

Perception and Cognition in Statistical Graphics

Susan VanderPlas

April 16, 2015

Statisticians produce graphics for a multitude of reasons: to understand the structure of raw data, to check model assumptions, to present model predictions in an informative fashion; all of these goals are facilitated by appropriate visualizations. Some of these goals are best served by quick-and-dirty representations of the data, while highly polished graphics may be more useful for other goals. Regardless, it is important to convey the data appropriately, which means we must understand how graphics are perceived on at least a basic level. Tukey focused on graphics as a tool for exploratory analysis, because pictures can often display the data in a more coherent fashion than a table or spreadsheet; his book, [Tukey \(1977\)](#) details many different types of graphics and the appropriate situations to utilize them. This chapter focuses on the interaction between graphics and the human visual system, with the goal of understanding how to most efficiently and effectively convey information using statistical graphics. We will first consider the physiology and psychology of the perceptual experience in general, and then address the issue of graph-specific perception.

1 The Human Visual System

In order to design graphics for the human perceptual system, we must understand, at a basic level, the makeup of the perceptual system. There are multiple levels of perception that must correctly function in order to perceive visual stimuli successfully, but a somewhat simplistic higher-level analogy would be that we must understand both the hardware and software of the human visual system to create effective graphics.

The “hardware”, in this analogy, consists of the neurons that make up the eyes, optic nerve, and the brain itself. The higher-level functions (object recognition, working memory, etc.) comprise the “software” component. In addition, much like computer software, there are different programs running simultaneously; these programs may interact with each other, run sequentially, or run in parallel. The following sections provide an overview of the grey-matter (hardware) components of the visual system as well as the higher-level cognitive heuristics (software) that order the raw input and construct our visual environment.

1.1 Hardware

The physiology of perception is complex; what follows is a brief overview of the physiology of perception, focusing on the areas most important to the perception of statistical graphics. This physiological information is important in understanding the difference between the sensation (i.e. the retinal image) and the perception (the corresponding mental representation), which is an important distinction in understanding how statistical graphics are perceived. This overview will entirely ignore the finer details of the organization of the brain: feature detector cells, specific processing units for certain types of visual stimuli, and most of the horrifying or amusing experiments and incidents that led to the understanding of the brain. A more thorough presentation of these aspects of perception can be found in [Goldstein \(2009\)](#).

The Eye The eye is a complex apparatus, but for our purposes, the most important component of the eye is the retina, which contains the sensory cells responsible for transforming light waves into electrical information in the form of neural signals. These sensory cells are specialized neurons, known as rods and cones, which perceive light intensity (brightness) and wavelength (color), respectively. One section of the

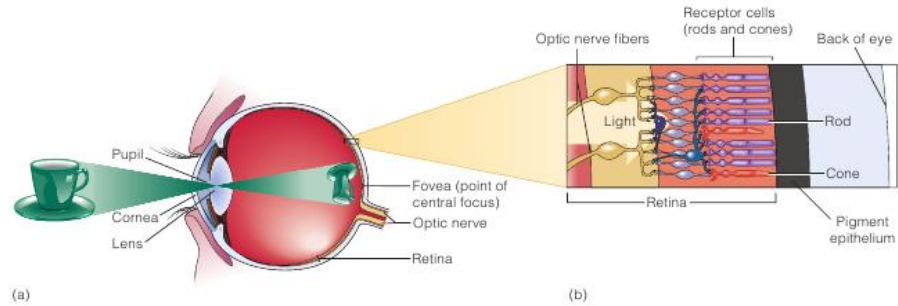


Figure 1: The human eye, with closeup of receptor cells in the retina (image from [Goldstein 2009](#), chap 3.1).

retina, known as the fovea, contains only cones; the rest of the retina contains a mixture of rods and cones. Figure 1 depicts the structure of the eye with a closeup of the retina.

Another important region of the retina is the blindspot, the area where the optic nerve exits the eye to connect the retina to the brain. There are no rods or cones in this region of the retina, and any vision in the region of space that maps onto this point is a result of two mechanisms: binocular vision (the other eye fills in the missing information) and your brain “filling in” what it believes should be there.

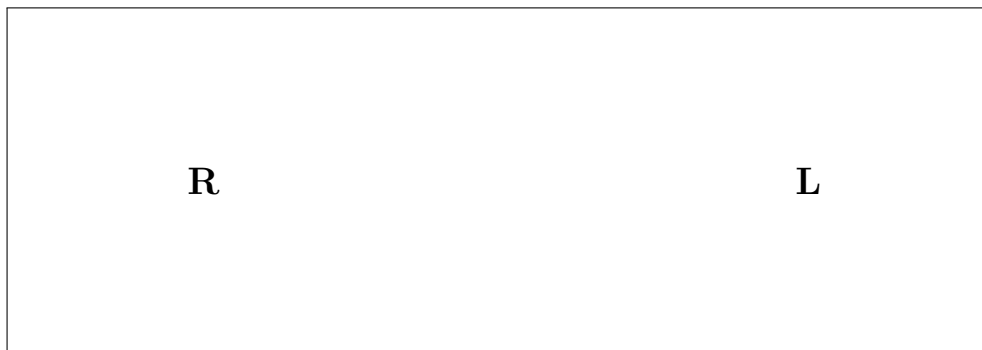


Figure 2: Illustration of the blind spot in the retina. Close one eye, and focus the other on the appropriate letter (R if the right eye is still open, L if the left eye is still open). Place the paper approximately 1 foot from your eye, and move the paper toward or away from your face until you notice the other letter disappear. Your brain “fills in” the other letter with the background.

Figure 3 shows the responsiveness of rods and each of the three types of cones to wavelengths of light in the visual spectrum. This image suggests that we have relatively good visual discrimination of the yellow-green portion of the color spectrum, but relatively poor discrimination of colors in the red and blue portions of the color spectrum.

As a result, rainbow-style color schemes are seldom appropriate for conveying numerical values, because the correspondance between the perceived information and the displayed information is not accurately maintained by the visual system ([Treinish and Rogowitz, 2009](#)). In addition, if any of the cones are missing or damaged as a result of genetic mutations, color perception is impaired, resulting in a smaller range of distinguishable colors. This set of impairments is known colloquially as color-blindness, and occurs in an estimated 5% of the population (approximately 10% of males, and less than 1% of females). Color blindness is discussed in more depth in section 1.3.2.

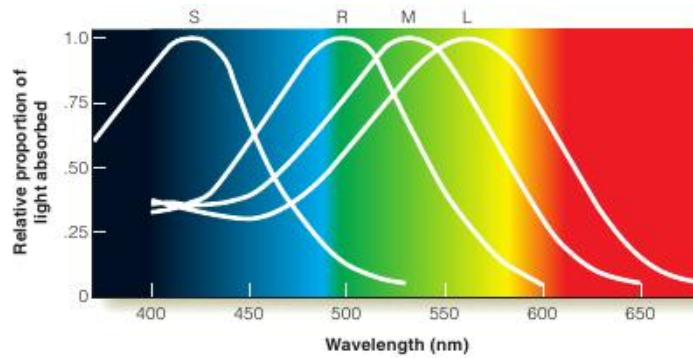


Figure 3: Absorption spectra of rods and short, medium, and long wave cones. (image from [Goldstein 2009](#), chap 3.3).

The Brain Once light hits the retina and causes a signal in the receptor cells, the information travels along the optic nerve and into the brain. Multiple neighboring rods are connected to the same neuron, where each cone is connected to a single neuron. The combined wiring of rod cells is responsible for the Hermann grid illusion and the Mach bands seen in figure 4. Both of these illusions are a product of lateral inhibition, which is a result of the wiring of rod cells in the retina. Essentially, neurons can only fire at a specific rate, so when neighboring cells are all stimulated simultaneously, the combined neuron cannot fire fast enough to pass on all of the signals, causing “inhibition”. The specifics of this response and its relationship with the wiring of the receptor cells are somewhat complex; a more thorough explanation can be found in [Goldstein \(2009\)](#), chapter 3.4.

Once neural impulses have left the retina through the optic nerve, they travel to the visual cortex by way of several specialized structures within the brain that process lower-level signals. Receptor cells in the visual cortex respond to specific angles, spatial locations, colors, and intensities, and arrays of these special ‘feature detector cells’ process the information into a form that higher-level processes can utilize. These higher-level processes are what we have previously called ‘software’: they are not directly related to the physical brain, but they do process information heuristically to produce higher-level reasoning and conclusions. In the next section, we explore some of the higher-level processes responsible for visual perception.

1.2 Software

Many of the processes for visual perception run simultaneously; in absence of a strict temporal ordering, we will start with the more basic tasks of visual perception and proceed towards higher-level processes. We will begin with attention.

1.2.1 Attention and Perception

In many tasks, it is necessary to pay attention to many parallel input streams simultaneously; this is particularly true for complex tasks like driving a car. These tasks demand divided attention; the brain must process many different sources of information in parallel. By contrast, most image recognition tasks require selective attention, that is, focusing on specific objects and ignoring everything else. The brain accomplishes this attention through several mechanisms.

Selective attention is accomplished by focusing the fovea (the area with the highest visual acuity) on the object. For instance, if the object is a page of text, each word will pass through the fovea, producing a focused stream of visual input. This stream of input consists of saccades (jumps between points of focus) and pauses in which the visual information is relayed to the brain. Figure 5 shows the saccades (lines) and pauses (circles) resulting when someone scans a paragraph of text. These saccades and pauses are utilized in eye-tracking technology to determine which parts of an image the observer is focusing on (and by extension,

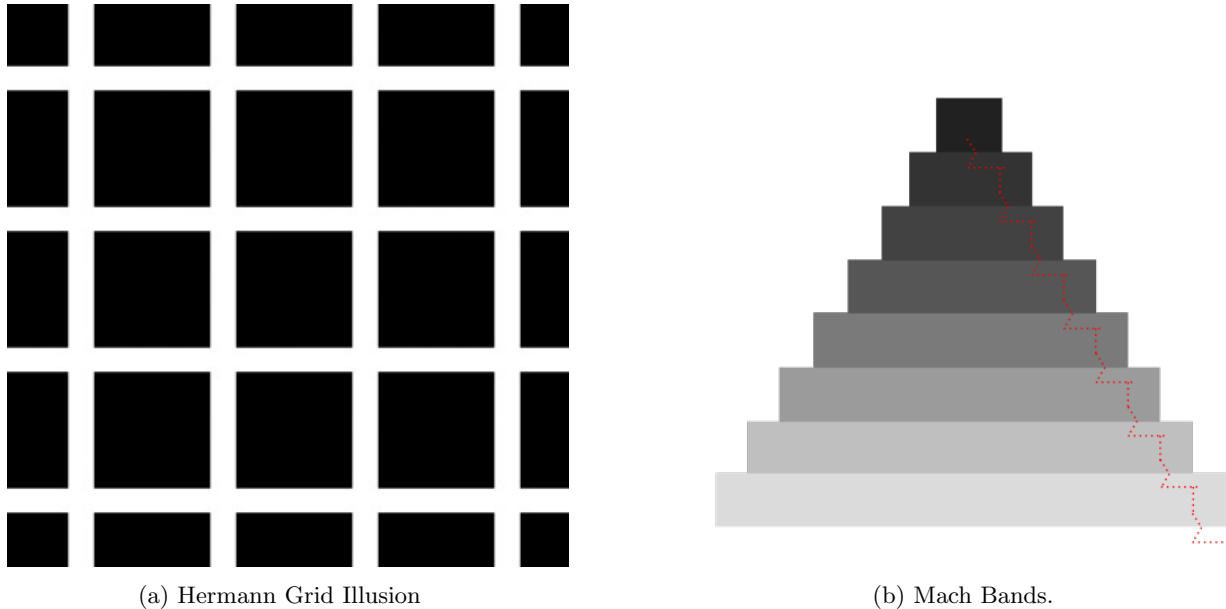


Figure 4: Optical Illusions resulting from lateral inhibition. The Hermann Grid illusion causes dark circles to appear at the intersection of white lines; the Mach bands illusion causes the borders of adjacent rectangles to appear more strongly defined.

which information is being encoded by the brain). Further discussion of eye tracking is provided in section 3.2.1.

Selective attention is generally necessary for perception to occur, though there is some information that is encoded automatically. Experiments such as the fairly famous “gorilla” film¹ demonstrate that even when there is attention focused on a task, information extraneous to that task is not always encoded, that is, even when participants focused on counting the number of passes between players in the basketball game, they did not notice the gorilla walking through the middle of the court. It is important to understand which parts of a visual stimulus are the focus of a given perceptual task, because most of the information encoded by the brain is a result of selective attention. Eye-tracking can be an important tool useful to understand these perceptual processes, but participants are often able to report which parts of a stimulus contributed to their decision as well.

