This is the title of a thesis submitted to Iowa State University

Note that only the first letter of the first word and proper names are capitalized

by

Susan Ruth VanderPlas

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:

Heike Hofmann, Major Professor

Di Cook

Sarah Nusser

Max Morris

Stephen Gilbert

Erin MacDonald

Iowa State University

Ames, Iowa

2014

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES v

CHAPTER 1. OVERVIEW 1

1.1 Goals of Statistical Graphics 1

1.2 The Human Visual System 1

1.2.1 Hardware 1

1.2.2 Software 4

1.2.3 Bugs and Peculiarities of the Visual System 9

1.2.4 Cognitive Load 18

1.3 Statistical Graphics 19

1.3.1 Low-level Perception of Statistical Graphics 19

1.3.2 Higher-level graphical perception 20

1.4 Testing Statistical Graphics 27

1.4.1 Basic Psychophysics Methodology 27

1.4.2 Testing Images using Psychological Paradigms 28

1.4.3 Testing Statistical Graphics 32

CHAPTER 2. SIGNS OF THE SINE ILLUSION – WHY WE NEED TO CARE 35

2.1 Introduction 35

2.1.1 The Sine Illusion in Statistical Graphics 38

2.1.2 Perceptual Explanations for the Sine Illusion 39

2.1.3 Geometry of the Illusion 41

2.2 Breaking the Illusion 43

2.2.1 Trend Removal 43

2.2.2 Transformation of the X-Axis 44

2.2.3 Transformation in Y 50

2.3 Transformations in Practice – a User Study 54

2.3.1 Study Design 56

2.3.2 Results 57

2.4 Application: US Gas Prices 62

2.5 Conclusions 64

CHAPTER 3. VISUAL REASONING AND STATISTICAL GRAPHICS . . 66

CHAPTER 4. STATISTICAL GRAPHICS AND FEATURE HIERARCHY 67

BIBLIOGRAPHY 68

LIST OF TABLES

1.1 Cleveland & McGill's ordering of graphical tasks 23

2.1 Sine Illusion Model: Fixed Effects 60

2.2 Sine Illusion Model: Random Effects 60

LIST OF FIGURES

1.1 The human eye, with closeup of receptor cells in the retina 2
1.2 Absorption spectra of retinal cells 2
1.3 Inhibition Illusions 3
1.4 Saccades and Pauses 5
1.5 Parallel and Serial Feature Detection 6
1.6 Rotation Task 8
1.7 Ambiguous Images 9
1.8 Colorblindness and Rainbow Color Schemes 12
1.9 Gestalt Illusions 14
1.10 Depth Illusions 15
1.11 Muller Lyer Real World Context 16
1.12 The Poggendorff Illusion 17
1.13 Cafe Wall Illusion 18
1.14 Preattentive Features and Interference 20
1.15 Chunking in Graphs 22
1.16 Visual Search Task 29
1.17 Eye Tracking Equipment 30
1.18 Three different plots of iris data, created using the grammar of graphics 32
1.19 ggplot2 code to produce Figure 1.18 33
1.20 Lineup for testing statistical graphics 34

2.1 Scatterplots of Ozone and Temperature in Houston, 2011 36
2.2 The original sine illusion 37

2.3	Trend, seasonality, and the sine illusion	38
2.4	Imports to and Exports from the East Indies in the 1700s	39
2.5	Three-dimensional origins of the sine illusion	40
2.6	Contextual origins of the sine illusion	41
2.7	Geometry of the sine illusion	42
2.8	Changing variability and nonlinear trend lines	44
2.9	X axis transformations	48
2.10	Original data and data after X transformation	49
2.11	General correction approach	50
2.12	Methods based on Approximations to f(x)	51
2.13	Quadratic Approximation	53
2.14	Screenshot of Data Collection Website	55
2.15	Overview of possible starting weights	57
2.16	Overcorrected transformations excluded from the analysis	58
2.17	Results from psychophysics analysis	59
2.18	Results from mixed model	61
2.19	US Gas prices from 1995 to 2014	62
2.20	Standard deviation of daily gas prices between 1995 and 2014.	62
2.21	Gas price data, X transformation	63
2.22	Gas price data, Y transformation	64

CHAPTER 1. OVERVIEW

1.1 Goals of Statistical Graphics

1.2 The Human Visual System

Basic overview of structure, to serve as a reference

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.

1.2.1 Hardware

The physiology of perception is complex; what follows is a high-level 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.

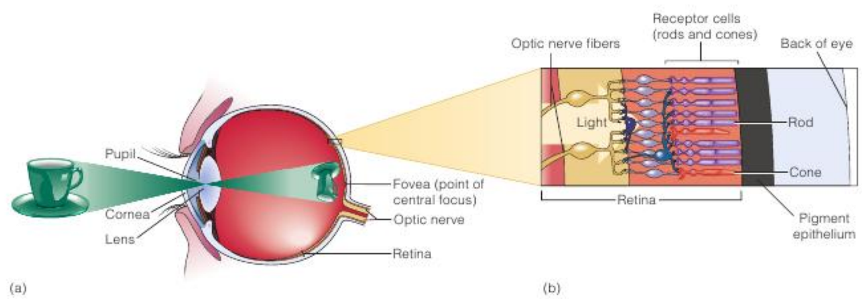


Figure 1.1: The human eye, with closeup of receptor cells in the retina (image from Goldstein 2009, chap 3.1).

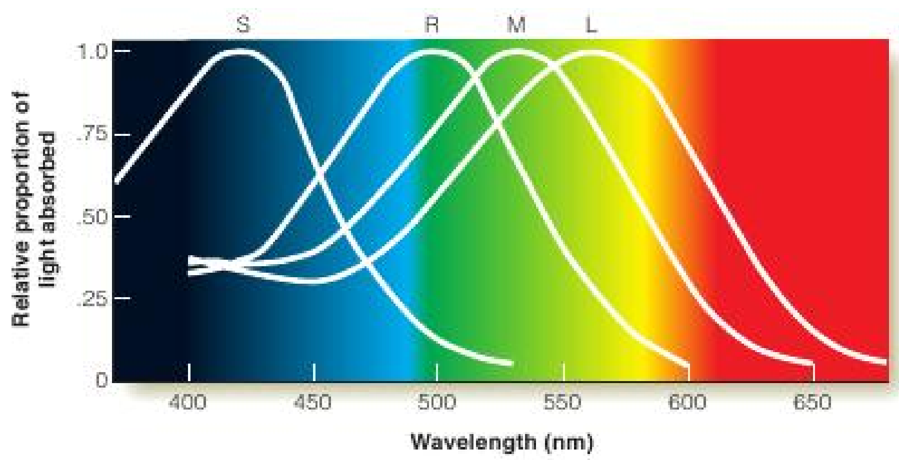
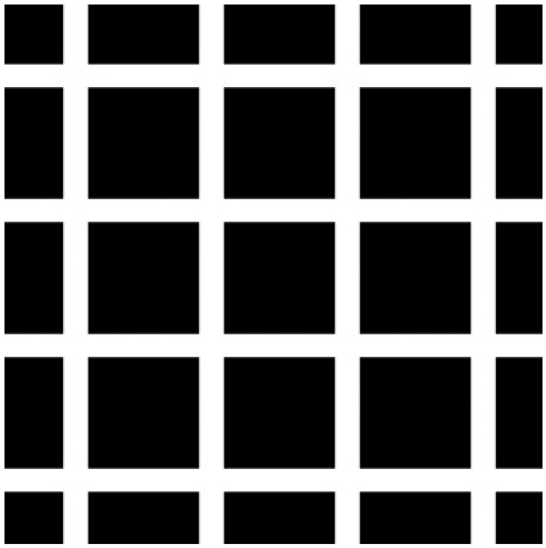


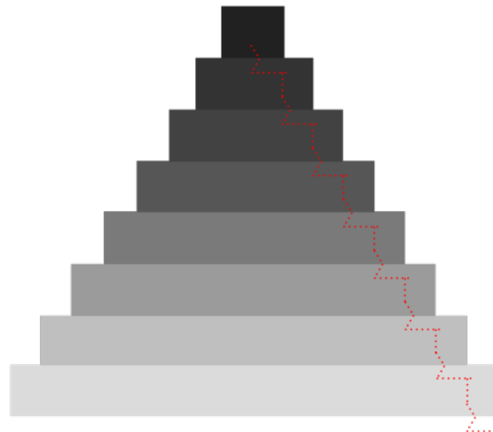
Figure 1.2: Absorption spectra of rods and short, medium, and long wave cones. (image from Goldstein 2009, chap 3.3).

The Eye The eye is a complex apparatus, but for our purposes, the primary 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 retina, known as the fovea, contains only cones; the rest of the retina contains a mixture of rods and cones. Figure 1.1 depicts the structure of the eye with a closeup of the retina.

Figure 1.2 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



(a) Hermann Grid Illusion



(b) Mach Bands.

Figure 1.3: 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.

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. 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 coloquially as color-blindness, and occurs in an estimated 5% of the population.

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 its' own neuron. The combined wiring of rod cells is responsible for the Hermann grid illusion and the Mach bands seen in Figure 1.3. Both of these illusions are a product of lateral inhibition, which is a result of the wiring of rod cells in the retina. The specifics of 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 cortext 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.

1.2.2 Software

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

1.2.2.1 Attention and Perception

In many tasks, it is necessary to pay attention to many different 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 1.4 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, which information is being encoded by the brain).

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"

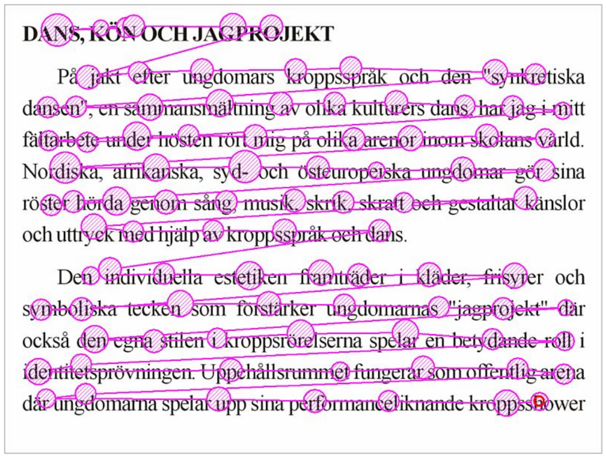


Figure 1.4: A plot of saccades made while reading text. Saccades, shown by the lines, indicate "jumps", while pauses are shown by circles, with size proportional to the time spent focusing on that area.

film[1] 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 of the basketball, they did not notice the obvious gorilla walking through the scene. 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

[1] http://www.theinvisiblegorilla.com/videos.html

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 1.5, taken from Chapter 6 of Helander et al. (1997). The importance of preattentive processing to statistical graphics is discussed in Section 1.3.1.

