overview for scalable genomic alignment visualization. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2392–2401.

Alexander, E., Kohlmann, J., Valenza, R., Witmore, M., & Gleicher, M. (2014). Serendip: Topic model-driven visual exploration of text corpora. In *2014 IEEE conference on visual analytics science and technology* (pp. 173–182). New York: ACM.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398.

Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. In *IEEE symposium on information visualization* (pp. 111–117). Washington, DC: IEEE Computer Society.

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

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.

Bartram, L., Ware, C., & Calvert, T. (2003). Moticons: Detection, distraction and task. *International Journal of Human-Computer Studies, 58*(5), 515–545.

Bauer, B. (2010). Does Stevens' power law for brightness extend to perceptual brightness averaging? *The Psychological Record, 59*(2), 171–185.

Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.

Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2376–2385.

Bulakowski, P. F., Bressler, D. W., & Whitney, D. (2007). Shared attentional resources for global and local motion processing. *Journal of Vision, 7*(10):10, 1–10, doi:10.1167/7.10.10. [PubMed] [Article]

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology, 18*(6), 425–428.

Callaghan, T. C. (1984). Dimensional interaction of hue and brightness in preattentive field segregation. *Perception & Psychophysics, 36*(1), 25–34.

Callaghan, T. C. (1989). Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception & Psychophysics, 46*(4), 299–311.

Chetverikov, A., & Kristjansson, Á. (2014). History effects in visual search for monsters: Search times, choice biases, and liking. *Attention, Perception, & Psychophysics, 77*(2), 402–412.

Chevalier, F., Dragicevic, P., & Franconeri, S. (2014). The not-so-staggering effect of staggered animated transitions on visual tracking. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 2241–2250.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404.

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67*(1), 1–13.

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*(7), 891–900.

Choo, H., Levinthal, B. R., & Franconeri, S. L. (2012). Average orientation is more accessible through object boundaries than surface features. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 585–588.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*(387), 531–554.

Corbett, J. E., & Melcher, D. (2014). Characterizing ensemble statistics: Mean size is represented across multiple frames of reference. *Attention, Perception, & Psychophysics, 76*(3), 746–758.

Correll, M., Alexander, E. C., & Gleicher, M. (2013). Quantity estimation in visualizations of tagged text. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2697–2706). New York: ACM.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A, 18*(5), 1016–1026.

Dakin, S. (2014). Seeing statistical regularities: Texture and pattern perception. In J. Wagemans (Ed.), *Handbook of perceptual organization* (pp. 150–166). New York: Oxford University Press.

Dakin, S., & Watt, R. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*(22), 3181–3192.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local association field. *Vision Research, 33*(2), 173–193.

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics, 70*(3), 456–464.

Franconeri, S. L. (2013). The nature and status of visual resources. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 147–162). New York: Oxford University Press.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences, 17*(3), 134–141.

Franconeri, S. L., Bemis, D., & Alvarez, G. (2009). Number estimation relies on a set of segmented objects. *Cognition, 113*(1), 1–13.

Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. E. (2012). Flexible visual processing of spatial relationships. *Cognition, 122*(2), 210–227.

Gleicher, M., Correll, M., Nothelfer, C., & Franconeri, S. (2013). Perception of average value in multiclass scatterplots. *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2316–2325.

Graham, D. J., Friedenberg, J. D., McCandless, C. H., & Rockmore, D. N. (2010). Preference for art: Similarity, statistics, and selling price. In *IS&T/SPIE electronic imaging* (pp. 75271A). Bellingham, WA: SPIE.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751–R753.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17*(7), 572–576.

Haroz, S., Kosara, R., & Franconeri, S. L. (2015). Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1191–1200). New York: ACM.

Haroz, S., & Whitney, D. (2012). How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics, 18*(12), 2402–2410.

Harrison, L., Yang, F., Franconeri, S., & Chang, R. (2014). Ranking visualizations of correlation using Weber's law. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1943–1952.

Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction, 3*(2), 107–135.

Healey, C. G., & Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics, 18*(7), 1170–1188.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM, 53*(6), 59–67.

Heer, J., Kong, N., & Agrawala, M. (2009). Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1303–1312). New York: ACM.

Hess, R. F., & Holliday, I. E. (1992). The coding of spatial position by the human visual system: Effects of spatial scale and contrast. *Vision Research, 32*(6), 1085–1097.

Hollingworth, A., Simons, D. J., & Franconeri, S. L. (2010). New objects do not capture attention without a sensory transient. *Attention, Perception, & Psychophysics, 72*(5), 1298–1310.

Hollt, T., Magdy, A., Zhan, P., Chen, G., Gopalakrishnan, G., Hoteit, I., & Hadwiger, M. (2014). Ovis: A framework for visual analysis of ocean forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics, 20*(8), 1114–1126.