Within the brain, attention is important because it allows different regions of the brain which process color, shape, and position to integrate these perceptions into a multifaceted mental representation of the object (Goldstein, 2009). This process, known as binding, is essential to coherently encode a scene into working memory. Feature integration theory (Treisman and Gelade, 1980) suggests that these separate streams of information are initially encoded in the preattentive stage of object perception; focusing on the object triggers the binding of these separate streams into a single coherent stream of information. Many single features, such as color, length, and texture are *preattentive*, because they can be pinpointed in an image without focused attention (and thus can be located faster), but specific combinations of color and shape require attention (because the features must be bound together) and are thus more difficult to search. Preattentive features are generally processed in parallel (that is, the entire scene is processed nearly simultaneously), while features requiring attention are processed serially. Examples of features processed serially and in parallel are shown in figure 6, taken from Chapter 6 of Helander et al. (1997). The importance of preattentive processing to statistical graphics is discussed in Section 2.1, which includes a demonstration of preattentive search in figure 16.

¹<http://www.theinvisiblegorilla.com/videos.html>

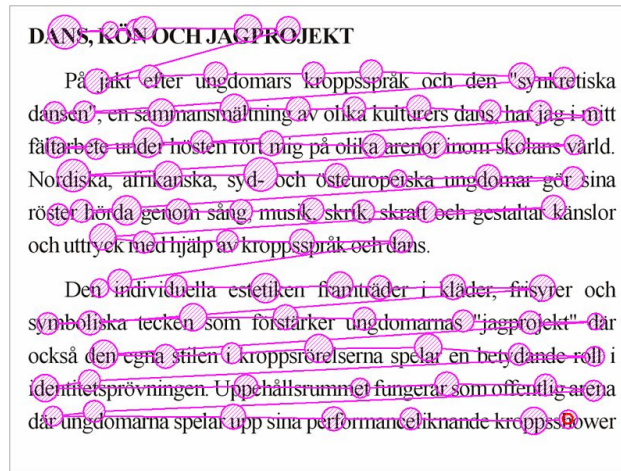


Figure 5: A plot of saccades made while reading text. Saccades, shown by the lines, indicate “jumps”, while pauses are shown by circles, with size proportional to the time spent focusing on that area.

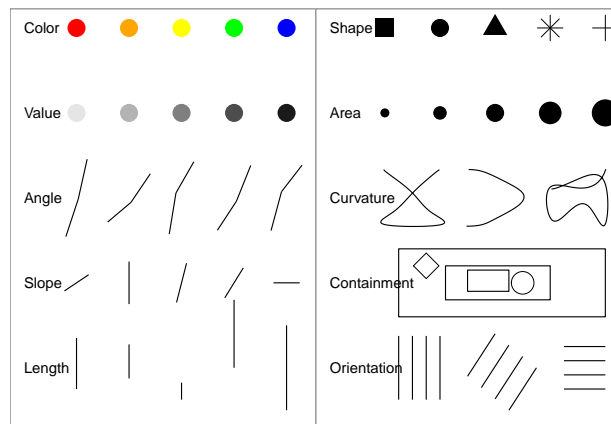


Figure 6: Examples of features detected serially or in parallel (Chapter 6, [Helander et al. 1997](#))

Feature integration as a result of attention enables the brain to process a figure holistically and integrate all of the separate aspects of the object into a single perceptual experience. This processing is important for the most basic visual processes we take for granted, including object perception.

1.2.2 Object Perception

The most basic task of the visual system is to perceive objects in the world around us. This is an inherently difficult task, however, because the retina is a flat, two-dimensional surface responsible for conveying a three-dimensional visual scene. This dimensional reduction means that there are multiple three-dimensional stimuli that can produce the same visual image on the retina. This is known as the inverse projection problem - an infinite number of three-dimensional objects produce the same two-dimensional image. Less relevant to statistical graphics, but still complicating the object perception process, a single object can be viewed from a multitude of angles, in many different situations which may affect the retinal image (lighting, partial obstruction, etc). In addition, we recognize objects even when they are partially obscured or viewed from an angle we have not previously seen. These problems mean that the brain must utilize many different heuristics to increase the accuracy of the perceived world relative to an ambiguous stimulus.

The most commonly cited set of heuristics for object perception (and the set most relevant to statistical graphics) are known as the *Gestalt Laws of Perceptual Organization* (Goldstein 2009, Chapter 5.2). These laws are related to the idea “the whole is greater than the sum of the parts”, that is, that the components of a visual stimulus, when combined, create something that is more meaningful than the separate components considered individually. The Gestalt laws are as follows (Goldstein, 2009):

- **Pragnanz - the law of good figure.** (Also referred to as the law of closure) Every stimulus pattern is seen so that the resulting structure is as simple as possible.
- **Proximity.** Things that are close in space appear to be grouped.
- **Similarity.** Similar items appear to be grouped together. The law of similarity is usually subordinate to the law of proximity.
- **Good Continuation.** Points that can be connected to form straight lines or smooth curves seem to belong together, and lines seem to follow the smoothest path.
- **Common Fate.** Things moving in the same direction are part of a single group.
- **Familiarity.** Things are more likely to form groups if the groups are familiar.
- **Common Region.** Things that are in the same region (container) appear to be grouped together
- **Uniform Connectedness.** A connected region of objects is perceived as a single unit.
- **Synchrony.** Events occurring at the same time will be perceived as belonging together.

Figure 7 shows examples of many of the gestalt laws, which when combined help to order our perceptual experience. We have discussed how we attend to visual stimuli and how they are recognized, but for the perceptual experience to be meaningful, previous experiences must be stored into memory in some coherent way. The next section discusses how visual scenes are stored into long-term memory.

1.2.3 Visual Memory

We have discussed how visual stimuli are perceived and how objects are recognized; we now must examine how visual stimuli are encoded into memory. Most researchers believe that visual perceptions are encoded in an analog fashion, so that the memory of an image is closely related to the perception of that same image (Matlin, 2005). Other theories suggest that visual perceptions are encoded semantically, that is, the

²From http://yusylvia.files.wordpress.com/2010/03/gestalt_illustration-01.jpg

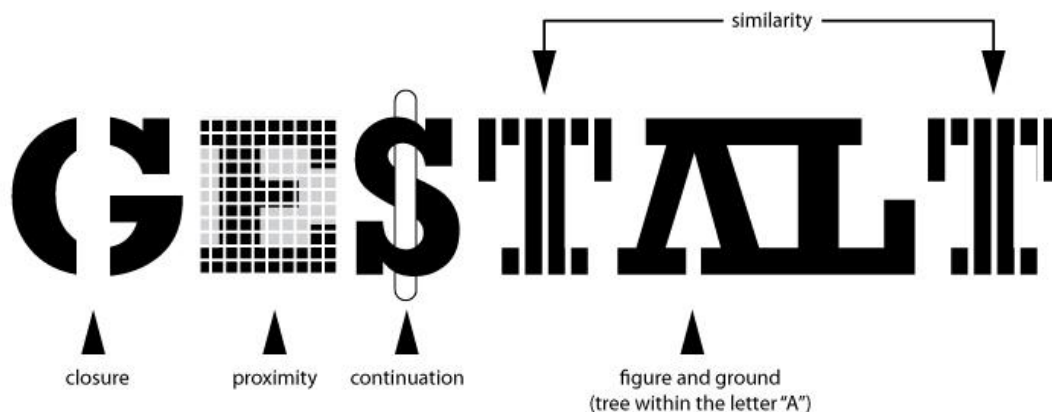


Figure 7: The gestalt laws of perception ²

description of a visual scene would be encoded, rather than a mental “image” of that scene. Both theories are likely at least partially correct, but the analog encoding of visual images is more relevant to statistical graphics because the accuracy of the stored image has the potential to affect recall of the contents of that image (and thus what people remember about a particular graphic). Experimental evidence for analog encoding includes the mental rotation task, where participants must determine whether or not a figure is a rotation of a target figure, as shown in figure 8. [Shepard and Metzler \(1988\)](#) showed that reaction time was proportional to the angle of rotation of the stimuli, which suggests that participants were mentally rotating the figure as they would rotate a three-dimensional figure in space.

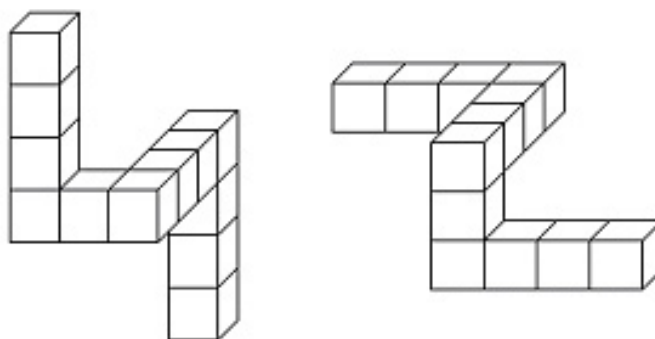


Figure 8: Rotation task ([Shepard and Metzler, 1988](#)). Are the two images the same?

In addition, [Kosslyn et al. \(1978\)](#) showed that mental representation of distances in a figure are accurate and that the time to encode those distances is proportional to the distances in the actual figure. These studies suggest that the memory of an image (statistical graphic or otherwise) is a reasonably accurate facsimile of the original image (though the accuracy of the mental representation is of course likely to be moderated by attention and recall ability).

Another facet of visual memory that will be important to understanding perception and memory of statistical graphics is that the “gist” of an image is stored along with the image. In these cases, recall ability is more consistent with the semantic encoding of images; that is, when shown an ambiguous figure and immediately asked for a description, participants could not give an alternate interpretation of the figure after the experiment was complete. In the case of figure 9, participants who initially said the figure was

a duck did not report having viewed a picture of a rabbit after the experiment, even though the image is consistent with either interpretation. This suggests that in some cases, verbal encoding of a figure (i.e. describing it as a duck) disrupts the mental representation of the picture. This is common in other types of memories as well: when the gist of a passage is stored, the actual content of the passage is no longer accessible. In other words, we would expect that if someone had to interpret a graph, they would remember the interpretation much more strongly than the actual graph, even if that interpretation was incorrect or incomplete.

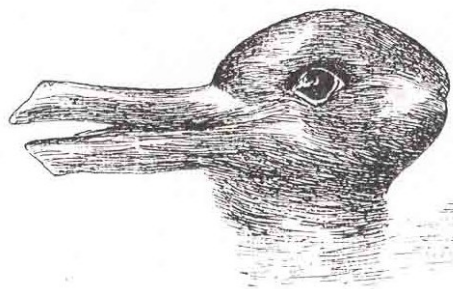


Figure 9: An ambiguous image that could be either a rabbit or a duck. When participants were asked to identify the image initially, they could not provide an alternate interpretation of the figure later.

The “software” of the visual system is of course more complex than the few modules listed here, but understanding attention, object perception, and how images are stored for later retrieval in the brain will make designing statistical graphics for the visual system easier and will also help with evaluating graphics based on the capabilities of the human visual system.

We have discussed the neural hardware of the visual system and some of the higher-level software that contributes to our ability to create and understand meaning in the world around us. Occasionally, our highly tuned perceptual system fails in unusual ways due to the heuristics and algorithms that were optimized for operation in a three-dimensional world where the main tasks were hunting, gathering, and avoiding predators. The next section will examine three interesting results of this tuning that are important to the design of statistical graphics as we transition from the psychophysics and cognitive psychology literature to statistics and human-computer interaction literature.

1.3 Bugs and Peculiarities of the Visual System

1.3.1 Logarithmic perception

One of the earliest psychophysics researchers, Ernst Weber, discovered that the difference threshold, the smallest detectable difference between two sensory stimuli, increased proportionately with the magnitude of the stimulus. This statement, known now as Weber’s Law, holds true for a large range of intensities of a number of senses. Numerically, Weber’s Law is stated as

$$\frac{\Delta S}{S} = K \quad (1)$$

where K is a constant called the Weber fraction, S is the value of the standard stimulus, and ΔS is the difference between the standard stimulus and the test stimulus. So if a participant is given a 100-g weight and a 102-g weight and can just barely tell the difference between the two, then $K = 0.02$ and we would assume that the difference between a 200-g weight and a 204-g weight would be just barely detectable as well (Chapter 1, [Goldstein 2009](#)). While this example concerns the ability to distinguish weight, the same law holds for the ability to distinguish sounds of different intensities as well as intensity of colors. The tendency of the brain to perceive stimuli in a logarithmic fashion is true across many perceptual domains.