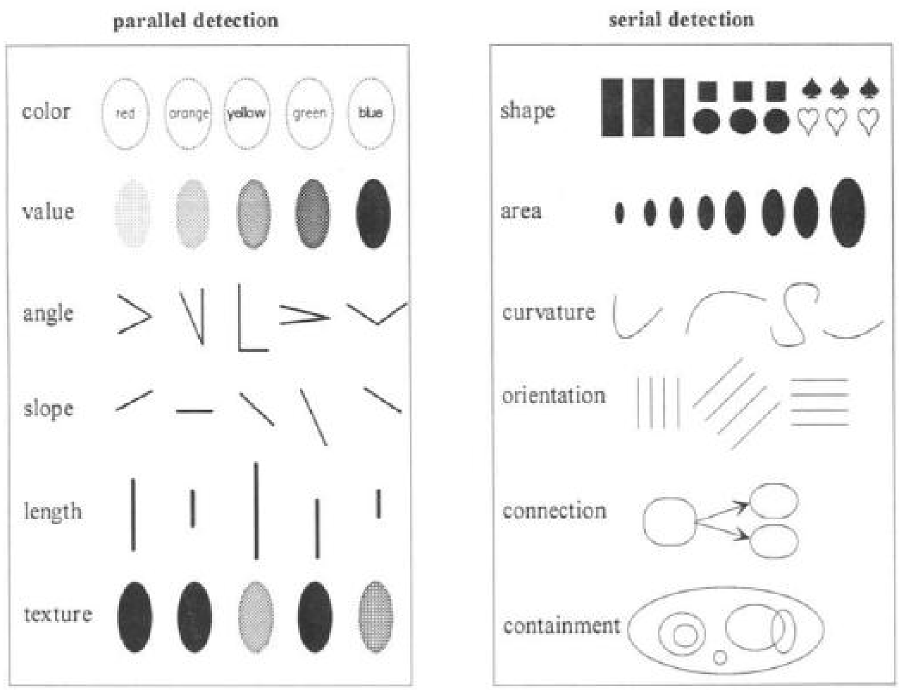


Figure 1.5: Examples of features detected serially or in parallel (Chapter 6, Helander et al. 1997)

Feature integration as a result of attention enables the brain to process a figure holistically. This processing is important for the most basic visual processes we take for granted, including object perception.

1.2.2.2 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). 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.

- **Pragnanz - the law of good figure**. 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.

1.2.2.3 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 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 accuracty 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 1.6. 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.

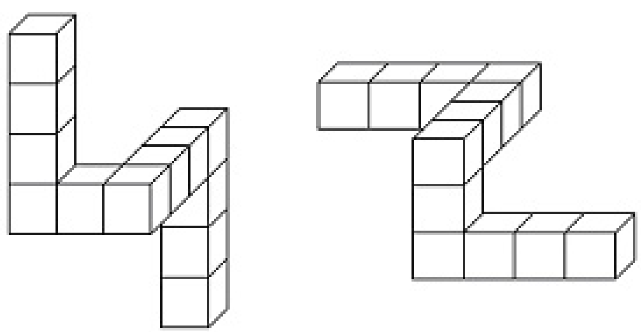


Figure 1.6: Rotation task (Shepard and Metzler, 1988). Are the two images the same?

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 this 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 (such as Figure 1.7) and asked to describe it initially, participants could not give an alternate interpretation of the figure after the experiment was complete. In the case of Figure 1.7, participants who initially said the figure was a duck could not describe the figure as a rabbit later, 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.

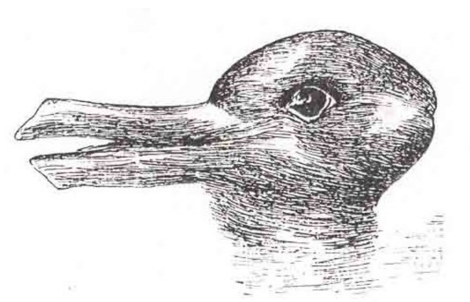


Figure 1.7: An ambiguous image that could be either a rabbit or a duck. When participants were asked to identify the image initially, they could not provide an alternate interpretation of the figure later.

The "software" of the visual system is of course more complex than the few programs listed here, but understanding attention, object perception concepts, 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.

1.2.3 Bugs and Peculiarities of the Visual System

We have discussed the neural "hardware" of the visual system and some of the higher-level processing 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. We 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.

1.2.3.1 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

$$\frac{\triangle S}{S} = K \tag{1.1}$$

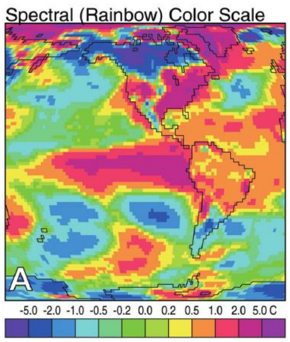
where K is a constant called the Weber fraction, S is the value of the standard stimulus, and $\triangle S$ is the difference between the standard stimulus and the test stimulus. So if a participant is given a 100-g weight and a 102-g weight and can just barely tell the difference between the two, then $K = 0.02$ and we would assume that the 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 as well (Varshney and Sun, 2013). In some sensory domains, even the scales used to measure stimuli such as sound intensity, earthquake intensity, and frequency along the electromatic 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) 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, as this provides more information discrimination ability and mimics natural human perceptual tendencies. The reasons for this heuristic are discussed more in section 1.2.4.

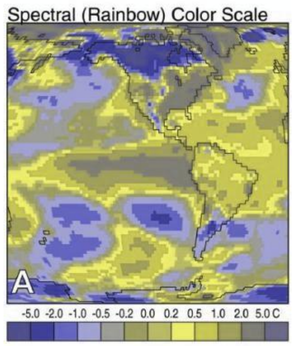
1.2.3.2 Colorblindness and color perception

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, but may be adaptive for seeing camoflauged 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.

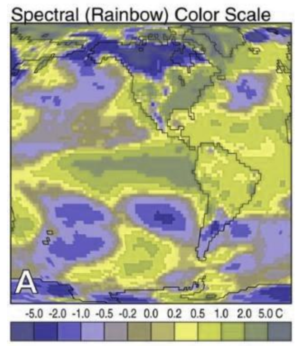
I can't tell a difference between the Protan* and Deuteran* pictures in Figure 1.8 at all. Is that just me?

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 camoflauged objects are easier to see), but these strategies fail when viewing abstract, constructed visual stimuli, such as graphics. Compounding this problem, the rainbow color schemes that are commonly used are particularly vulnerable to misinterpretation by colorblind viewers. Figure 1.8 shows a map using a rainbow color scheme (first shown in Light and

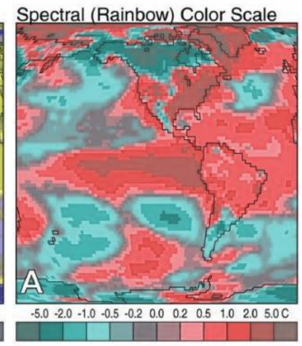




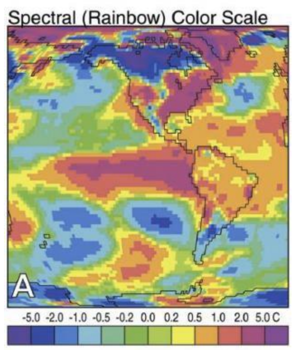
(a) Protanopia



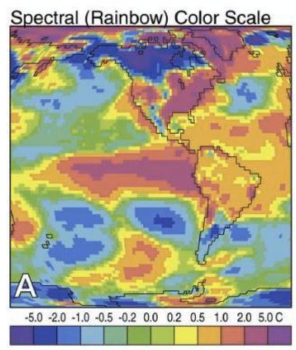
(b) Deuteranopia



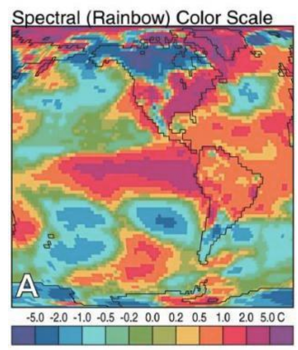
(c) Tritanopia



(d) Deuteranomaly



(e) Protanomaly



(f) Tritanomaly

Figure 1.8: 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.

Bartlein (2004)) and simulated images showing what that map would look like to those with missing cones (-anopia) and cones with altered wavelengths (-anomaly). , These simulations[2] show that rainbow color schemes are incredibly difficult for color-deficient individuals to read. 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

[2]provided by http://www.color-blindness.com/coblis-color-blindness-simulator/

a form appropriate for the human perceptual system.

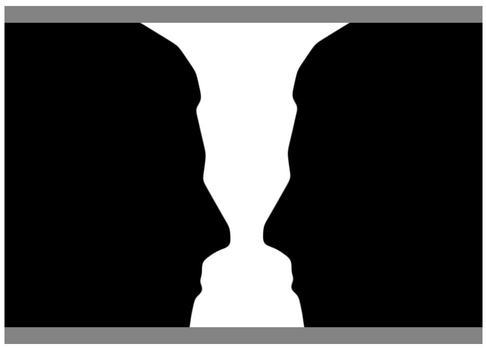
1.2.3.3 Optical Illusions

The "software" programs presented in section 1.2.2 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. As with most heuristic-based algorithms, though, these approaches produce suboptimal results when applied to more artificial tasks, such as reading statistical graphics. As such, 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 occurence based on the visual system.

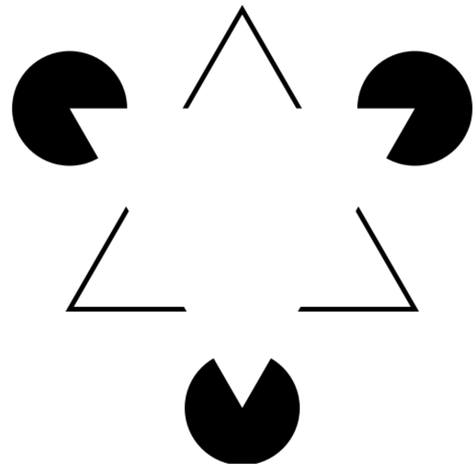
Physiological Illusions The illusions shown in Figure 1.3 are illusions which occur due to the wiring of the brain. These illusions can generally be avoided in statistical graphics, but are difficult to counteract once they occur.

Gestalt Illusions Some illusions occur due to a conflict of gestalt principles. Two of these illusions are shown in Figure 1.9: the figure/ground illusion, and the illusory contour illusion.

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 the vase and the faces. 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



(a) Figure-Ground Illusion



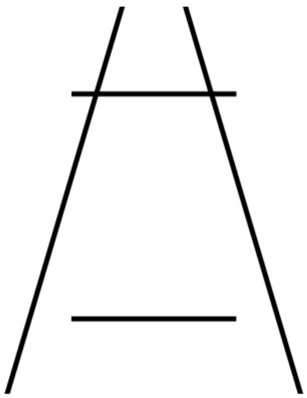
(b) Kanizsa Triangle Illusion

Figure 1.9: Illusions due to misapplied or ambiguous Gestalt rules.