Huang, L., & Pashler, H. (2002). Symmetry detection and visual attention: A binary-map hypothesis. *Vision Research, 42*(11), 1421–1430.

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review, 114*(3), 599–631.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093–1123.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203.

Javed, W., McDonnel, B., & Elmqvist, N. (2010). Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 927–934.

Levinthal, B. R., & Franconeri, S. L. (2011). Common-fate grouping as feature selection. *Psychological Science, 22*(9), 1132–1137.

Lin, J., Keogh, E., Lonardi, S., Lankford, J. P., & Nystrom, D. M. (2004). Visually mining and monitoring massive time series. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 460–469). New York: ACM.

Lustig, A. G., & Beck, D. M. (2012). Task-relevant and task-irrelevant dimensions are modulated independently at a task-irrelevant location. *Journal of Cognitive Neuroscience, 24*(9), 1884–1895.

Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research, 39*(17), 2929–2946.

Meyer, M., Munzner, T., & Pfister, H. (2009). MizBee: A multiscale synteny browser. *IEEE Transactions on Visualization and Computer Graphics, 15*(6), 897–904.

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A "dipper" function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11):9, 1–8, doi:10.1167/8.11.9. [PubMed] [Article]

Morgan, M., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision Research, 31*(12), 2075–2083.

Nakayama, K., & Silverman, G. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320*(6059), 264–265.

Neisser, U. (1964). Visual search. *Scientific American, 201*(6), 94–102.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

Oyama, T. (1961). Perceptual grouping as a function of proximity. *Perceptual and Motor Skills, 13*(3), 305–306.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12):12, 1160–1169, doi:10.1167/4.12.12. [PubMed] [Article]

Prinzmetal, W., & Taylor, N. (2006). Color singleton pop-out does not always poop out: An alternative to visual search. *Psychonomic Bulletin & Review, 13*(4), 576–580.

Rensink, R. A. (2014). On the prospects for a science of visualization. In W. Huang (Ed.), *Handbook of human centric visualization* (pp. 147–175). New York: Springer.

Rensink, R. A., & Baldridge, G. (2010). The perception of correlation in scatterplots. *Computer Graphics Forum, 29*(3), 1203–1210.

Rock, I., & Palmer, S. (1990). Gestalt psychology. *Scientific American, 263*, 84–90.

Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision, 10*(2):10, 1–8, doi:10.1167/10.2.10. [PubMed] [Article]

Roth, R. E. (2012). Cartographic interaction primitives: Framework and synthesis. *The Cartographic Journal, 49*(4), 376–395.

Saenz, M., Buraĉas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research, 43*(6), 629–637.

Schulz, H.-J., Nocke, T., Heitzler, M., & Schumann, H. (2013). A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics, 19*(12), 2366–2375.

Scimeca, J. M., & Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(2), 109–118.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science, 16*(2), 114–122.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings, IEEE symposium on visual languages* (pp. 336–343). Boulder, CO: IEEE.

Sweeny, T. D., Haroz, S., & Whitney, D. (2012). Reference repulsion in the categorical perception of biological motion. *Vision Research, 64*, 26–34.

Trick, L. M., & Enns, J. T. (1997). Clusters precede shapes in perceptual organization. *Psychological Science, 8*(2), 124–129.

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision, 15*(4):8, 1–14, doi:10.1167/15.4.8. [PubMed] [Article]

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica, 146,* 7–18.

Uttal, W. R., & Tucker, T. E. (1977). Complexity effects in form detection. *Vision Research, 17*(3), 359–365.

Ware, C. (2008). *Visual thinking for design*. Burlington, MA: Morgan Kaufman.

Ware, C. (2013). *Information visualization: Perception for design* (3rd ed.). Burlington, MA: Morgan Kaufman.

Ware, C., & Plumlee, M. D. (2013). Designing a better weather display. *Information Visualization, 12*(3–4), 221–239.

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32*(5), 931–941.

Watamaniuk, S. N., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research, 29*(1), 47–59.

Webster, J., Kay, P., & Webster, M. A. (2014). Perceiving the average hue of color arrays. *Journal of the Optical Society of America A, 31*(4), A283–A292.

Whitaker, D., McGraw, P. V., Pacey, I., & Barrett, B. T. (1996). Centroid analysis predicts visual localization of first- and second-order stimuli. *Vision Research, 36*(18), 2957–2970.

Whitney, D., Haberman, J., & Sweeny, T. (2014). From textures to crowds: Multiple levels of summary statistical perception. In. J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 685–709). Cambridge, MA: MIT Press.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*(4), 160–168.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(3), 473–494.

Wolfe, J. (1998). What can 1,000,000 trials tell us about visual search? *Psychological Science, 9*(1), 33–39.