In fact, when kindergarden children are asked to place numbers 1-10 along a number line, they place 3 in about the middle, just as one would expect from a logarithmic perspective. This ability disappears with mathematical education, but persists in those who are not given a formal education in mathematics, indicating that our brains are naturally wired to perceive numbers logarithmically (Varshney and Sun, 2013). Some sensory domains utilize logarithmic scales to measure stimuli; scales such as sound intensity, earthquake intensity, and frequency along the electromagnetic spectrum are logarithmic. Information theory suggests that logarithmic scaling provides optimal compression of information to minimize relative errors in perception while accounting for limits in our neural bandwidth. Sun et al. (2012); Varshney and Sun (2013) showed that a bayesian model for perception would result in a model that mimics the logarithmic relationship in Weber’s Law. This suggests that the logarithmic nature of human perception is a result of an heuristic that increases processing power by reducing the neural bandwidth necessary to process information through quantization of continuous information and compression of discrete information. From a statistical graphics point of view, then, log-transformed scales should be used instead of linear scales for continuous color scales and for contour plots with a fixed number of contours, as this provides more information discrimination ability and mimics natural human perceptual tendencies. The reasons for this guideline are discussed more in section 1.4.

1.3.2 Colorblindness and color perception

Another common “bug” in the visual system are mutations that change (or remove entirely) the cones in the retina. Such mutations are commonly termed “colorblindness” and encompass many different types of mutations, shifts and deletions that affect color perception in the visual system. These mutations affect up to 5% of the population, and are generally more common in males than they are in females, as two of the three genes producing cones are found on the X chromosome. Evolutionarily, these mutations are maladaptive for gathering plants (as finding red berries within green leaves is more difficult with the most common types of colorblindness), but may be adaptive for seeing camouflaged objects (Morgan et al., 1992). In statistical graphics, however, these mutations often disrupt perception of standard color schemes used in maps, heatmaps, and divergent color scalings.

In the natural world, many strategies can be used to compensate for colorblindness; the most common of these strategies is to look for textural variation instead of color variation (which may be why camouflaged objects are easier to see), but these strategies fail when viewing abstract, constructed visual stimuli, such as graphics, which may not have textural variation which corresponds to color variation. Compounding this problem, the rainbow color schemes that are commonly used are particularly vulnerable to misinterpretation by colorblind viewers. Figure 10 shows a map using a rainbow color scheme (first shown in Light and Bartlein (2004)) and simulated images showing what that map would look like to those with missing cones (-anopia) and cones with altered wavelengths (-anomaly). Simulations³ show that rainbow color schemes are incredibly difficult for color-deficient individuals to read, because the opposite ends of the color spectrum appear extremely similar with mutations to the first or second cones. In addition, categorical rainbow color schemes require longer fixation times to identify regions of importance (Lewandowsky et al., 1993) and result in decreased recall accuracy compared to monochrome categorical color schemes. Light and Bartlein (2004) provides color schemes that are more appropriate for those with color deficiencies, but not all of these schemes are appropriate for all types of color blindness. Silva et al. (2011) suggest many tools to recommend appropriate color schemes for colorblind users as well as tools to preview graphics as they might look to color-deficient or colorblind users.

Appropriately colored maps and graphs are not only useful for those who have impaired color vision, they can also be much easier to read for those with normal vision. Treinish and Rogowitz (2009) suggest that color schemes which utilize the range of human color vision appropriately produce more aesthetically pleasing graphs and more accurately convey data in a form appropriate for the human perceptual system. Even some dual-color schemes may be problematic, as some evidence (Lewandowsky et al., 1993) suggests that these schemes are not infrequently inverted when encoded in memory, and are suboptimal for those

³provided by <http://www.color-blindness.com/coblis-color-blindness-simulator/>

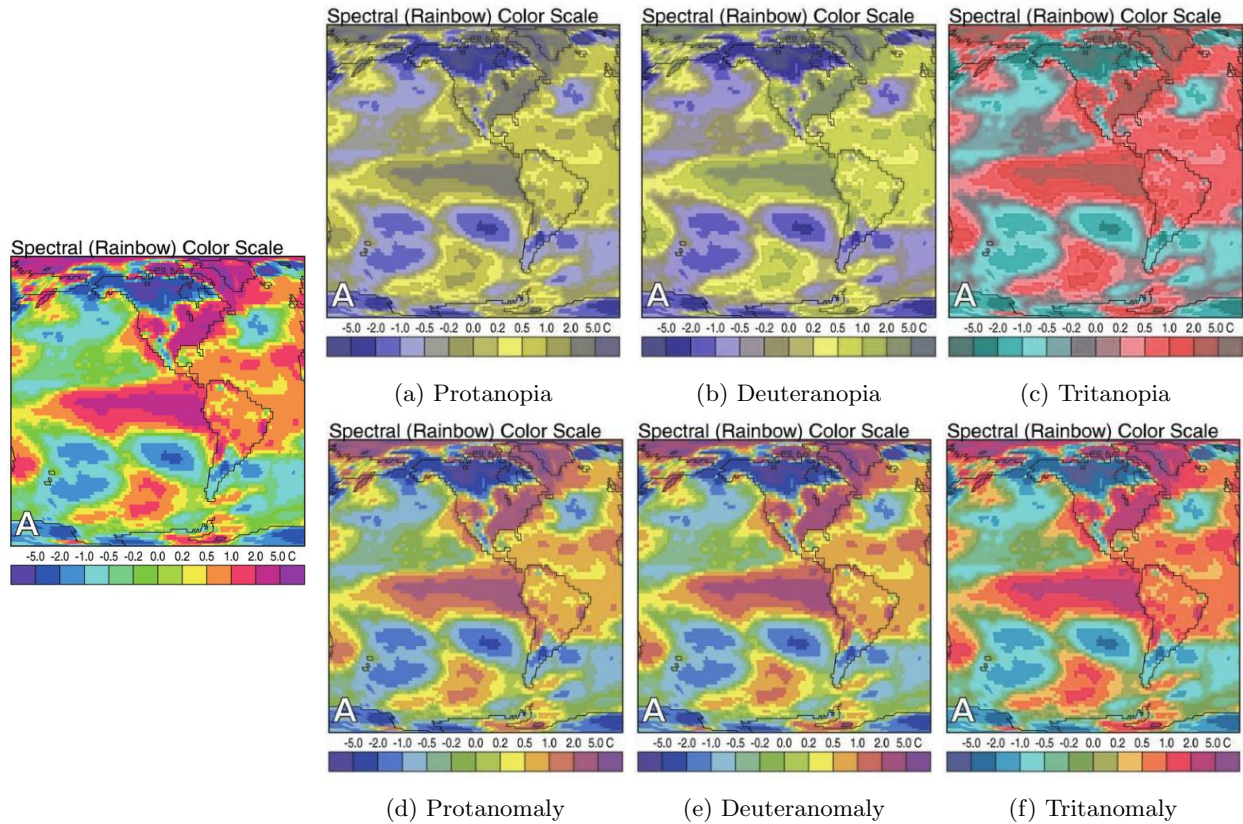


Figure 10: Rainbow color scheme with simulations of each of the six common types of color deficiency. The original image, on the left, is from [Light and Bartlein \(2004\)](#). The top row of pictures on the right show simulations of the map when cones are entirely missing. The bottom row of pictures show simulations of the map when each cone is altered due to genetic mutation.

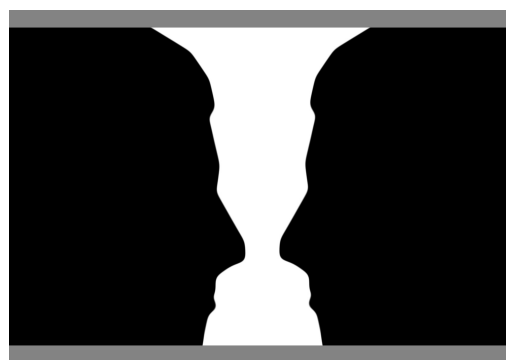
with limited color perception as well as when printed in monochrome. Psychologically, color schemes which utilize both hue and intensity (for instance, transitioning from blue to red through white) require binding of two features, increasing encoding latency and the possibility of recall errors. Where possible, schemes utilizing only one feature (generally intensity, to preserve accessibility) are preferable.

1.3.3 Optical Illusions

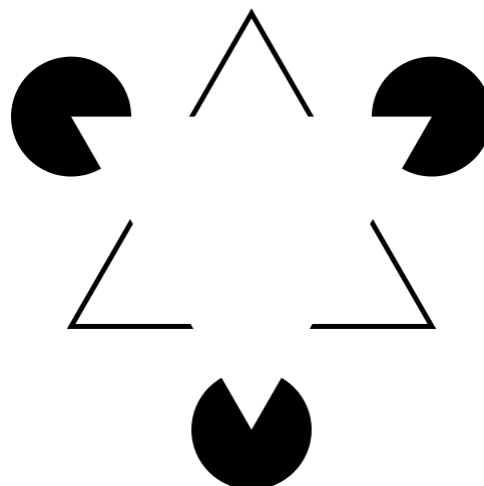
The “software” programs presented in section 1.2 are generally efficient at completing everyday tasks: navigating the environment, avoiding predators (lions or cars, as the case may be), and identifying situations and objects relevant to the task at hand. These heuristic-style approaches produce suboptimal results when applied to more artificial tasks, such as reading statistical graphics. As a result, it is important to understand where conflicts between sensation and perception may occur, so that these conflicts can be dealt with or avoided entirely. In this section, we will discuss several optical illusions and explanations for their occurrence based on the visual system.

Physiological Illusions The illusions shown in figure 4 are illusions which occur due to the wiring of the brain. These illusions can generally be avoided in statistical graphics, but are difficult to counteract once they occur, as the illusion is literally hard-wired into the brain.

Gestalt Illusions Some illusions occur due to a conflict of gestalt principles. Two of these illusions are shown in figure 11: the figure/ground illusion, and the illusory contour illusion.



(a) Figure-Ground Illusion



(b) Kanizsa Triangle Illusion

Figure 11: Illusions due to misapplied or ambiguous Gestalt rules.

The figure/ground illusion depends on the color of the top and bottom edges of the picture; if the edges are black, the vase appears to be the central part of the image; if the edges are white, the faces appear to be the central part of the image. When the edges are omitted, the image seems to oscillate between a vase and the profile view of two heads. This is a result of ambiguity in identifying which part of the image is the background; when the top and bottom edges are present, that cue is sufficient to resolve the illusion. The Kanizsa triangle demonstrates the Gestalt principles of good form and continuity: We perceive objects that are partially obscured by a floating white triangle, even though no such triangle actually exists. The illusory triangle produces an image that is much simpler (3 circles, a black triangle outline, and a white triangle) than the objects that are actually displayed (3 partial circles and three V shapes arranged pointing in toward

a central point). In addition to the existence of the illusory contour, we also perceive some depth to the image; that is, the white triangle is perceived as being above the other components of the image (Coren and Porac, 1983), as this is the only way to make sense of the set of stimuli in a simple fashion. In general, these gestalt principles make sense of the natural world, but when applied to artificial contexts, they occasionally produce unexpected results.

Depth illusions Other optical illusions occur due to the optimization of the visual system for three-dimensional perception. These three-dimensional heuristics can produce unexpected or misleading results when applied to two dimensional objects.