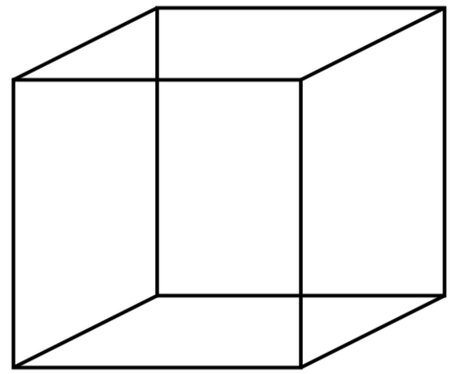
central point). In addition to the existence of the illusory contour, we also perceive some depth to the image; that is, the white triangle is perceived as being above the other components of the image (Coren and Porac, 1983), as this is the only way to make sense of the set of stimuli in a simple fashion. 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.

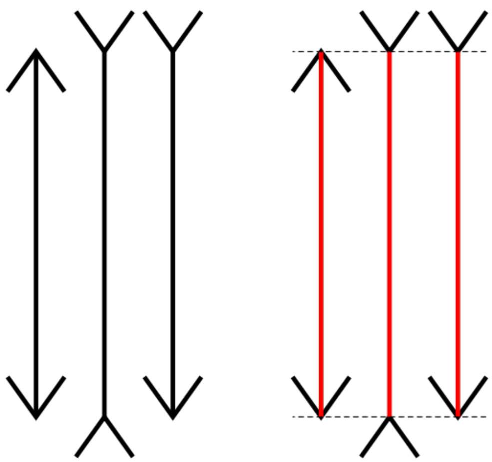
Figure 1.10 contains four of the more interesting optical illusions that result from ambiguous figures that trigger depth cues. The Ponzo illusion (Figure 1.10a) suggests that the top line is longer than the bottom line, because of the implied convergence of the two vertical lines (to understand the natural scenario behind this illusion, consider railroad tracks converging at the horizon). The Necker Cube, shown in figure 1.10b, 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 loses focus on the image momentarily(Gregory, 1997). The Muller-Lyer illusion (Figure 1.10c) is generally believed to result from misapplied depth cues as well - the left-most image would occur in nature



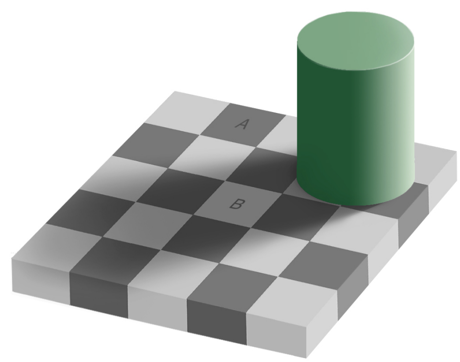
(a) Ponzo Illusion



(b) Necker Cube



(c) Muller-Lyer Illusion



(d) Color Constancy Illusion

Figure 1.10: Illusions due to misapplied depth perception.

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 1.11 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 1.10d 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 1.11: 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; Chapter 2 provides more information on a class of statistical graphics that trigger three-dimensional heuristics in the brain, producing misleading conclusions.

There are other optical illusions that have the potential to appear in statistical graphics but are not easily classified (or necessarily easily explained). These illusions are detailed in the next section.

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 1.12, is one such illusion.

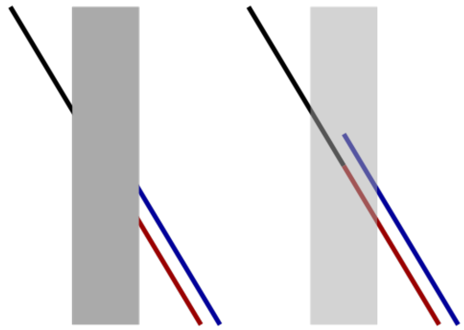
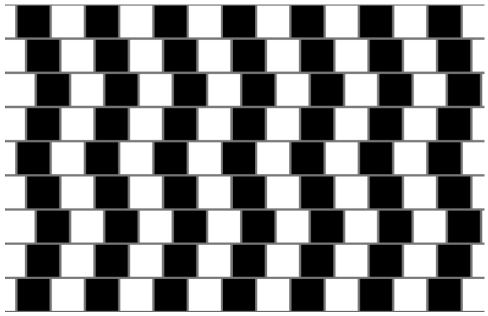


Figure 1.12: The Poggendorff Illusion. The left figure shows a black line which is obscured by a grey rectangle; it appears that the blue line would intersect the black line if the rectangle were not present. In fact, the black line and the red line are co-linear, and the blue line is parallel to both other lines.

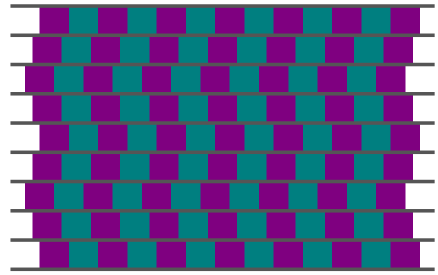
Gregory (1963, 1997) suggests the Poggendorff illusion is similar to the Muller-Lyer illusion in that it results from misapplied depth cues, but Green and Hoyle (1963); Ward et al. (1977) found no evidence that participants viewed this illusion in any three-dimensional context. Instead, Green and Hoyle (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 certain graphs (Amer, 2005; Poulton, 1985) if proper precautions are not taken.

The second of these illusions is the cafe wall illusion, shown in figure 1.13, named because this tile pattern is apparently common in cafes.

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 directly relevant to statistical graphics, shows that even simple (and pleasant)



(a) Original Cafe Wall Illusion



(b) Isoluminant Cafe Wall Illusion

Figure 1.13: 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.

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 it still has not been fully explained by psychologists or neuroscientists. It is on this cautionary note that we move into the research on graphical perception in statistics, as opposed to that in psychology.

The psychology research in this domain is typically very low-level, concerned with the underlying mechanisms of effects within the brain. In statistics, the literature is somewhat more variable; Cleveland and McGill (1985) produced the seminal paper on the subject, but outside of their 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. 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.

1.2.4 Cognitive Load

Short term memory and other considerations

Add short term memory and importance of freeing up brain cells to make it easier to retain information from graphics

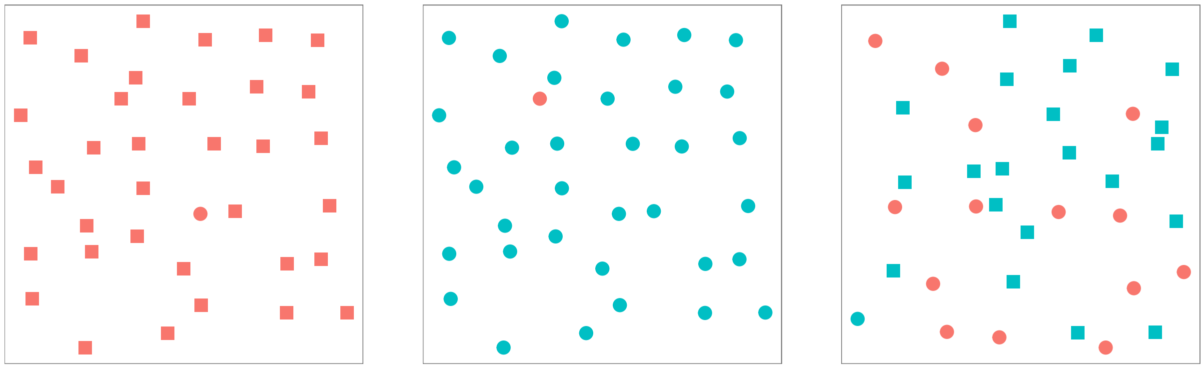
(Gattis and Holyoak, 1996) "Integrating information across dimensions normally imposes a heavy cognitive load, but graphs reduce this load by integrating values on dimensions by means of points or lines that simultaneously represent a value or a set of values on more than one dimension. Such visual integration, or "chunking"... allows people to reason about relations between two or more sets of data on those two or more dimensions."

1.3 Statistical Graphics

1.3.1 Low-level Perception of Statistical Graphics

Much of the lower-level research within the statistical graphics and information visualization community 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 1.2.2.1 and demonstrated in figure 1.14, 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), as shown in figure 1.14c.

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



(a) Shape (b) Color (c) Interference

Figure 1.14: 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.

be preattentive to be useful. 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.

1.3.2 Higher-level graphical perception

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 does the information get transferred from the page to working memory in the brain? Several psychological 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.

1.3.2.1 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

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)

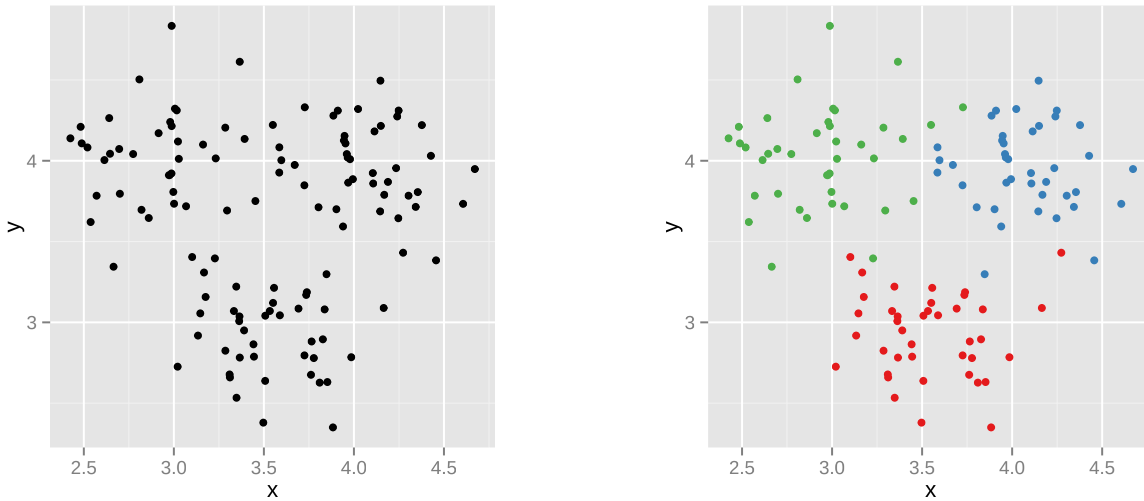
3. Once information is found, the focus shifts between the important part of the graph and the legend several times in order to keep the relevant information in working memory (conceptual relations produce quantitative meaning from visual features)

4. The question is answered and the participant moves on to another task (the question is related to the encoded quantitative features)

Create a simple graph and illustrate the steps of graph comprehension as described in Shah and Miyake (2005) pg 430.

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 this sequence to allow for more complex graphical relationships to be assimilated, generally as a participant cycles between stages (2) and (3) several times for different portions of the graph in order to understand relationships between different graphical elements. The time required for each of these steps may also change, depending on how familiar the reader is 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 in graphs. The gestalt law of proximity and similarity dictates 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 1.15 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. Of graphics that present information of similar complexity, graphics that require less effort to understand and search for relevant information are preferrable (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



(a) Proximity-based grouping

(b) Proximity and Similarity-based grouping

Figure 1.15: 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.

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 then be integrated into the viewer's working conceptual understanding of the dataset. Additionally, color schemes and appropriate labeling of graph features that reduce the amount of work necessary to integrate numerical information from a legend into the visual representation of the graph are likely to 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 psychologocial mechanisms involved in processing simple graphics are generally 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 or 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.

1.3.2.2 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 1.1. 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 1.1: 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; these contradictory results also suggest that the type of graph 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 (C 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). 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. While specific numerical judgments would still require selective attention, the "gist" of a graph using such schemes may be understood fairly quickly.

Other graph features can also influence the inferences viewers make: 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, 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 Chapter 2). 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).

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 require higher cognitive load and may promote recruitment of 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.

1.3.2.3 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 meterologists 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 meterologist'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 meterologist speaking aloud as they process each graph and map to assimilate information), the evidence is consistent enough 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 more difficult. One tool that facilitates such work in statistical graphics despite the complications described above 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.

1.4 Testing Statistical Graphics

1.4.1 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 (for an example, see Chapter 2). 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 are used, with each trial starting at either the lower or upper range of intensity and incrementally moving towards the opposite end of the range; the observer indicates the point at which they either begin to detect the stimulus or the point at which they can no longer detect the stimulus. At the end of several trials, the detection limits are averaged to produce a measured absolute threshold.

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.

Measuring the Difference Threshold The difference threshold, discussed in section 1.2.3.1, is the smallest detectable difference between two stimuli. This threshold can be measured using either the method of limits or the method of adjustment, but instead of increasing the absolute intensity of the stimulus, two stimuli are given, one with constant intensity and one whose intensity may vary either continuously or incrementally. The participant is instructed to identify the point at 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).

1.4.2 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. There are experimental methods such as visual search and eye tracking, but there are also experimental control procedures that may be important in graphics studies that are similar to cognitive psychology studies. We will first address the experimental methods, and then briefly discuss some common control procedures that may be applicable to the study of statistical graphics.

1.4.2.1 Experimental Methods

There are straightforward methods commonly used in psychological and statistical research; asking participants to make numerical judgments based on presented stimuli, for instance. 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 (Figure 1.14c), but 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 1.16.

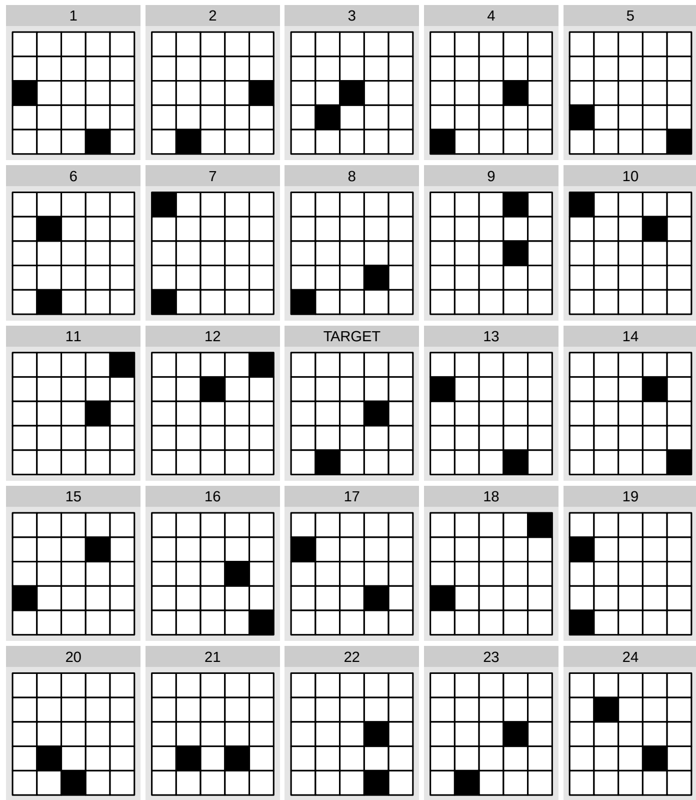


Figure 1.16: Visual Search Task. Participants were asked to locate the figure (1-24) most similar to the central "target" figure.

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 well below 100% performance on the task.

Eye Tracking Eye tracking studies are often utilized in order to understand which parts of an image participants focus on, and in what order they examine the image. 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 different parts of the graph in order to assimilate all of the represented information into working memory. Figure 1.17 shows one lightweight eye-tracking assembly. The camera allows researchers to track the direction of the pupil and thus infer gaze.

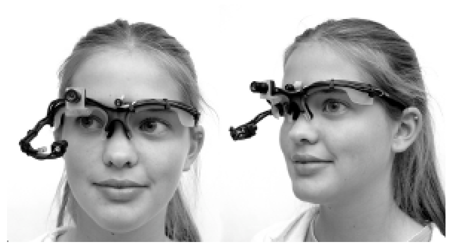


Figure 1.17: Eye tracking equipment (Babcock and Pelz, 2004). The cameras allow researchers to determine what part of a scene the wearer is viewing.

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.

In psychological experiments, the obvious experimental approach 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.

1.4.2.2 Experimental Control Procedures

While not all of these procedures 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. As a result, to avoidi habituation, experiments should generally consist of somewhat varied tasks in order to maintain participant attention.

Masking Images can persist on the retina for a period after the image is no longer available; this phenomenon is called <u>persistence of vision</u>. 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).

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.

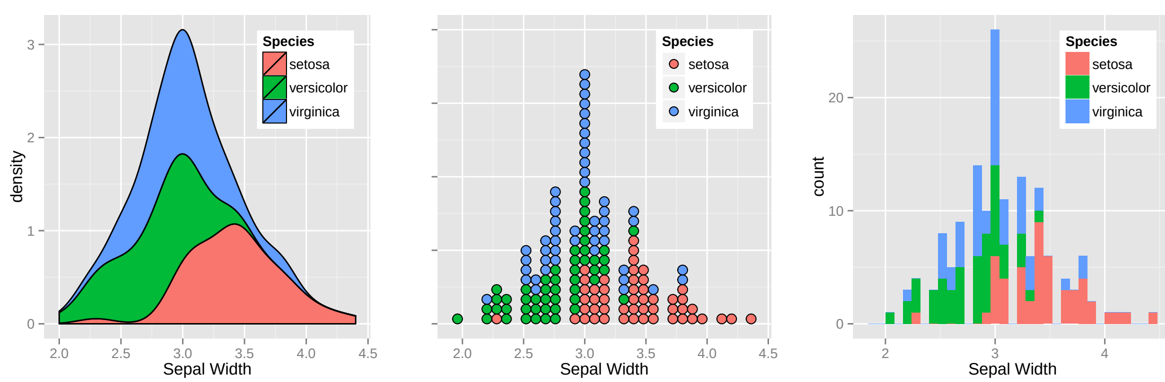
There are many other psychological mechanisms that may impact participant performance; the mechanisms presented here are simply some of the more salient considerations in experimental design for statistical graphics.

1.4.3 Testing Statistical Graphics

Tukey, Cleveland & McGill

1.4.3.1 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 1.18 shows three plots created using the same data and different geometric objects, and Figure 1.19 provides the ggplot2 code to create the plots [3]. 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.



(a) Stacked Density Plot (b) Dotplot (c) Histogram

Figure 1.18: Three different plots of iris data, created using the grammar of graphics