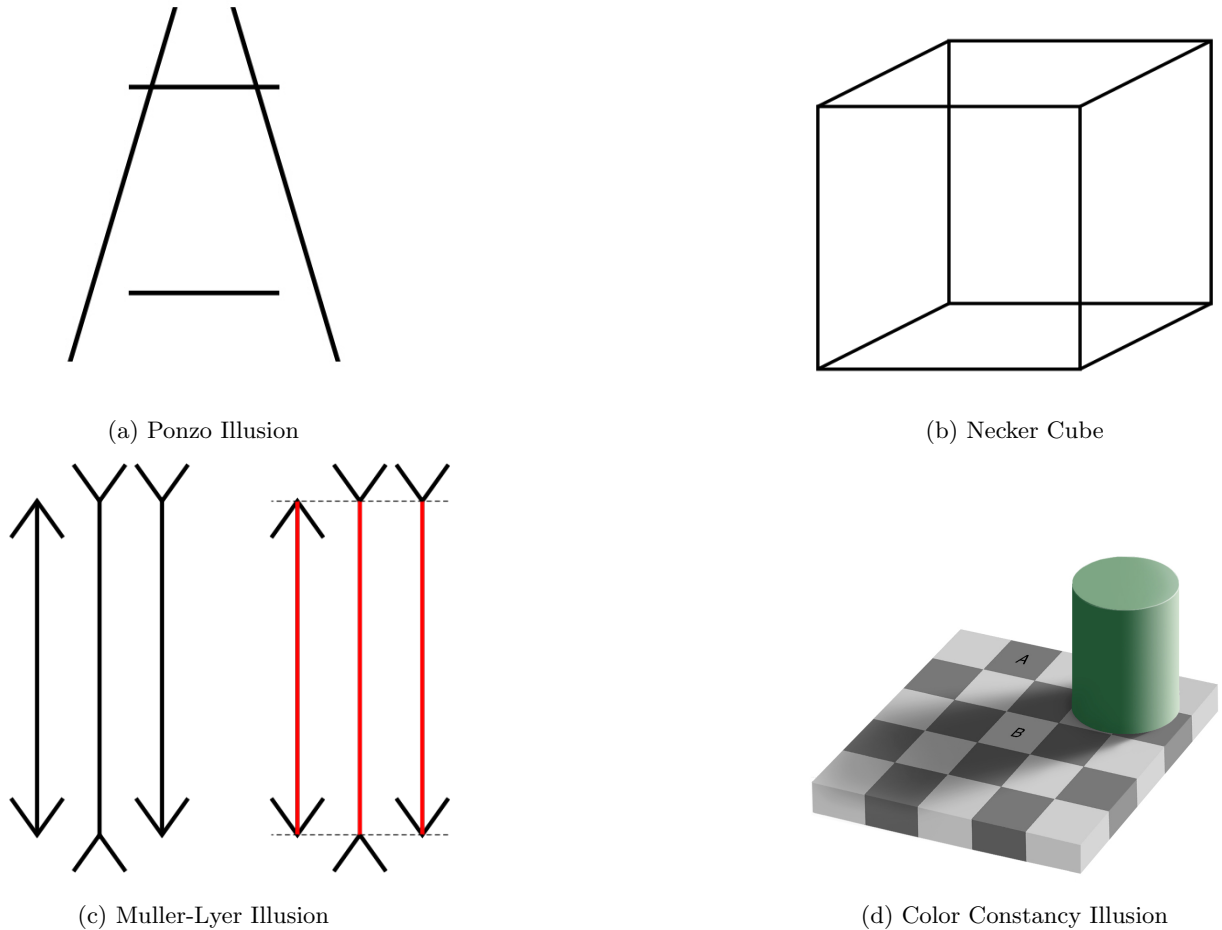


Figure 12: Illusions due to misapplied depth perception.

Figure 12 contains four of the more interesting optical illusions that result from ambiguous figures that trigger depth cues. The Ponzo illusion (figure 12a) suggests that the top line is longer than the bottom line, because of the implied convergence of the two vertical lines (to understand the natural scene that produces this illusion, consider railroad tracks converging at the horizon). The Necker Cube, shown in figure 12b, can be seen such that the top-right face is closest to the viewer or alternately such that the bottom-left face is closest to the viewer. Due to the ambiguity in the image, it will often seem to “flip” when the viewer temporarily loses focus on the image (Gregory, 1997). The Muller-Lyer illusion (figure 12c) is generally believed to result from misapplied depth cues as well - the left-most image would occur in nature as the

exterior corner of a building, the middle image would occur when viewing an interior corner of the same building, further away from the viewer (Ward et al., 1977; Gregory, 1968; Fisher, 1970). As a result of the illusion, the middle line appears to be longer than the first or third lines. Figure 13 shows the first two parts of the illusion in a context which removes the ambiguity through additional depth cues. The additional cues result in the resolution of the illusion. Finally, the color constancy illusion shown in figure 12d suggests that the square marked A is much darker than the square marked B, even though the two squares are the same color. This illusion results from our experiences with depth and shadows: square B is perceived to be the same color as the lighter-colored squares outside the shadow, while square A is perceived to be the same color as the other dark squares in the tile pattern, regardless of the actual color due to the shadow.



Figure 13: The Muller-Lyer illusion in a non-ambiguous three-dimensional context.

Depth illusions in particular result from a conflict between our experience with the three-dimensional world and the appearance of two-dimensional ambiguous stimuli. The Necker cube “flips” because there are two physical objects that could produce the same retinal image, the Muller-Lyer illusion exists because our experience with the three-dimensional world is harnessed inappropriately for a two-dimensional figure, and the color-constancy illusion exists because our brains automatically correct a pseudo three-dimensional image to represent the reality of that image in the real world. These conflicts occur in statistical graphics as well: certain graphical configurations can trigger a three-dimensional heuristic in the brain and lead to misleading conclusions (VanderPlas and Hofmann, 2014; Hofmann and Vendettuoli, 2013).

Add more information on Sine Illusion and other illusions?

There are other optical illusions that have the potential to appear in statistical graphics but are not easily classified (or necessarily easily explained). Two of these illusions are considered below.

Other Important Optical Illusions Certain illusions do not lend themselves to simple classification. While many illusions are the result of multiple concurrent processes in the brain, these illusions may not even be fully understood. The Poggendorff illusion, shown in figure 14, is one such illusion.

Gregory (1963, 1997) suggests the Poggendorff illusion is similar to the Muller-Lyer illusion in that it results from misapplied depth cues, but Green and Hoyle (1963); Ward et al. (1977) found no evidence that participants viewed this illusion in any three-dimensional context. Instead, Green and Hoyle (1963) suggests that the illusion results from a tendency to perceive acute angles as less acute and obtuse angles as less obtuse than the image suggests. As the illusion disappears as the angle of the line segment approaches horizontal, this seems to be a reasonable explanation, but it is almost certainly not complete (Morgan, 1999), as the illusion survives in forms which do not preserve the acute angle intersections. Regardless, this illusion can make it difficult to read line graphs (Amer, 2005; Poulton, 1985) if proper precautions (use of reference lines and grid lines) are not taken.

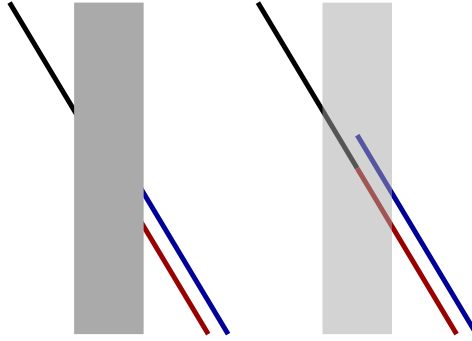


Figure 14: The Poggendorff Illusion. The left figure shows a black line which is obscured by a grey rectangle; it appears that the blue line would intersect the black line if the rectangle were not present. In fact, the black line and the red line are co-linear, and the blue line is parallel to both other lines.

The second of these illusions is the cafe wall illusion, shown in figure 15, named because this tile pattern is apparently common in cafes.

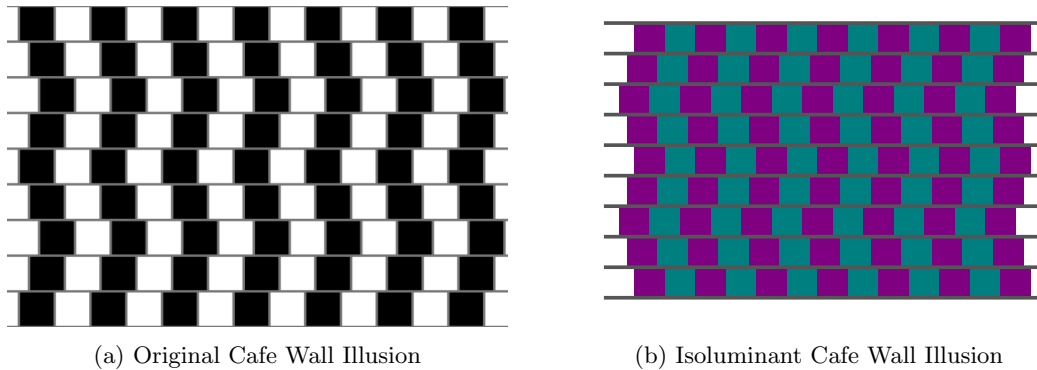


Figure 15: The Cafe Wall illusion. The lines between rows of black and white tiles are parallel but appear to be tilted. The second image shows the isoluminant version, which mitigates some of the illusion but does not entirely eliminate the effect.

The cafe wall illusion is in part due to the contrast between light and dark zones (as in the mach bands and hermann grid illusion), and much of the illusion is resolved when the black and white tiles are replaced with isoluminant colors, but some of the illusion still remains. [Westheimer \(2007\)](#) suggests that this portion of the illusion is because the position of a black-white border will be biased to appear closer to the black side of its physical location, an effect which is compounded in the cafe wall illusion to produce the appearance of tilted lines. This illusion, while not explicitly found in most statistical graphics, shows that even simple (and pleasant) configurations of geometric objects can wreak havoc in the brain under the right circumstances. In fact, the illusion is so simple that it is also known as the “Kindergarten illusion”, but its cause is sufficiently complex that it has not been fully explained by psychologists or neuroscientists.

1.4 Cognitive Load

It is fairly well established that statistical graphics are useful in part because they can replace long tables of data, summarizing information in a form that is much easier to understand and mentally manipulate; as such, it would not be out of line to suggest that graphics require less cognitive load than tables of the

same data. When we discuss cognitive load, we typically mean the limits in short-term memory, cognitive manipulation, attention, and mental bandwidth that bound our ability to take in new information and draw conclusions from that information. In this section, we will briefly discuss some of these considerations.

Short Term Memory A famous paper in memory and cognition (Miller, 1956) suggests that active memory can contain only 7 (plus or minus two) chunks of information. A chunk of information could be a single letter or number, a meaningful collection of several letters or numbers (e.g. a word or an area code), or an association. This limitation is important in designing information for graphical consumption. For instance, the number of categories in legends should be limited to 7, to allow a viewer to store the associations within the legend and then use that information to understand the graph. Abuse of this limitation is referred to as a “color mapping attack” in Conti et al. (2005), a paper detailing the various ways to “attack” a human visualization system. Similarly, viewers should not be expected to remember more than 7 “chunks” of information from a single graph. Due to these limitations in memory, when a single color scale is used to represent more than one order of magnitude of variation, using a logarithmic scale provides more optimal information scaling than using a linear color scale (Sun et al., 2012; Varshney and Sun, 2013).

Information Integration Integrating multiple dimensions of information (or mentally combining multiple graphics) is another area which can strain the ability of the brain to utilize information effectively. Well-constructed graphs can help the brain to integrate information by connecting points across dimensions (through the use of regression lines, clustering, etc.), which creates “chunks” of information that can then be stored in memory in a more compressed format. These chunks are useful because they allow people to draw conclusions from multiple sets of data across multiple dimensions (Gattis and Holyoak, 1996). Poorly created graphics may make this task harder or even promote the encoding of misleading chunks; for instance, data that is overplotted may obscure the important trend and may also produce chunks which lead to the wrong associations being stored in memory. This integration limitation is very much related to short-term memory, but is also constrained by mental effort limitations and processing capacity. As a result, it is important to reduce the effort required to integrate multiple graphics.

Attention Human attention is limited; thus visualizations which do not focus attention on important aspects of the data are likely to confuse the reader. Tukey (1977) said “The greatest value of a picture is when it forces us to notice what we never expected to see”. When there are too many salient features to notice anything in particular, attention is split too many ways to gain useful information from the picture. Graphics should present data in a controlled fashion, so that focused attention is rewarded with useful information taken from the graph. Conti et al. (2005) describes graphs that do not follow this principle as “processing attacks”, in that the overload the “CPU” with needless calculations and mental manipulations that are ultimately futile to understanding the data.

The consequence of the limits of human perception and processing capacity is that there is a limited amount of information one can expect to portray graphically; thus graphics should be designed to most efficiently communicate information so that this cognitive overload does not occur. The next section presents studies which examine the perception of graphs and charts directly across a wide range of perceptual levels and experimental conditions.

2 Statistical Graphics

Psychologists who study graphical perception are generally concerned with the underlying mechanisms of effects within the brain, and thus study very simple graphics and lower-level perceptual effects. In statistics, the literature is somewhat more variable; Cleveland and McGill (1985) produced the seminal paper on the subject, but outside of that work, there are relatively few papers that examine the accuracy of judgments made from graphs through user studies that mimic the way graphs are used in practice. There have been a few papers in other disciplines; business and communications researchers occasionally study graphs and

charts as well. As a result, the literature in this area is scattered across many disciplines. In order to organize this section effectively, we will begin with the lower-level graphical perception literature and conclude with studies that have more external validity and are applicable to statistical practice. For the purposes of the following sections, lower-level perception research is research which involves either extremely simple graphical elements (or non-graphical research which applies directly to graph perception processes), or which does not require attention (pre-attentive features of graphics). These experiments provide some information, but are less informative than experiments which utilize more complex graphics in realistic scenarios.

2.1 Preattentive Perception of Statistical Graphics

Much of the lower-level research within the statistical graphics literature has been performed by those within the psychological community that study human information processing. In particular, [Healey and Enns \(1999, 2012\)](#) have produced several papers studying the accuracy of conclusions viewers can make after less than 1/2 of a second of viewing an image. As discussed in section 1.2.1 and demonstrated in figure 16, certain features do not require individual focus to process; these features are called *preattentive* and can be detected on the first glance (typically within 250ms). Healey’s work focuses on determining which features can be detected in a pre-attentive fashion, and whether a hierarchy of features exists when these features are combined. Healey suggests that for three-dimensional displays, the 3d layout is determined first, surface structure and volume are determined next, followed by object movement (if present), luminance gradients, and color; he suggests that if there are conflicts between these 5 levels, priority is given to an earlier process ([Healey and Enns, 2012](#)). [Healey and Enns \(1999\)](#) showed that if visualizations are carefully constructed to conform to the architecture of the human visual system (isoluminant colors, removing certain background patterns from textural arrays), visual estimation tasks can be performed preattentively. The experiment also revealed an interference effect between texture and color that corresponds to previously documented interference between preattentive features ([Treisman, 1985](#)). Figure 16c demonstrates the interference effect; it is much easier to locate the target point in figure 16a or figure 16b than it is to locate the target in figure 16c.

Healey’s work on preattentive perception is interesting, and provides a reasonable approach to creating graphics compatible with the human visual system, but his work is largely focused on multidimensional displays and his focus on preattentive processes limits the applicability of his work to statistical graphics. In particular, most graphics are created with the idea that a viewer will spend more than one second looking at the graph, so not all features need to be preattentive to be useful. In addition, many of the multidimensional displays he designs are very load-intensive to understand; with so many additional dimensions, the viewer must spend considerable time understanding the scales and legends which correspond to each variable (feature integration over multiple dimensions is time intensive and generally not preattentive). In the next section, we will examine the literature concerning higher-level graphical perception, including perception at the attentive level and which types of graphs are more accurately perceived by viewers.

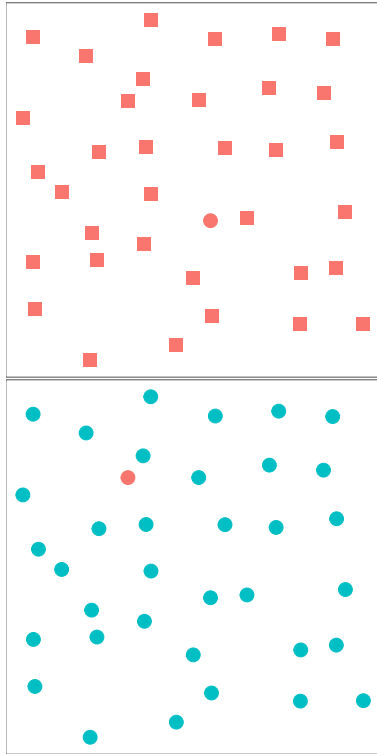
2.2 Conscious Perception of Statistical Graphics

Graph perception from a statistician’s point of view is more focused on the attentive stage of perception: When asked to answer a question using a graph, what parts of the graph are useful, and how is the information transferred from the image to working memory in the brain? Several models have been proposed to describe this process; of these, the set of “task models” and “integration models” seem to be most consistent with empirical evidence.

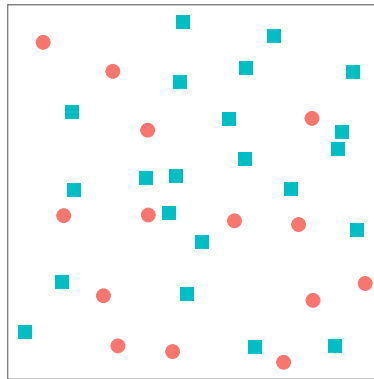
2.2.1 Models of Graph Perception

These task-based models suggest that task-based graphical perception, e.g. using a graph to answer a specific question involves several stages of information processing ([Ratwani et al., 2008](#)).

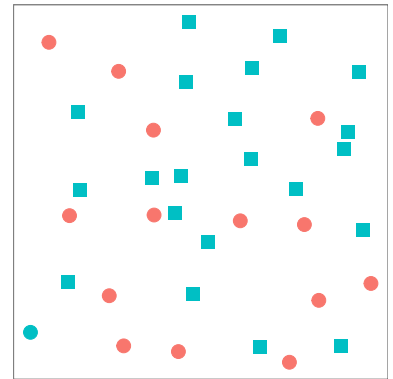
1. Parts of the question are read several times



(a) Shape



(b) Color



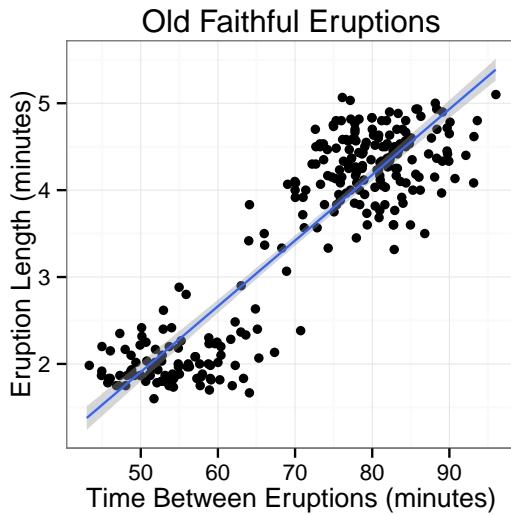
(c) Interference

Figure 16: Shape and color are detected preattentively in figures (a) and (b), but interfere in figure (c) so that location of the target in (c) is no longer preattentive.

2. The graph is searched for relevant information, with focus shifting from the graph axes to the main part of the graph and back again (pattern recognition)
3. Once information is found, the focus shifts between the important part of the graph and the legend several times in order to keep the relevant information in working memory (conceptual relations produce quantitative meaning from visual features)
4. The question is answered and the participant moves on to another task (the question is related to the encoded quantitative features)

These steps are illustrated in figure 17.

Question: What is the relationship between the length of the eruption and the time between eruptions for Old Faithful?



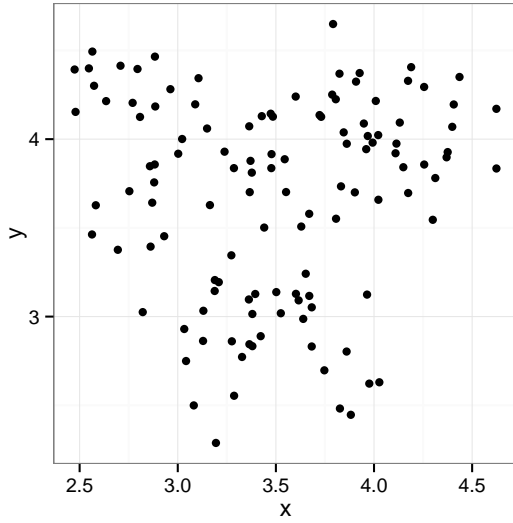
Sample mental steps:

1. Understand the question, identify “length of eruption” and “time between eruptions” as things to search for in the graph.
2. Look for “length of eruption” on the axes, determine that the y coordinate contains that information. Look for “time between eruptions” on the axes and determine that the x coordinate contains that information. Verify that these quantities are indeed what is being sought by re-reading the question.
3. Establish that as the time between eruptions increases, the length of the eruption increases. Note that there seems to be a bimodal distribution of points.
4. Answer the question: As the time between eruptions increases, the length of the eruption seems to increase.

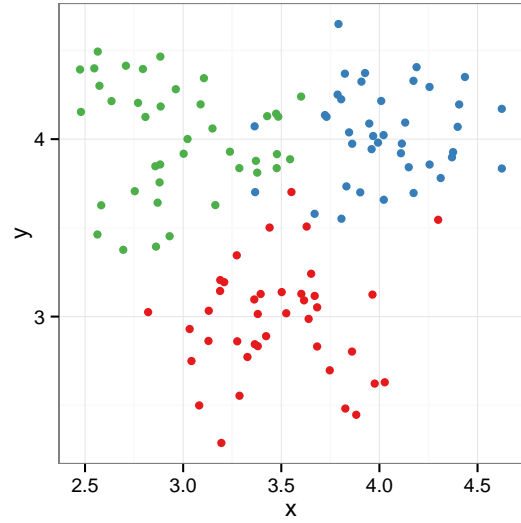
Figure 17: Task analysis of a simple graph-based task. The graph shows the length of the eruption of Old Faithful as a function of the waiting time between eruptions, with a corresponding sample “mental dialog” of the perceptual tasks involved in answering the question in response to the graph, according to [Shah and Miyake \(2005\)](#).

Working within this task-oriented framework, researchers have explored the “search” portion of the task-based model, the information integration portion of the model, and the types of graphs which facilitate both the “search” and “integration” portions of the task. Integration models modify the above sequence to allow for more complex graphical relationships to be assimilated, such that a viewer cycles between stages (2) and (3) several times in order to encode different portions of the graph. The time required for each of these steps may also change in accordance with the reader’s familiarity with the task and graphic style; those who are more familiar with similar graphics may be able to encode information faster and in larger chunks and thus answer the question more quickly ([Carpenter and Shah, 1998](#)).

Analyzing graphs using task-based models emphasizes the importance of spatial relationships between graphical elements. The gestalt laws of proximity and similarity dictate that items which are close together or physically similar (the same shape or color) are perceived as a group; this spatial perception creates “chunks” of the graph which may be encoded as single objects and thus reduce the mental bandwidth necessary to process the image. Figure 18 shows the advantage of “chunks” in graphs, as the second graph shown is much easier to describe and understand than the first graph, even without the contextual meanings of the variables.



(a) Proximity-based grouping



(b) Proximity and Similarity-based grouping

Figure 18: The utility of chunking in graph perception. Graph (a) could potentially be described as three distinct clusters of points rather than 120 individual points, but it is much easier to draw conclusions from graph (b), which has colored points that clearly show the grouping structure in the data. The second figure would more probably be encoded and described as three groups of 40 points, which serves as a form of mental data compression.

Of graphics that present information of similar complexity, graphics that require less effort to understand and search for relevant information are preferable (Cleveland and McGill, 1985). More complex models of the graphical perception process suggest that data are integrated on a visual level and then further integrated on a cognitive level, to form successive clusters of information. Once these clusters are formed, further information can be integrated by comparing and contrasting different clusters to understand the higher-level meaning in the graph (Ratwani et al., 2008). Graph types which cater to this hierarchical clustering mechanism may be more easily understood by viewers than graphs that do not provide information in a manner easily assimilated by the human brain. Based on this information, facets of graphs may be particularly useful for mapping multidimensional data to provide “chunks” of information in a relevant manner that can be integrated into the viewer’s working conceptual understanding of the dataset. Additionally, color schemes and appropriate labeling of graph features which reduce the amount of work necessary to integrate numerical information from a legend into the visual representation of the graph facilitate graphical inference (Carpenter and Shah, 1998).

From a statistical perspective, much of the literature involved in understanding graph perception from a task-analysis point of view focuses on simple graphics, such as side-by-side bar graphs or line graphs, and straightforward tasks of reading information from the graph accurately, rather than examining model assumptions or making inference beyond the data. The psychological mechanisms involved in processing simple graphics are perceptual, and typically require direct comparisons rather than mental manipulation in order to satisfy the tasks posed by the researchers (Trickett and Trafton, 2006). More complicated graphics and more sophisticated tasks may require comparison of two or more distinct graphs and may utilize working memory and spatial reasoning; these situations are not as well studied (Shah and Miyake, 2005). We begin first with the simple tasks of graph comprehension, and will summarize work with more complicated graphics at the end of this section.