[3]These 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).

```r
# 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 1.19: ggplot2 code to produce Figure 1.18

1.4.3.2 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 1.20 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.

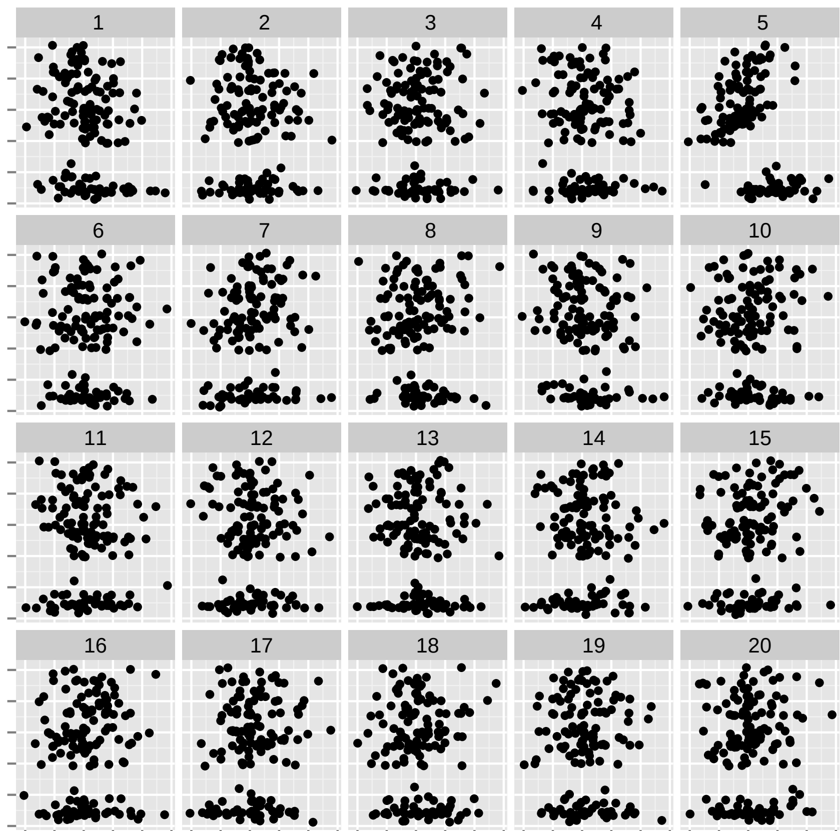


Figure 1.20: Lineup of the iris data, comparing sepal width to petal width. The target data is in plot 5, other plots generated by permuting petal width.

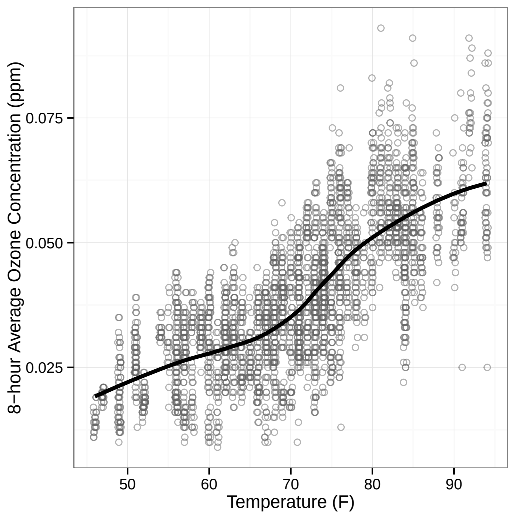
CHAPTER 2. SIGNS OF THE SINE ILLUSION – WHY WE NEED TO CARE

2.1 Introduction

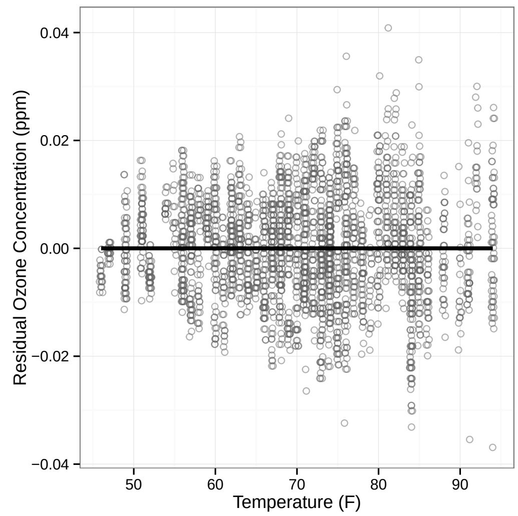
Graphics are powerful tools for summarizing large or complex data, but they rely on the main premise that any graphical representation of the data has to be "true" to the data (see e.g. Tufte (1991); Wainer (2000); Robbins (2005)). That is, a measurable quantity of a graphical element in the representation has to directly reflect some aspect of the underlying data. Generally, we see a lot of discussion on keeping true to the data in the framework of (ab)using three dimensional effects in graphics. Tufte (1991) goes as far as defining a *lie-factor* of a chart as the ratio of the size of an effect in the data compared to the size of an effect shown, with the premise that any large deviations from one indicate a misuse of graphical techniques. Computational tools help us ensure technical trueness – but this brings up the additional question of how we deal with situations that involve innate inability or trigger learned misperceptions in the audience. In this paper we want to raise awareness for one of these situations, show that it occurs frequently in our dealings with graphics and provide a set of strategies for solving or avoiding it.

As a first example let us consider the relationship between ozone concentration and temperature. Ozone concentrations were measured from 21 locations in the Houston area (Environmental Protection Agency, 2011), and temperature data is provided by the NCDC (National Climate Data Center, 2011) site at Hobby International Airport, located near the center of Houston.

Figure 2.1a shows daily measurements of 8-hour average ozone concentration and temperature at several sites in Houston, for days in 2011 with temperatures above 45°F and dew points



(a) Scatterplot of Ozone and Temperature in Houston, 2011. A loess fit is overlaid to show the overall trend.



(b) Scatterplots of Ozone and Temperature de-trended according to the loess fit in (a).

Figure 2.1: Scatterplots of Ozone and Temperature in Houston, 2011. The increase in variability over the temperature range is more pronounced in the de-trended plot on the right.

of less than 60°F. A loess smooth line is added for reference. These types of plots are often used to give an overview of the relationship between two variables. The trendline summarizes this relationship, while the points show raw measurement to allow an assessment of the overall size of the data, the amount of (marginal) variability presented, as well as the (conditional) variability along the trendline. It is the latter task that we cannot satisfactorily complete. While we might agree that there is an increase in variability of ozone concentrations for temperatures above 80°F, we will not doubt homogeneity elsewhere based on figure 2.1a.

This evaluation changes when considering figure 2.1b: the scatterplot shows a loess based de-trended residual of temperature. A previously almost invisible increase in variability of ozone measurements with increasing temperatures now becomes apparent.

This phenomenon, caused by the change in the slope of the trend line, is known as the *sine illusion* in the literature on cognition and human perception or *line width illusion* in the statistical graphics literature.

The illusion is a frequent occurrence in statistical graphics, and displays should therefore be thoughtfully considered to minimize its effect visually and acknowledge its influence. In the cognitive literature, Day and Stecher (1991) first documented the illusion in the context of vertical lines along a sinusoidal curve. Figure 2.2 shows a sketch of this: line segments are centered evenly spaced along the curve. Line segments are of equal length but appear longer in the peaks and troughs due to the illusion. The parameters that influence the strength of the illusion are the amplitude of the curve and the length of the line segments. As the length of the line segments increases, the apparent difference in the length of the line segments decreases. Any modification that increases the change in slope under which the curve appears, such as an increase in the amplitude of the curve or a more extreme aspect ratio, reinforces the apparent difference in line lengths.

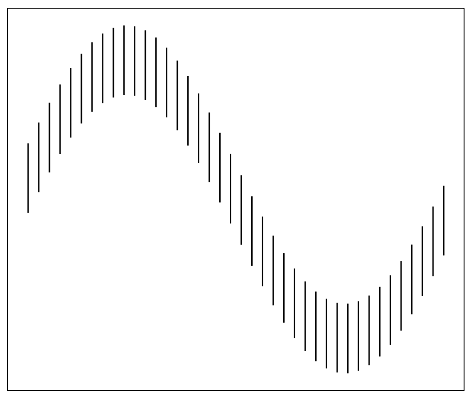
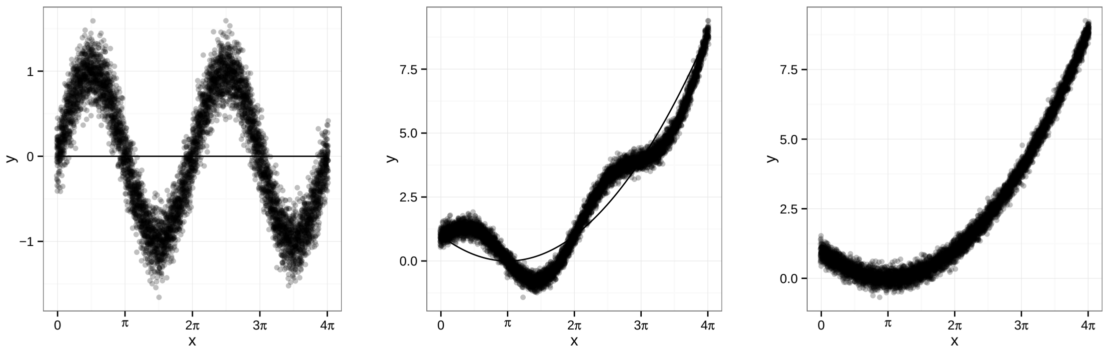


Figure 2.2: The original sine illusion, demonstrated on evenly spaced vertical lines centered around a sinusoidal curve of $f(x) = \sin(x)$. The lines in the peak and trough of the curve appear to be longer than in the other regions.

More recently the illusion has been shown in non-sinusoidal curves (Cleveland and McGill, 1984; Schonlau, 2003; Robbins, 2005; Hofmann and Vendettuoli, 2013), but the underlying effect seems to be the same, in the sense that the illusion is not triggered by the periodic nature of the underlying trendline but only by changes to its slope. Figure 2.3 shows three panels, which all exhibit the illusion. From left to right, the trend stems from (a) a periodic function, (b) a periodic component added to an exponential function, and (c) an exponential function on its own. While all three graphs seem to show nonconstant variance along the main trend; in reality, it is constant. Clearly, the illusion does not rely on the periodicity of the function for which it was named, but is a symptom of the change in curvature that comes with the periodicity.



(a) Seasonality, No Trend (b) Seasonality and Trend (c) Trend, No Seasonality

Figure 2.3: Set of three scatterplots of simulated data with constant variance. Plot (a) shows seasonality without any underlying trend, (b) shows seasonality superimposed on a quadratic trend, and (c) shows a quadratic trend without seasonality. Though all three sets of simulated data have constant variance, none of the variances appear constant due to the sine illusion.

Next, we give an overview of the perceptual and statistical literature regarding this illusion.

2.1.1 The Sine Illusion in Statistical Graphics

The sine illusion demonstrated in figures 2.1 and 2.2 has been frequently noted in statistical graphics, though usually not as an optical illusion. Rather, the problem is typically identified as the difficulty of visually subtracting two curves, and the resulting erroneous conclusions when this process goes awry. Figure 2.4 presents the possibly oldest example of this common phenomenon (Playfair, 1786; Playfair et al., 2005): Playfair's chart of the balance of trade between England and the East Indies shows time series of the trade value for imports and exports between the countries in the 18th century. The shaded area on the chart is named "balance against England", suggesting that the difference between the lines is of main importance. This difference in trading values is encoded as the difference between the lines along the vertical axis. However, the vertical distance between two lines provides a much less visually salient cue than the orthogonal width between the lines. This results in an underestimation (Cleveland and McGill, 1984) of the difference in trades around 1763, which is of a much higher (about 1.5 fold) magnitude as around 1770, but appears much smaller. In more modern visualizations, bivariate area charts and "stream graphs" (Byron and Wattenberg, 2008) commonly produce the illusion (see an example at http://bl.ocks.org/mbostock/3894205).

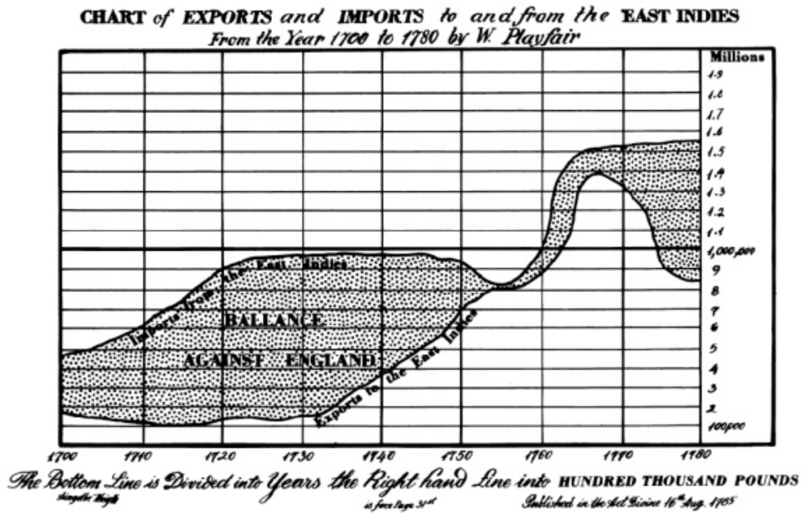


Figure 2.4: Playfair's graph of exports to and imports from the East Indies demonstrates that the line width illusion is not only found on sinusoidal curves but is present whenever the slope of the lines change dramatically. The increase in both imports and exports circa 1763 does not appear to portray as large of a deficit as that in 1710, even though they are of similar magnitude.

2.1.2 Perceptual Explanations for the Sine Illusion

While not thoroughly examined in the sensation and perception literature, the sine illusion has been classified as part of a group of geometrical optical misperceptions related to the Müller-Lyer illusion (Day and Stecher, 1991) or the Poggendorf illusion (Weintraub et al., 1980), which puts the illusion into the framework of context-based illusions. Day and Stecher (1991) suggest that the sine illusion occurs due to misapplication of perceptual experience with the three-dimensional world to a two-dimensional "artificial" display of data.

Experience with real-world objects suggests that the stimulus of figure 2.2 is very similar to a slightly angled top view of the 3-dimensional figure of a strip or ribbon describing waves in a third dimension, such as e.g. a road does on rolling hills. This is sketched out in figure 2.5a. Our experience suggests immediately that changes in the width of the road are unlikely and resolves the illusion. While figure 2.5a shows the line segments slightly angled towards each other, figure 2.5b shows a variation of the same plot with a vanishing point set further away from the viewer. This makes the line segments almost parallel to each other and therefore more closely resembles the sketch of figure 2.2, in which the sine illusion was originally presented.

(a) Perspective plot of sine illusion

(b) Perspective plot, vanishing point near infinity.

Figure 2.5: Two different perspective projections of the same data responsible for the sine illusion. The first projection angles the lines and appears more natural, but the second projection suggests that the lines do not need to be angled to create the same three-dimensional impression.

Recreating the three-dimensional context of the sine illusion might resolve the distortion, even if increasing the dimensionality of a graph is generally not recommended (Tufte, 1991; Cleveland and McGill, 1984) (though Spence (1990) suggests that in certain cases additional dimensions are not misleading). While creating a three-dimensional projection of two-dimensional data might counteract the illusion, the process of projecting the data accurately into a higher dimension is not simple. The projection that best resolves the illusion likely is highly subjective and influenced by choices of angle and color gradient for depth cues. As there is not a single three-dimensional projection that corresponds to the two-dimensional data, this approach would only produce further visual ambiguity.

Further complicating the situation, the illusion itself is insidious – we trust our vision implicitly, to the point that when we understand something, we say “I see”. This trust in our visual perception is seldom called into question, for our perception is optimized for interaction with a three-dimensional world. Artificial two-dimensional situations (such as graphs and pictures) may accurately represent the data and still produce a misleading perceptual experience.

The contextual cues of the overall trend are critical to the sine illusion’s effect; the illusion only holds when a substantial portion of the graph is considered simultaneously, which triggers our innate ability of perceiving one whole rather than the individual parts it consists of (principle of grouping; Wolfe et al. (2012)). Considering only two line segments at a time

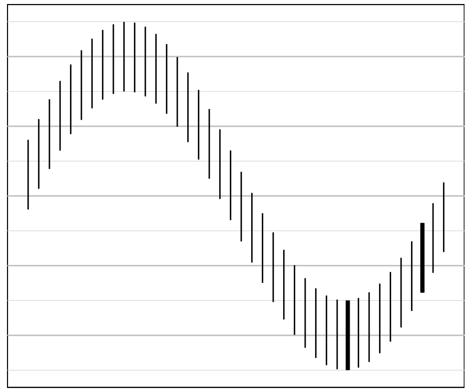


Figure 2.6: The sine illusion with two individual lines highlighted. Horizontal grid lines do not help to resolve the illusion, even though they provide a clear basis for comparison of line lengths. Readers are much better at assessing the length of the two singled out line segments; they are equal.

resolves the illusion. The bold lines in figure 2.6 are clearly of the same length. Comparisons of individual line lengths is visually a fairly simple task, and is done with a relatively high accuracy (Cleveland and McGill, 1984). Day and Stecher (1991) contains a more thorough discussion of how much surrounding context is required for the illusion to persist.

2.1.3 Geometry of the Illusion

In figure 2.2 we have seen that the our preference in evaluating line width is to assess *orthogonal* width rather than the difference along the vertical axis. Figure 2.7 demonstrates the change in orthogonal width as the slope of the line tangent to the graph of f changes; these changes correspond to our perception of apparent line length.

The illusion is most pronounced in regions where the angle between the orthogonal and the vertical line is large. Changes to the aspect ratio therefore have a major impact on the strength of the sine illusion. Any change that alleviates the difference between perceived width and the perpendicular width, such as banking to 45° (Cleveland et al., 1988), will alleviate the effect but not completely overcome it. The perceived length of the vertical line changes with the angle of the line perpendicular to the slope of $\sin(x)$, suggesting that the sine illusion stems from a conflict between the visual system's perception of figure width and the mathematical judgement necessary to determine the length of the vertical lines.

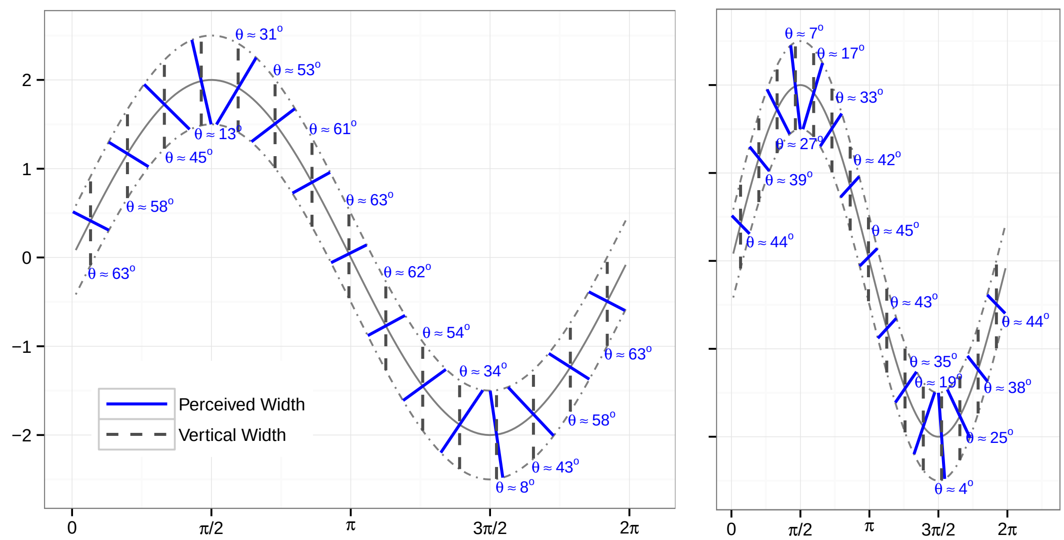


Figure 2.7: The sine illusion with lines orthogonal to the tangent line at $f(x)$. The perception that the vertical length changes with $f(x)$ corresponds to changes in actual orthogonal width due to the change in the visual (plotted) secant angle. The strength of the perceptual effect depends in part on the aspect ratio of the graph, as shown in the second image, which has an aspect ratio of 2 compared to the first figure's aspect ratio of 1. This correspondingly multiplies the strength of the effect by 2.

Our preference for assessing figure width based on the orthogonal width suggests that the underlying illusion may be a function of geometry rather than some unknown visual or neural process that occurs subconsciously. In this case it may be possible to correct the graphical display for the illusion to minimize its misleading effect. A geometrical correction that –at least temporarily– counteracts the illusion would be a valuable tool in visual analysis, as this illusion very persistently affects our judgment of very common tasks such as e.g. the assessment of conditional variability of data along a trend line.

Simply raising people's awareness of the presence of this illusion is not enough, as it is incredibly difficult, if not impossible, to overcome this illusion even when we are aware of its presence: our brains simply cannot "un-see" it.

What follows is a compilation of several approaches to correct for or mitigate the effect of the illusion. Our primary intention here is to demonstrate the persaviness of the illusion is and the extreme measures necessary to remove its effect.

2.2 Breaking the Illusion

The sine illusion is caused by a conflict between vertical width, which is the width that we want onlookers to assess visually, and orthogonal width, which is the width that the onlooker perceives. This difference can be expressed as a function in the slope of the underlying trend line. This provides the basis for adjusting the vertical width for the perceived orthogonal width.

We consider the following three approaches:

1. separating the trend and the variability,

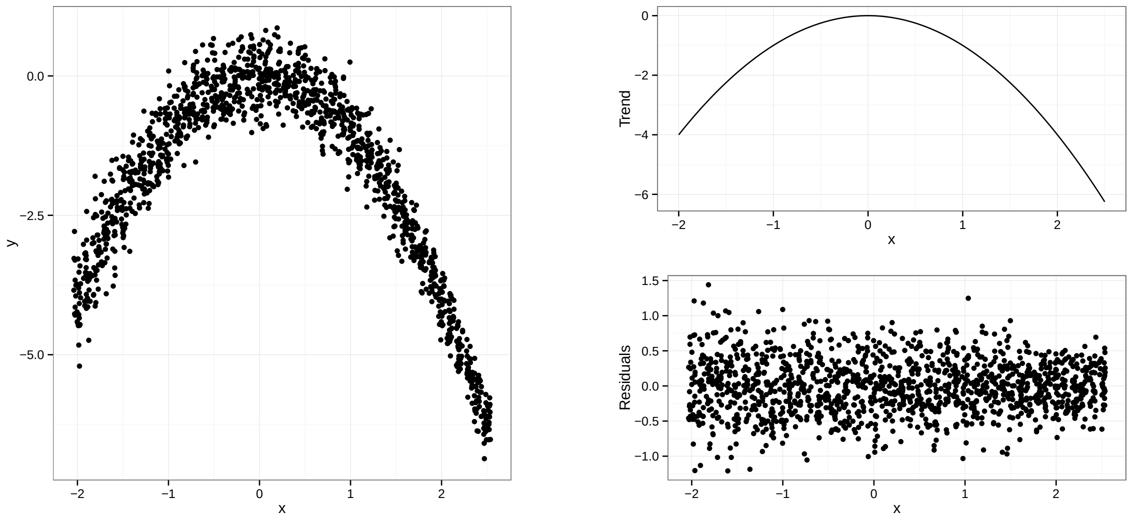
2. transformation of x: adjusting the slope to be constant by reparametrizing the x axis, and

3. transformation of y: adjusting y values to make conditional variability appear correctly by adjusting accordingto orthogonal width.

Each of these ideas is discussed in more detail in this section.

2.2.1 Trend Removal

Cleveland and McGill (1984, 1985) discuss the perceptual difficulty of judging the difference between two curves plotted in the same chart, and alternatively, recommend to display the difference between the two curves directly. This is in line with recommendations for good graphics to 'show the data' rather than make the reader derive some aspect of it (e.g. Wainer (2000)). In particular, de-trending data to focus on residual structure is the generally accepted procedure for assessing model fit. Figure 2.8(a) shows a scatterplot of data with a trend. A loess smooth is used to estimate the trendline. A visual assessment of variability along this trendline might result in a description such as 'homogeneous variance or slightly increasing variance for negative x, followed by a dramatic decrease in vertical variability for positive x'. Once the residuals are separated from the trendline as shown on the right hand side of the figure, it becomes apparent that this first assessment of conditional variability was not correct, and the decreasing variance along the horizontal axis becomes visible.