2.2.2 Perception of Simple Graphs

A series of experiments by [Cleveland and McGill \(1984, 1985\)](#) studied basic perceptual tasks in graphical perception to produce a relative ordering of graphical elements by the accuracy of participant conclusions. This ranking is shown in Table 1. Other researchers ([Kosslyn, 1994](#)) have collapsed this ranking into position/length/angle and area/volume, as the difference in accuracy between categories 1, 2, and 3 is small compared to categories 4 and 5.

Rank	Task
1	Position (common scale)
2	Position (non-aligned scale)
3	Length, Direction, Angle, Slope
4	Area
5	Volume, Density, Curvature
6	Shading, Color Saturation, Color Hue

Table 1: [Cleveland and McGill \(1984, 1985\)](#) ordering of graphical tasks by accuracy (adapted from both papers and [Shah and Miyake 2005](#)). Higher-ranking tasks are easier for viewers than low-ranking tasks and should be preferred in graphical design.

The particular task required of participants in experiments also has an effect; [Simkin and Hastie \(1987\)](#) found that readers were more accurate in determining position when presented with a bar graph, but when readers were presented with a pie chart, they were more accurate at determining proportional judgments (using angles). This finding contradicts [Cleveland and McGill \(1984\)](#) to some degree and suggests that the experimental design and specific task are important in evaluating these sorts of user studies; the contradictory results also suggest that graph type is an important influence in determining what information viewers encode from the graph. This conflict also illustrates that the user’s attention and past experience influence the judgments they make from a given graph: when participants were asked to provide a summary of the graphic, their answers depended on the type of display: bar charts elicited a comparison judgment, pie charts elicited proportional judgments ([Simkin and Hastie, 1987](#)). Similarly, when presented with a line graph, viewers are more likely to summarize the graph in terms of the slope of the trend line (even when the x-axis is discrete); when presented with a bar graph, viewers summarize the information using discrete comparisons ([Carswell and Wickens, 1987](#); [Shah and Miyake, 2005](#)). The task and the graph format interact to influence viewer perceptions, thus, when creating graphics, statisticians should match appropriate graphical formats to meaningful conclusions about the data.

The task requirements are mediated by the limits of human processing ability. Chernoff faces, once proposed for visualizing multidimensional data, are difficult to read because viewers are unable to store the legend and the image in working memory; comparisons must be made serially and with conscious attention ([Shah and Miyake, 2005](#)), and data features may not map well to relatable facial features ([Lewandowsky and Spence, 1989b](#)). Similarly, while color does not generally correspond to precise quantitative information, certain color schemes utilizing hue, saturation, and brightness together can provide an implicit numerical ordering that does not place exceptional demands on working memory ([Shah and Miyake, 2005](#)). Color schemes which correspond to everyday situations (e.g. using a blue to red scale for low to high temperatures) may also reduce the demand on working memory (though such a scale may be problematic due to principles of color perception). While specific numerical judgments would still require selective attention, the “gist” of a graph using such schemes may be understood fairly quickly.

It is important to consider working memory when constructing graphical scales (particularly when utilizing a discrete scale for categorical data), but it is also important to consider feature selection and discriminability as well. Color is generally believed to be preferable for representing strata on a scatterplot ([Cleveland and McGill, 1984](#)), but [Lewandowsky and Spence \(1989a\)](#) found that if color is not available or appropriate, shapes, intensity, or discriminable letters may be utilized without a significant decrease in accuracy. Discriminable letters are those which do not share physical features such as closure and symmetry,

such as the letters H, Q, and X; confusable letters, such as H, E, and F, are associated with significantly less accurate perception. Demiralp et al. (2014) synthesized experimental evaluations of stimuli to create “perceptual kernels” describing the perceived distance between values; multidimensional scaling of the resulting distance matrix produces four distinct groups of shapes which share features (triangles with various degrees of rotation, squares and diamonds, and non-convex shapes such as x, +, and *). This separation suggests that feature integration underlies many of the processing speed effects found in studies examining discrete palettes and scales: palettes which are composed of confusable shapes, letters, or colors will require more processing time (and decrease accuracy) compared with discriminable palettes.

Other graph features can also influence viewer inferences: multiple studies suggest that our mental schematic for a graph is most consistent with a 45° trend line (Cleveland et al., 1988; Tversky and Schiano, 1989). “Banking to 45°” is a commonly-cited recommendation for optimal graphics (it is also quite old, according to Wickham (2013)), and does have some limited utility in reducing the strength of the line-width illusion (a more thorough discussion of this heuristic is provided in VanderPlas and Hofmann (2014)). Axis scale transformations can make it easier for viewers to spot outliers of data conforming to skewed distributions (though this does require some domain-specific knowledge of statistical distributions), and appropriately labeled graphs can reduce the working memory requirements by reducing the number of “back-and-forth” comparisons required to pass information into working memory (Shah and Miyake, 2005).

While graph perception is commonly limited by working memory considerations, there is some evidence that we perceive and process graphical information differently than numerical algorithms: Bobko and Karren (1979) found that participants underestimated correlation coefficients when estimating correlation strength, particularly under unequal variance; their estimates were in fact more closely aligned with r^2 . In addition, visual estimation tends to discount the effect of perceived outliers, producing a more robust estimator than numeric estimators designed for that purpose (Lewandowsky and Spence, 1989b). Finally, some evidence suggests that when visually estimating lines of best fit, we fit the slope of the first principal component rather than the least squares regression line; that is, we consider variability in x and y rather than only considering variability in y (Mosteller et al., 1981). While these studies do not offer theoretical explanations or attempt to provide causal explanations, human perception of data displays does appear to differ from computational exploration in meaningful ways.

These studies indicate that it is important to consider the cognitive processing of statistical graphics as well as the data used to generate these graphics: the type of graph, color scheme, annotations, aspect ratio, legends, and axis transformations can all influence the amount of mental processing required to draw conclusions from a graph, as well as the types of conclusions that graph viewers are likely to draw. Many of these features were studied in relative isolation, using simple graphs that may lack real-world context. More complex, domain specific graphs may create higher cognitive load and recruit previously acquired knowledge; experiments using simple, bland graphics may not be applicable to more complex graphics meant for experts. What follows is a summary of the relatively sparse literature on these sorts of real-world graphics.

2.2.3 Perception of Complex, Domain-Specific Graphs

Carpenter and Shah (1998) showed that graph comprehension time increased when the number of distinct x-y functions (i.e. nonparallel sloped lines) increased, even if the same data was represented. The density of these functions also had an impact: dense graphs with multiple intersecting trend lines took more time to interpret than dense graphs with parallel trends or sparse graphs with intersecting trend lines. This supports the idea that the information conveyed in the graph must be read into working memory before the graph can be described or used for inference; more complex graphs would take more time to understand and internalize. Additional factors can also influence the ease with which graphs are perceived and understood in real-world scenarios. Gattis and Holyoak (1996) found that graphs were more accurately perceived when the dependent variable was on the y axis and the independent variable was on the x axis, even when the perceived IV and DV were manipulated using a cover story. In even more complex visualizations, Trickett and Trafton (2006) found that meteorologists and other domain experts would mentally superimpose graphs from memory on visible graphs, utilizing spatial processing rather than manipulating a physical interface. These interactions demonstrated complex spatial manipulation to assimilate information from multiple graphs,

particularly when the information provided in the graphs conflicted with prior information, either from the meteorologist’s own domain knowledge or verbal information provided during the course of the study. While the procedures used in this study rely on verbal descriptions of mental processes (i.e. the meteorologist speaking aloud as they process each graph and map to assimilate information), the evidence is sufficient to suggest that in addition to working memory and the visual processing performed by the brain, some complex graphs also utilize spatial processes (and the corresponding brain regions) to perform complicated overlays and mental transformations. By designing such complicated graphs to more easily facilitate such mental operations, it is possible that more effective spatial visualizations could make these graphs more accessible.

Complicating the research into more complex graphs, there are many different types of complexity that can affect graphics. There may be differences in how processing occurs for large amounts of data, but it could also be that more complex x-y relationships could also require more mental effort. In addition, multiple relationships can be depicted simultaneously, either because of underlying groups in the data or because multiple related trends are depicted on the same graph (though this is widely acknowledged as bad practice in statistical graphics). Finally, the mental complexity of the task required of the graph viewer can also factor into the amount of time and effort required to complete a task using a graph. These different types of complexity interact with the graph format; for instance, line graphs are less effected by increasing complexity than bar graphs (Tan, 1994), and bar graphs are more affected by increasing complexity than pie charts for ratio judgments (Hollands and Spence, 1998).

Finally, complex graphs often facilitate different types of participant tasks; rather than simple numerical judgments or information lookup, complex graphs may encourage (or require) viewers to use prior knowledge and interpretation skills. These additional complications make experimental study of complex or domain-specific graphs more difficult. Many of these problems (types of complexity, expanded tasks, prior knowledge) make further work in this area somewhat difficult. One concept that facilitates studies examining the relationship between complex data and graph format in statistical graphics is the *grammar of graphics*, which we discuss in the next section, along with other experimental methodology useful for understanding how people perceive statistical graphics.

3 Testing Statistical Graphics

3.1 Basic Psychophysics Methodology

Psychophysics studies are generally concerned with the ability to detect a stimulus (or a difference between two stimuli). Many classic psychophysical methods are still used in studies today (VanderPlas and Hofmann, 2014). Several of these methods are mentioned here; for a more thorough review, see Goldstein (2009).

Method of Limits The method of limits seeks to determine the level of intensity at which a stimulus is just barely detectable. A series of trials is used, with each trial starting at either the lower or upper range of intensity and incrementally moving towards the opposite end of the range; for each point, the observer indicates whether they can detect the stimulus. At the end of several trials, the detection limits are averaged to produce a measured *absolute threshold*. Figure 19 demonstrates this process.

Method of Adjustment This method is similar to the method of limits, except that the stimulus intensity is adjusted by the observer (not the experimenter) in a continuous manner until the observer can just barely detect the stimulus. This procedure may be repeated several times, with trials averaged to produce a mean value for the absolute threshold. Figure 20 demonstrates this procedure.

Measuring the Difference Threshold The difference threshold, discussed in section 1.3.1, is the smallest detectable difference between two stimuli. This threshold can be measured using either the method of limits or the method of adjustment. Rather than increasing the absolute intensity of the stimulus as discussed above, two stimuli are given: one with constant intensity and one whose intensity may vary either continuously or incrementally (depending on the method utilized). The participant is instructed to identify the point at

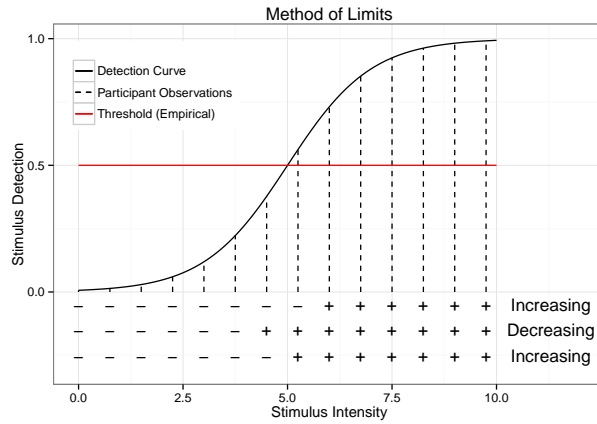


Figure 19: Demonstration of the method of limits. In this experiment, three trials were performed, two trials starting from 0 and increasing, and one trial starting at 9.5 and decreasing. The empirical detection threshold is the threshold at which detection occurs 50% of the time, and is shown in red.

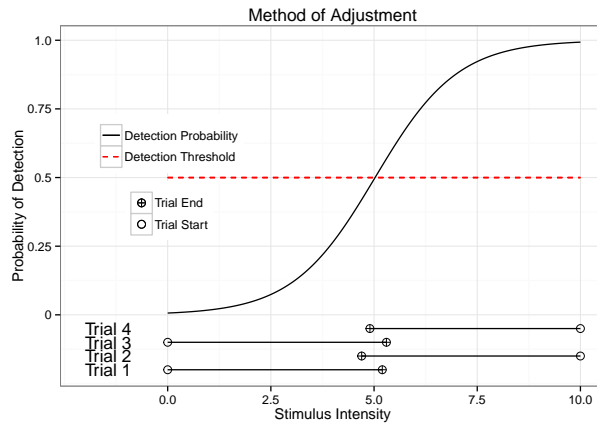


Figure 20: Demonstration of the method of adjustment. In this experiment, four trials were performed, two trials starting from 0 and increasing, and two trials starting at 10 and decreasing. The empirical detection threshold is the threshold at which detection occurs 50% of the time, and is shown in red.

which the two stimuli are indistinguishable (if the varied stimulus is approaching the constant stimulus) or the point at which the two stimuli are distinguishable (if the varied stimulus is diverging from the constant stimulus).