(a) Data

(b) Trend and Residuals

Figure 2.8: Describe the conditional variability of the points along the x axis in (a). Is your description consistent with the residual plot in (b)?

While the illusion is not apparent when trend line and variability in the residual structure are shown separately, the separation makes it more difficult to evaluate the overall pattern in the data, as we must base any judgment on two charts; either by combining information from two graphs or by mentally re-composing the original graph (at which point, the sine illusion becomes a factor). To minimize cognitive demands we ideally want to tell the whole story with a single graph, in particular because in many situations we may not be able to show multiple graphs.

Additionally, removing the trend requires an initial model, making any plots produced using that fit conditional on the assumptions necessary to obtain that model fit. In many situations, this may be undesireable. In particular, we typically view the data before fitting even a rudimentary model, and the sine illusion may influence even these initial modeling decisions.

2.2.2 Transformation of the X-Axis

As the sine illusion is driven by changes in the slope of trends between variables, we can counteract the illusion by removing these changes, transforming the x axis such that the absolute value of the slope is constant and forcing the corresponding orthogonal width to represent the conditional variability. In order to describe this transformation of the x axis mathematically, let us assume that the relationship between variables X and Y is given by a model of the form

$$y = f(x) + \varepsilon,$$

where f is some underlying function (either previously known or based on a model fit). Further let us assume that f is differentiable over the region of observed data.

Ideally, the correction would force all lines to appear under the same slope, i.e. we want to find a transformation $T(x)$ of x, such that $f(T(x))$ is a piece-wise linear function, where each piece has the same absolute slope. This transformation has an effect similar to "banking to 45°" in a piecewise manner.

Let a and b be the minimum and maximum of the x-range under consideration. Then for any value $x \in (a, b)$ the following transformation results in a function with constant absolute slope:

$$(f \circ T)(x) = a + (b-a)\left(\int_a^x |f'(z)|dz\right) / \left(\int_a^b |f'(z)|dz\right), \tag{2.1}$$

2.2.2.1 Derivation of the X Transformation

As the slope is determined by the aspect ratio, we are free to choose it and w.l.o.g. we get for each piece T_i:

$$f(T_i(x)) = \pm ax + b_i.$$

This means that T_i is essentially an inverse of function f, with each piece defined by the intervals on which the inverse of f exists: let $\{x_0 = \min(x), x_1, ..., x_{K-1}, x_K = \max(x)\}$ be the set of values with local extrema enhanced by the boundaries of the x-range, i.e. $f'(x_i) = 0$ for $i = 1, ..., K-1$ and $f'(x) \neq 0$ for any other values of x. Then each interval of the form (x_{i-1}, x_i) defines one piece T_i of the transformation function $T(x)$. We will define T_i now as a combination of a linear scaling function and the inverse of f, which we know exists for interval (x_{i-1}, x_i).

Let function $s = {}_{[a,b]}s^{[c,d]}$ be the linear scaling function that maps the interval (a, b) linearly to the interval (c, d). This function is formally defined as

$$s(x) = {}_{[a,b]}s^{[c,d]}(x) = (x - a)/(b - a) \cdot (d - c) + c \text{ for all } x \in (a, b).$$

Note that the slope of function s is given as

$$s'(x) = (d - c)/(b - a).$$

Two scaling functions can be evaluated one after the other, only if the image (i.e. y-range) of the first coincides with the domain (i.e. x-range) of the second. This consecutive execution results in another linear scaling:

$${}_{[e,f]}s^{[c,d]}\left({}_{[a,b]}s^{[e,f]}(x)\right) = {}_{[a,b]}s^{[c,d]}(x)$$

In our situation let the scaling function s be given as:

$${}_{[c,d]}s^{f([x_{i-1},x_i])}(x) = f(x_{i-1}) + (x - c)/(d - c) \cdot (f(x_i) - f(x_{i-1}))),$$

where $f([x_{i-1}, x_i])$ is defined as the interval given by $(\min(f(x_{i-1}), f(x_i)), \max(f(x_{i-1}), f(x_i)))$. Note that s has either a positive or negative slope depending on whether $f(x_{i-1})$ is smaller or larger than $f(x_i)$, respectively.

Then the transformation in the x-axis, $T(x)$ is defined piecewise as a combination of T_i, where each T_i is given as:

$$T_i(x) \quad = \quad f^{-1}\left({}_{[c_i,d_i]}s^{f([x_{i-1},x_i])}(x)\right). \tag{2.2}$$

Using this definition for the transformation makes $f(T(x))$ a piece-wise linear function with parameters c_i and d_i, i.e. for $x \in (c_i, d_i)$ we have

$$f(T(x)) = f(f^{-1}({}_{[c_i,d_i]}s^{f([x_{i-1},x_i])}(x))) = {}_{[c_i,d_i]}s^{f([x_{i-1},x_i])}(x).$$

Correspondingly, the slope of $f(T_i(x))$ is $(f(x_i) - f(x_{i-1})))/(d_i - c_i)$. In order to make the slope the same on all pieces T_i of T, we need to define c_i and d_i with respect to the function values on the interval (x_{i-1}, x_i). There are various options, depending on how closely the x-range of T should reflect the original range: for $[c_i, d_i] = \text{range}\,(f([x_{i-1}, x_i]))$ the new x-range is the range of f on (x_{i-1}, x_i), but with the advantage that the scaling function simplifies to the identity or a simple shift.

In order to preserve the original x-range, we need to invest into a bit more work for the scaling. With an identity scaling, each T_i maps from the range of f on (x_{i-1}, x_i) to the same range. Overall we can therefore set up the function T to map from the interval given by the sum of the function's 'ups' and 'downs', i.e. $(0, \sum_{i=0}^{K} |f(x_i) - f(x_{i-1})|)$, to the range of f on (x_0, x_K). This ensures that all pieces $f(T_i)$ have the same slope (of $|1|$).We can then use another - global - linear scaling function to map from the range of x, i.e. interval (x_0, x_K) to $(0, \sum_{i=0}^{K} |f(x_i) - f(x_{i-1})|)$, yielding a transformation function T of

$$T(x) = (f^{-1} \circ {}_{[c_i,d_i]}s^{f([x_{i-1},x_i])} \circ {}_{(x_0,x_K)}s^{(0,\sum_{i=0}^{K}|f(x_i)-f(x_{i-1})|)})(x),$$

where c_i and d_i are given as

$$c_i = \sum_{j=0}^{i-1} |f(x_j) - f(x_{j-1})| \text{ and } d_i = \sum_{j=0}^{i} |f(x_j) - f(x_{j-1})|.$$

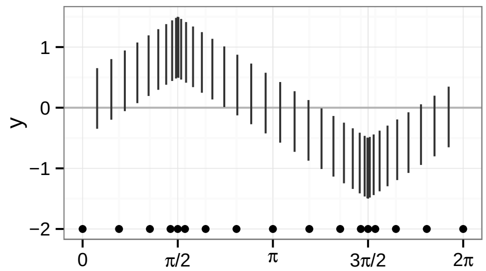
We can write the difference $|f(x_j) - f(x_{j-1})|$ as $\int_{x_{j-1}}^{x_j} |f'(z)| dz$. This shows equation (2.1).

2.2.2.2 Weighting the X Transformation

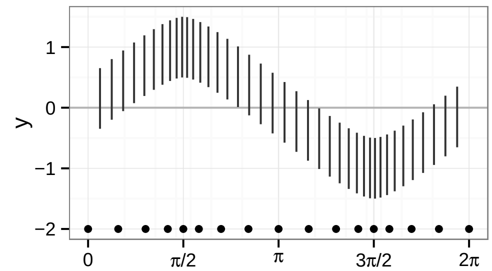
As the sine illusion depends on changing slope in the overall trend, re-parametrizing the x-axis in terms of the slope will make the data appear under a constant slope, thereby removing the effect of the illusion, while the transformed x-axis is changed from a linear representation of the x values to a 'warped' axis that continuously changes the scale of x to compensate for the changes in the slope. To emphasize this change in scale along the x axis, dots are drawn at the bottom of the chart to show the transformation's effect on equally spaced points along the x-axis.

Results from this transformation are demonstrated in Figure 2.9a.

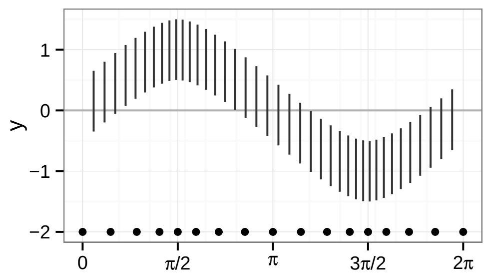
While the transformation in equation (2.1) effectively removes the appearance of changing line lengths, we can see in practice that the illusion can be broken by a much less severe transformation of the x axis. For that we introduce a shrinkage factor $w \in (0, 1)$ that allows a



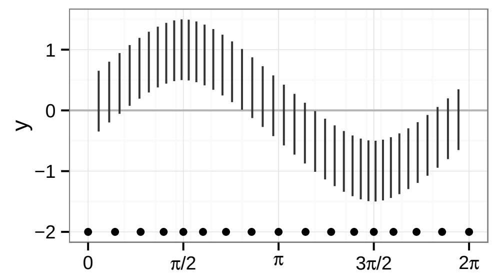
(a) X axis transformation based on eqn. (2.1), corresponding to weighting of $w = 0$.



(b) Weighted Transformation, $w = 1/2$ (based on eqn. (2.3))



(c) Weighted Transformation, $w = 1/3$



(d) Weighted Transformation, $w = 1/4$