Magnitude Estimation Magnitude estimation studies are used to measure the perceptual intensity of two different stimuli. For example, a participant might be shown a series of two lights, and asked to assign a number to describe how bright each light is. These numerical values would then be compared to the actual light intensity (as measured by a digital sensor or by the input voltage) to determine how perceived brightness corresponds to actual intensity. Many stimuli measured this way produce power-law functions that exhibit response compression (doubling the actual intensity corresponds to a much smaller change in perceived brightness) or response expansion (doubling the actual intensity corresponds to a change in perceived intensity that is more than double the original intensity).

3.2 Testing Images using Psychological Paradigms

In addition to the psychophysics methods outlined above, there are testing paradigms within psychology that are applicable to the study of statistical graphics. These include experimental methods such as visual search and eye tracking, as well as experimental control procedures that may be important in graphics studies that are similar to cognitive psychology studies. We will first discuss the experimental methods, and then briefly discuss some common control procedures that may be applicable to the study of statistical graphics.

3.2.1 Experimental Methods

Many psychological experiments utilize straightforward methods to address hypotheses in perception, such as asking participants to make numerical judgments based on presented stimuli. These methods are quite useful, but not particularly difficult or domain-specific. In this section, we discuss two domain-specific methods for understanding perception of visual stimuli: visual search and eye tracking.

Visual Search Simply put, visual search methods involve presenting a participant with many distractor stimuli and one or more target stimuli, and asking the participant to locate the target stimuli. Time is measured between the initial stimulus presentation and the participant’s answer; participant accuracy is also considered in more complicated visual search settings. This procedure allows researchers to measure simple preattentive stimuli and can also be utilized for more complicated tasks that require attention ([Anderson and Revelle, 1983](#)). One example of a visual search task is shown in figure 21; a common preattentive search task is shown in figure 16c.

Visual search tasks can be used to measure the efficiency of a participant’s visual search abilities (and focus on a task) to serve as a baseline for more complicated visual tasks. They can also be used to examine feature binding and common mistakes that may indicate relevant distractor stimuli. Even when reaction time is not directly measured, these tasks are typically given under time pressure, to establish a baseline performance that is below 100% performance on the task. This time pressure allows experimenters to avoid response compression, so that the number of questions completed within the time limit provides an approximate measure of response time.

Eye Tracking Eye tracking studies are often utilized in order to understand which parts or features of an image participants focus on, and in what order they examine the image components. Eye tracking studies were heavily used in order to refine the task-based models of graphical perception; they allowed researchers to understand that participants had to iterate between the question and different parts of the graph in order to assimilate all of the represented information into working memory. Figure 22 shows one lightweight eye-tracking assembly. The camera allows researchers to track the direction of the pupil and thus infer gaze direction.

Eye tracking studies have been performed on statistical graphics as well ([Zhao et al., 2013](#)), utilizing a visual search task and examining which graphics participants compared to determine the target plot.

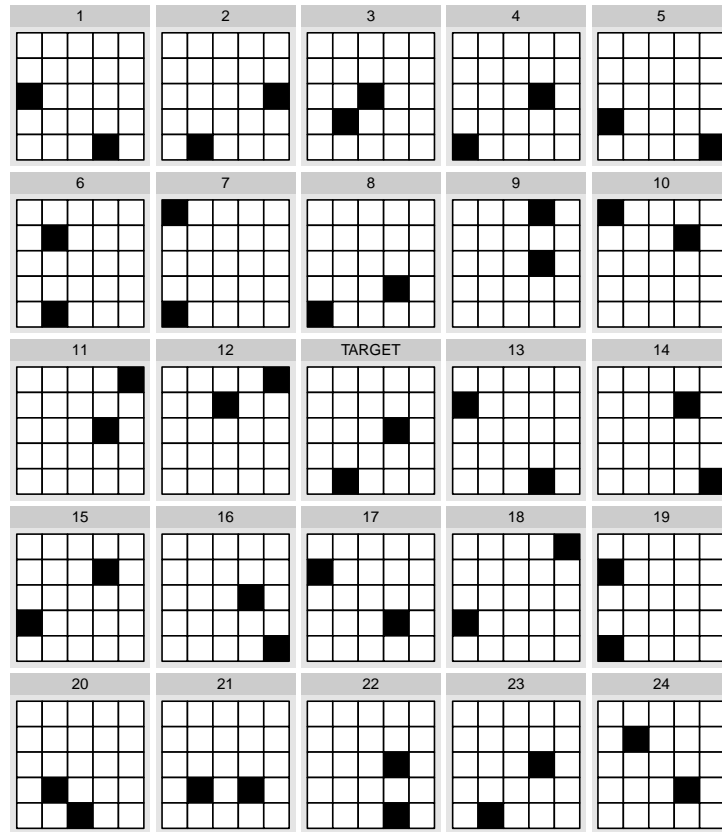


Figure 21: Visual Search Task. Participants were asked to locate the figure (1-24) most similar to the central “target” figure.

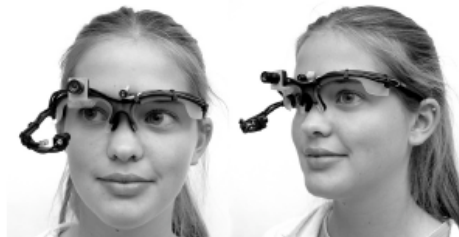


Figure 22: Eye tracking equipment (Babcock and Pelz, 2004). The cameras allow researchers to determine what part of a scene the wearer is viewing.

In some psychological experiments, the straightforward approach to a task can produce biased responses from participants. Human perception is highly reliant on expectations and past experience, and as a result, experimenters must take care to reduce undesired biasing effects in order to appropriately control experiments. Some of these considerations are discussed in the next section.

3.2.2 Experimental Control Procedures

While not all of the experimental control procedures discussed here are appropriate for every experiment, they do demonstrate the degree of control many experiments require to measure small psychological effects. The variation in the human brain and in cognitive strategies (and the sample size constraints of testing in humans) requires a large degree of experimental control in order to minimize the effects of population variance. Some of the biases of the human brain as well as strategies to address these biases are described below.

Habituation The human visual system is attracted to novelty; odd, bizarre, or new sights attract more attention than ordinary, run of the mill scenes. Habituation describes the process of becoming less interested in a stimuli; as this occurs, the mind begins to enter “auto-pilot” and attention to the task at hand becomes less focused. In infants, this habituation process is used to determine whether there is a perceived difference between two stimuli; in adults, this process is not typically as useful to the experimenter. To avoid habituation, experiments should generally consist of somewhat varied tasks in order to maintain participant attention. The psychophysics experiments described in figures 19 and 20 can be vulnerable to this problem; to overcome habituation, trials often alternate in direction or start at random points along the intensity spectrum.

Masking Images can persist on the retina for a period after the image is no longer available; this phenomenon is called *persistence of vision*. In order to control the time in which the stimulus is visible, psychological experiments often will show a mask immediately after an image in order to “erase” the retina. This degree of control is often useful in experiments which focus on the preattentive stage of perception, but persistence of vision is not likely to affect experiments which take place in the attentive stage of perception (i.e. images shown for more than .5 seconds). A sample mask is shown in figure 23.

```
## Warning: Removed 200 rows containing missing values (geom_point).
```

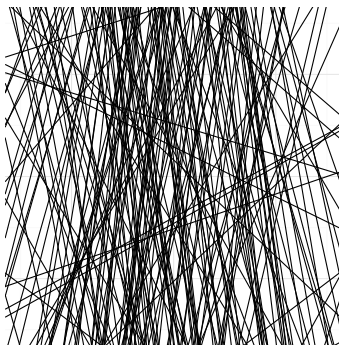


Figure 23: Sample mask used to “erase” the retina in some psychological experiments which test stimuli for short time periods (<1 s). The mask removes any afterimage the participant might have, ensuring that the stimuli is only available for the specified period.

Priming Broadly, priming is a technique that can be used to subconsciously bias a participant towards a certain conclusion. In cognitive psychology research, priming can be used to test word association (i.e. participants are quicker to identify an apple if they have just heard the word “fruit” than if they heard the unrelated word “pen”); in statistical graphics, priming effects are more likely to occur due to instructions or examples provided to participants at the start of a testing session. If an initial example contains notable outliers, participants are more likely to look for and recognize graphs with outliers than graphs with other notable features. Examples must be designed in such a way to avoid activating these priming affects as much as possible.

There are many other psychological mechanisms that may impact participant performance; the mechanisms presented here are simply some of the more salient considerations in experimental design for statistical graphics.

3.3 Testing Statistical Graphics

This section details the tools specific to the testing of graph perception within the field of statistics. [Cleveland and McGill \(1985\)](#) studied statistical graphics from a largely psychological perspective, but their findings have been widely utilized in the field of statistics; however, it has been 20 years since that paper, and the field has developed within statistics since that time. Two major developments, the grammar of graphics and the lineup protocol, are particularly important for future research into the perception of statistical graphics.

3.3.1 The Grammar of Graphics

The grammar of graphics, detailed in [Wilkinson et al. \(2006\)](#), is a framework for describing a graphic in terms of its basic component pieces. An implementation of the grammar of graphics for R, `ggplot2` ([Wickham, 2009, 2010](#)), provides a useful tool for manipulating graphics to test in an experimental setting. Using the grammar of graphics, it is easy for experimenters to compare different types of charts using the same data, as the underlying structure of the graph remains the same. Figure 24 shows three plots created using the same data and different geometric objects, and figure 25 provides the `ggplot2` code to create the plots ⁴. Comparing these graphics experimentally would be reasonably simple, and the grammar of graphics helps to control the extraneous variables introduced by utilizing different plot types. In addition, the grammar of graphics approach to transformations and scales allows us to easily test judgments made utilizing different axis transformations and color scales to compare perceptual accuracy ([Hofmann et al., 2012](#)).

3.3.2 Testing Statistical Graphics using Lineups

One useful tool for testing statistical graphics is the concept of a lineup. Lineups combine the psychological notion of visual search tasks with the statistical concept of hypothesis testing: Participants are provided with a number of plots of the same form, one using real data and the rest generated using resampling methods. If participants identify the target plot (the plot with real data), this is considered similar in nature to a significant hypothesis test at a given α level (generally, there are 20 plots, so $\alpha = 0.05 = 1/20$). Figure 26 shows a sample lineup.

In addition to the visual inference protocols lineups were designed to fulfill ([Buja et al., 2009](#)), they also provide a method to easily quantify (on a statistical level) the “power” of a plot; if two lineups are generated from the same data, but one allows participants to more frequently detect the target plot, then that lineup provides more perceptual power. The lineup protocol provides a useful tool for examining some of the issues discussed for complex, domain specific graphs. When combined with the grammar of graphics approach ([Wickham et al., 2010](#)), lineups have the potential to be extremely useful for studying the perception of graphs which present the same data in different forms.

⁴These plots are terrible from a psychological perspective, but serve to illustrate the versatility of the grammar of graphics. In general, stacked density plots, histograms, and dot plots are bad for making numerical comparisons ([Cleveland and McGill, 1985](#)).

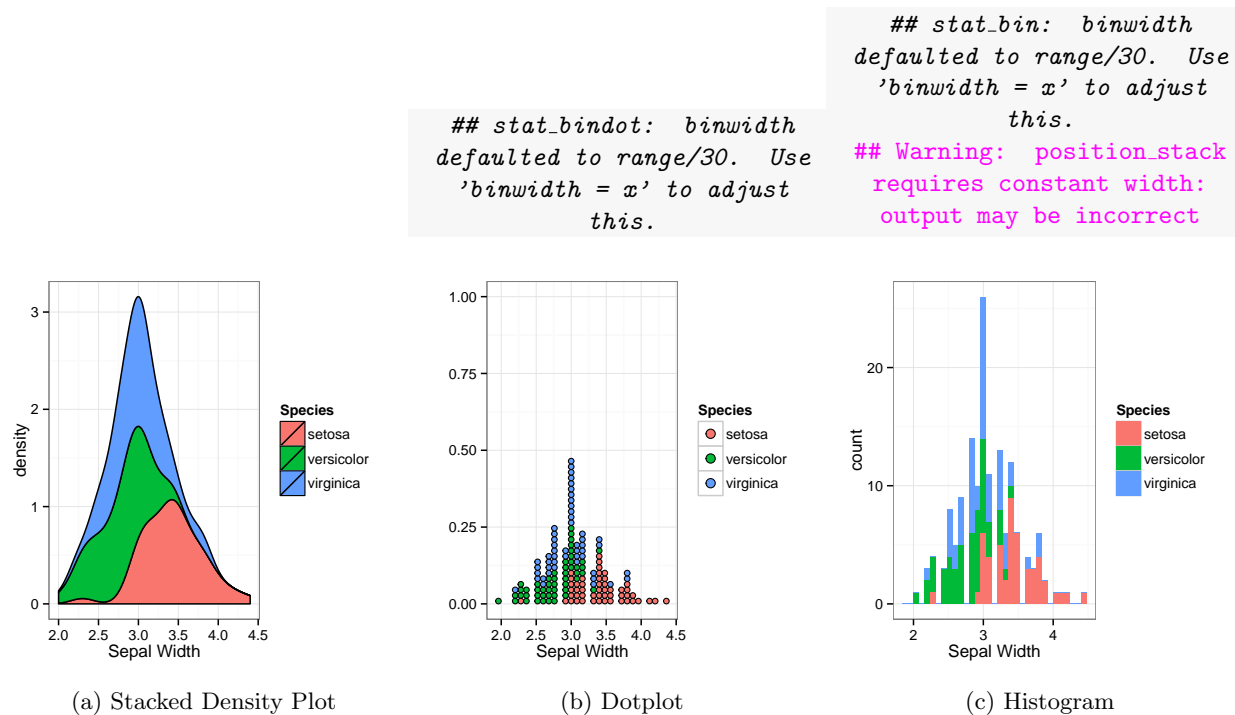


Figure 24: Three different plots of iris data, created using the grammar of graphics

```
# Stacked Density plot
ggplot(data=iris, aes(x=Sepal.Width, fill=Species)) +
  geom_density(position="stack")

# Dotplot
ggplot(data=iris, aes(x=Sepal.Width, fill=Species)) +
  geom_dotplot(method='histodot', stackgroups=TRUE)

# Histogram
ggplot(data=iris, aes(x=Sepal.Width, fill=Species)) +
  geom_histogram(position="stack")
```

Figure 25: ggplot2 code to produce figure 24

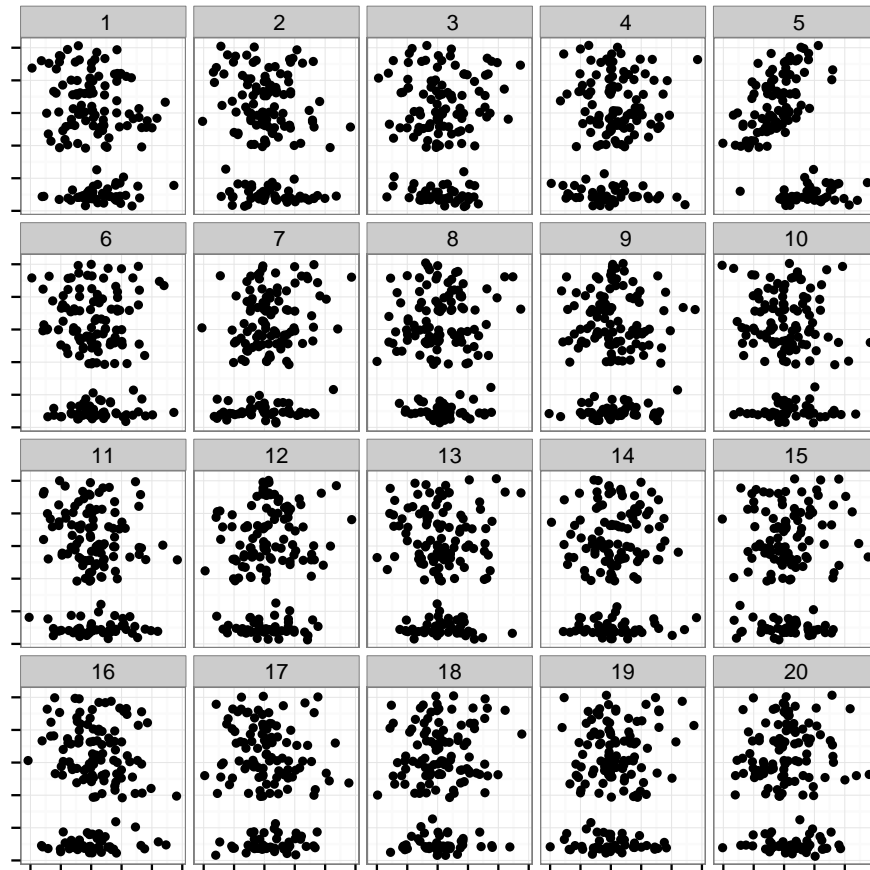


Figure 26: Lineup of the iris data, comparing sepal width to petal width. The target data is in plot 5, other plots generated by permuting petal width.

References

- Amer, T. (2005). Bias due to visual illusion in the graphical presentation of accounting information. *Journal of Information Systems*, 19(1):1–18.
- Anderson, K. J. and Revelle, W. (1983). The interactive effects of caffeine, impulsivity and task demands on a visual search task. *Personality and Individual Differences*, 4(2):127–134.
- Babcock, J. S. and Pelz, J. B. (2004). Building a lightweight eyetracking headgear. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ETRA '04, pages 109–114, New York, NY, USA. ACM.
- Bobko, P. and Karren, R. (1979). The perception of pearson product moment correlations from bivariate scatterplots. *Personnel Psychology*, 32(2):313–325.
- Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D. F., and Wickham, H. (2009). Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4361–4383.
- Carpenter, P. A. and Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2):75.
- Carswell, M. C. and Wickens, C. D. (1987). Information integration and the object display an interaction of task demands and display superiority. *Ergonomics*, 30(3):511–527.
- Cleveland, W. S., McGill, M. E., and Robert, M. (1988). The Shape Parameter of a Two-Variable Graph. *Journal of the American Statistical Association*, 83(402):289–300.
- Cleveland, W. S. and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):pp. 531–554.
- Cleveland, W. S. and McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716):828–833.
- Conti, G., Ahamad, M., and Stasko, J. (2005). Attacking information visualization system usability overloading and deceiving the human. In *Proceedings of the 2005 symposium on Usable privacy and security*, pages 89–100. ACM.
- Coren, S. and Porac, C. (1983). Subjective contours and apparent depth: A direct test. *Attention, Perception, & Psychophysics*, 33(2):197–200.
- Demiralp, C., Bernstein, M., and Heer, J. (2014). Learning perceptual kernels for visualization design.
- Fisher, G. H. (1970). An experimental and theoretical appraisal of the perspective and size-constancy theories of illusions. *The Quarterly journal of experimental psychology*, 22(4):631–652.
- Gattis, M. and Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1):231.
- Goldstein, E. B. (2009). *Sensation and Perception*. Thomson Wadsworth, Belmont, CA, eighth edition.
- Green, R. and Hoyle, E. (1963). The poggendorff illusion as a constancy phenomenon. *Nature*, 200:611–612.
- Gregory, R. (1968). Perceptual illusions and brain models. *Proc. Roy. Soc. B*, 171:279–296.
- Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, 199(678-91):1.

- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358):1121–1127.
- Healey, C. G. and Enns, J. T. (1999). Large datasets at a glance: Combining textures and colors in scientific visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 5(2):145–167.
- Healey, C. G. and Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *Visualization and Computer Graphics, IEEE Transactions on*, 18(7):1170–1188.
- Helander, M. G., Landauer, T. K., and Prabhu, P. V. (1997). *Handbook of human-computer interaction*. Elsevier.
- Hofmann, H., Follett, L., Majumder, M., and Cook, D. (2012). Graphical tests for power comparison of competing designs. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2441–2448.
- Hofmann, H. and Vendettuoli, M. (2013). Common Angle Plots as perception-true visualizations of categorical associations. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2297–2305.
- Hollands, J. and Spence, I. (1998). Judging proportion with graphs: The summation model. *Applied Cognitive Psychology*, 12(2):173–190.
- Kosslyn, S. M. (1994). *Elements of Graph Design*. W.H. Freeman and Company.
- Kosslyn, S. M., Ball, T. M., and Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of experimental psychology: Human perception and performance*, 4(1):47.
- Lewandowsky, S., Herrmann, D. J., Behrens, J. T., Li, S.-C., Pickle, L., and Jobe, J. B. (1993). Perception of clusters in statistical maps. *Applied Cognitive Psychology*, 7(6):533–551.
- Lewandowsky, S. and Spence, I. (1989a). Discriminating strata in scatterplots. *Journal of the American Statistical Association*, 84(407):682–688.
- Lewandowsky, S. and Spence, I. (1989b). The perception of statistical graphs. *Sociological Methods & Research*, 18(2-3):200–242.
- Light, A. and Bartlein, P. J. (2004). The end of the rainbow? color schemes for improved data graphics. *Eos, Transactions American Geophysical Union*, 85(40):385–391.
- Matlin, M. W. (2005). *Cognition*. Wiley, 6th edition.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81.
- Morgan, M. (1999). The pogendorff illusion: a bias in the estimation of the orientation of virtual lines by second-stage filters. *Vision Research*, 39(14):2361 – 2380.
- Morgan, M., Adam, A., and Mollon, J. (1992). Dichromats detect colour-camouflaged objects that are not detected by trichromats. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 248(1323):291–295.
- Mosteller, F., Siegel, A. F., Trapido, E., and Youtz, C. (1981). Eye fitting straight lines. *The American Statistician*, 35(3):150–152.
- Poulton, E. (1985). Geometric illusions in reading graphs. *Perception & psychophysics*, 37(6):543–548.
- Ratwani, R. M., Trafton, J. G., and Boehm-Davis, D. A. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied*, 14(1):36.

- Shah, P. and Miyake, A. (2005). *The Cambridge handbook of visuospatial thinking*. Cambridge University Press.
- Shepard, S. and Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1):3.
- Silva, S., Sousa Santos, B., and Madeira, J. (2011). Using color in visualization: A survey. *Computers & Graphics*, 35(2):320–333.
- Simkin, D. and Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82(398):454–465.
- Sun, J. Z., Wang, G. I., Goyal, V. K., and Varshney, L. R. (2012). A framework for bayesian optimality of psychophysical laws. *Journal of Mathematical Psychology*, 56(6):495–501.
- Tan, J. K. (1994). Human processing of two-dimensional graphics: Information-volume concepts and effects in graph-task fit anchoring frameworks. *International Journal of Human-Computer Interaction*, 6(4):414–456.
- Treinish, L. A. and Rogowitz, B. E. (2009). Why should engineers and scientists be worried about color?
- Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31(2):156 – 177.
- Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136.
- Trickett, S. B. and Trafton, J. G. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In *Diagrammatic representation and inference*, pages 286–300. Springer.
- Tukey, J. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.
- Tversky, B. and Schiano, D. J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General*, 118(4):387.
- VanderPlas, S. and Hofmann, H. (2014). Signs of the sine illusion - why we need to care. *Journal of Computational and Graphical Statistics*.
- Varshney, L. R. and Sun, J. Z. (2013). Why do we perceive logarithmically? *Significance*, 10(1):28–31.
- Ward, L. M., Porac, C., Coren, S., and Girgus, J. S. (1977). The case of misapplied constancy scaling: Depth associations elicited by illusion configurations. *American Journal of Psychology*, 90(4):609–620.
- Westheimer, G. (2007). Irradiation, border location, and the shifted-chessboard pattern. *Perception*, 36(4):483.
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.
- Wickham, H. (2010). A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1):3–28.
- Wickham, H. (2013). Graphical criticism: some historical notes. *Journal of Computational and Graphical Statistics*, 22(1):38–44.
- Wickham, H., Cook, D., Hofmann, H., and Buja, A. (2010). Graphical inference for infovis. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):973–979.
- Wilkinson, L., Wills, D., Rope, D., Norton, A., and Dubbs, R. (2006). *The grammar of graphics*. Springer.
- Zhao, Y., Cook, D., Hofmann, H., Majumder, M., and Chowdhury, N. R. (2013). Mind reading: Using an eye-tracker to see how people are looking at lineups. *International Journal of Intelligent Technologies & Applied Statistics*, 6(4).