Figure 2.9: Examples of X axis transformations in the sine curve. Dots at the bottom of the graph show the transformation's effect on equally spaced points along the x-axis. Different amounts of weighting w correspond to differently strong corrections. In (a), x-spacing of the lines changes the extant width such that the absolute value of the slope is uniform across the whole range of the x axis resulting in the largest amount of correction. (b) - (d) reduce the correction in (a) towards successively more uniform spacings in x while still breaking the effects of the illusion.

weighted approach in counteracting the illusion as:

$$(f \circ T_w)(x) = (1 - w) \cdot x + w \cdot (f \circ T)(x) \tag{2.3}$$

Note that for $w = 1$ the x-transformation is completely warped, while smaller values of w indicate a less severe adjustment against the sine illusion. Under weaker transformation,s the data more closely reflect the original function $f(x)$. Figures 2.9b - 2.9d show the effect of different shrinkage coefficients w. As w decreases, the lines become more evenly spaced and the illusion begins to return.

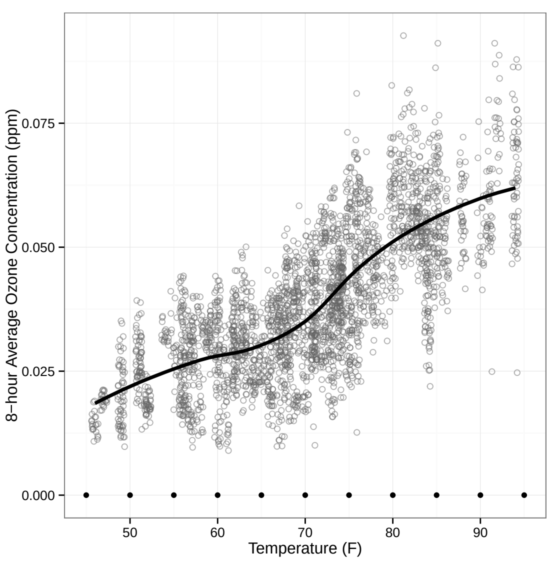
The extent to which we can shrink the adjustment back to the original function varies with the aspect ratio of the chart and the shape of the function. It might also be influenced by the audience's experience with the sine illusion, resulting in very subjective choices of an "optimal weighting" for specific situations which minimizes distortion and maximizes the correspondence between inferences made from the data and inferences made using the visual display.

Note, that we only make use of the transformation T in the form of $f \circ T$. This allows us to avoid an explicit calculation of the transformation T, which in particular involves a computation of the inverse of f leading to potentially computation-intense solutions.

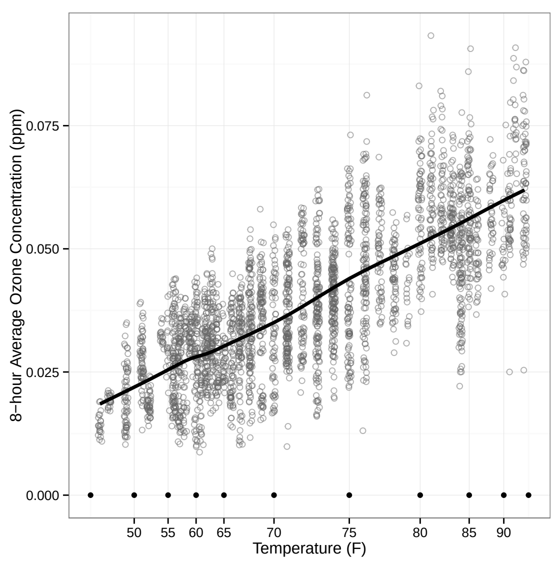
2.2.2.3 X Transformation Demonstration

In the example of the Ozone data shown in figure 2.1, we can base a transformation of the x-axis on a loess fit of ozone concentration in daily temperature. Loess is particularly convenient for this transformation, as it enforces continuity conditions including differentiability of the fitted function; software allows us to obtain fits of both the function values and their derivatives.

Figure 2.10 shows the original data side-by-side with the transformed x-axis, demonstrating not only the effect of transformation of the x-axis, but also that the transformation is not overly misleading in this example. The granularity of the data in this example provides an implicit measure of the strength of the transformation along the x-axis and the transformation is also clearly evident in the labels along the x-axis.



(a) Original Data



(b) Transformed X Axis

Figure 2.10: Original data and data after x-transformation. The increasing variance is easier to see when x has been transformed, because the slope is now uniform.

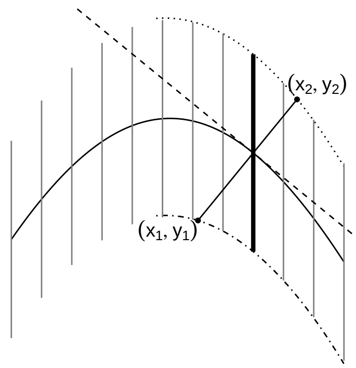


Figure 2.11: General correction approach. This approach may require numerical optimization to obtain exact solutions for (x_1, y_1) and (x_2, y_2).

2.2.3 Transformation in Y

Understanding the geometry of the sine illusion leads to another approach to counteracting the conflict between the orthogonal width and the vertical length of the segment.

Let again the function f describe the general relationship between variables X and Y.

As sketched out in figure 2.11 we want to first find the orthogonal (extant) width in a point $(x_0, f(x_o))$ on the graph, which corresponds to the perceived width, and then correct the vertical width accordingly to match with the audience's expectation.

The orthogonal width (see sketch in figure 2.11) is given as the line segment between endpoints $(x_1, f_1(x_1))$ and $(x_2, f_2(x_2))$, where f_1 and f_2 denote the vertical shifts of function f by $-\ell/2$ and $\ell/2$, respectively, where ℓ is defined as the overall line length, $\ell > 0, \ell \in \mathbb{R}$. These endpoints are determined as the intersection of the line orthogonal to the tangent line in $(x, f(x))$ and the graphs resulting from the vertical shifts of f.

The function describing the orthogonal line through $(x_o, f(x_o))$ is given in point-vector form as

$$\begin{pmatrix} x_o \\ f(x_o) \end{pmatrix} + \lambda \begin{pmatrix} f'(x_o) \\ 1 \end{pmatrix},$$

for any real-valued λ. The advantage of using point vector form is that it allows us to solve for parameter λ easily, which gives us easy access to the extant (half-)widths, as:

$$|\lambda| \sqrt{1 + f'(x_o)^2}. \tag{2.4}$$

Eqn. (2.4) describes the quantity that we perceive rather than the quantity that we want to display ($\ell/2$), which leads us to a general expression of the correction factor as

$$\ell/2 \cdot \left(|\lambda|\sqrt{1+f'(x_o)^2}\right)^{-1}.$$

Note that this yields in general two solutions: one for positive, one for negative values of λ corresponding to upper and lower (half-)extant width.

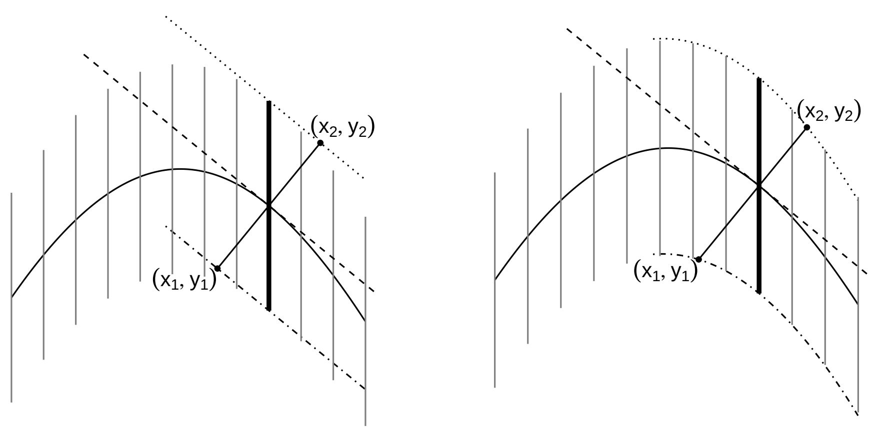
In order to get actual numeric values for λ, we need to find end points of the extant line width as solutions of intersecting the orthogonal line and the graphs of f_1 and f_2. We find these end points as solutions in x and λ of the system of equations:

$$x - x_o = \lambda f'(x_o) \tag{2.5}$$

$$f(x) - f(x_o) = -\lambda \pm \ell/2 \tag{2.6}$$

Note that the above system of equations involves function values $f(x)$, which implies that solving this system requires numerical optimization for any but the most simple functions f.

In the following two sections we make use of Taylor approximations of first and second order to find approximate solutions to end points as sketched out in figure 2.12.



(a) Linear Approximation

(b) Quadratic Approximation

Figure 2.12: (a) uses a first-order Taylor series approximation to $f(x)$ and (b) uses a second-order Taylor series approximation to $f(x)$. The intersection of the function $f(x) \pm \ell/2$ and the orthogonal line, $(x_1, y_1), (x_2, y_2)$ must be obtained to determine the necessary correction factor.

2.2.3.1 Linear Approximation to $f(x)$

For the linear approximation we make use of $f(x) \approx f(x_0) + (x - x_0)f'(x_0)$, which together with equations 2.5 and 2.6 yields a correction factor in x_0 of

$$\ell_{\text{new}}(x_0) = \ell_{\text{old}}\sqrt{1 + f'(x_0)^2}.$$

Note that the linear method gives the same result as a varying slope extension from a trigonometric approach suggested by Schonlau (2003) and used in Hofmann and Vendettuoli (2013)

A second-order Taylor polynomial approximation to $f(x)$ additionally accounts for the asymmetry in the extant widths on either side of the center trendline.

2.2.3.2 Quadratic Approximation to $f(x)$

Using the approximation $f(x) \approx f(x_0) + f'(x_0)(x - x_0) + 1/2 f''(x_0)(x - x_0)^2$, the system of equations 2.5 and 2.6 simplifies to the following quadratic equation in λ:

$$f''(x_0)f'(x_0)^2\lambda^2 + 2(f'(x_0)^2 + 1)\lambda \pm \ell = 0,$$

which leads us to corrections for the half lengths as:

$$\ell_{\text{new}_1}(x_0) = 1/2 \cdot \left(v + \sqrt{v^2 + f''(x_0)f'(x_0)^2 \cdot \ell_{\text{old}}}\right) \cdot v^{-1/2} \tag{2.7}$$

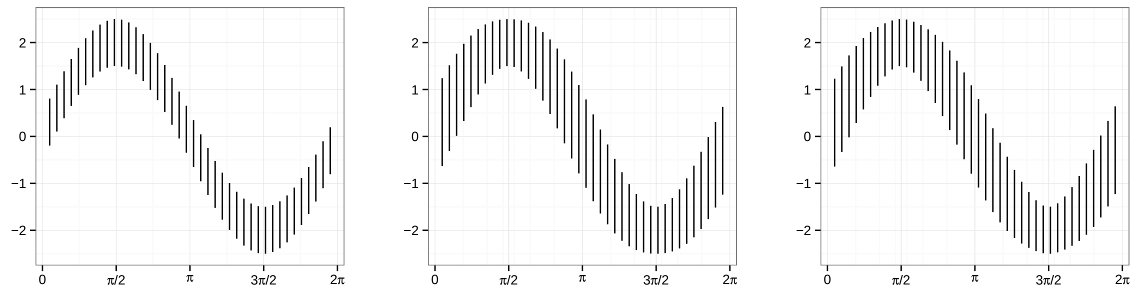
$$\ell_{\text{new}_2}(x_0) = 1/2 \cdot \left(v + \sqrt{v^2 - f''(x_0)f'(x_0)^2 \cdot \ell_{\text{old}}}\right) \cdot v^{-1/2} \tag{2.8}$$

where $v = 1 + f'(x_0)^2$.

2.2.3.3 Reformulation of the quadratic approximation

A quadratic equation in λ of the form

$$a\lambda^2 + b\lambda + c = 0, \tag{2.9}$$



(a) Uncorrected (b) Linear correction (c) Quadratic correction

Figure 2.13: In the quadratic approximation top and bottom segments of the vertical lines are adjusted separately.

where a, b, and c are real-valued parameters the solutions take on the form

$$\lambda_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \stackrel{*}{=} 2c\left(-b \pm \sqrt{b^2 - 4ac}\right)^{-1}.$$

* if $b \neq \pm \sqrt{b^2 - 4ac}$, i. e. $a, c \neq 0$.

Application to quadratic approximation to f: in the example, we have the following equivalencies:

$$\begin{aligned} a &= f''(x_0) f'(x_0)^2 \\ b &= 2(1 + f'(x_0)^2) \qquad > 0 \text{ for all } x \\ c &= \pm \ell \end{aligned}$$

For a valid solution for the correction factor, we have to assume that λ is a factor that extends the original extant width (in absolute value).

$$\lambda_{1/2} = \ell\left(v + \sqrt{v^2 \pm f''(x_0) f'(x_0)^2 \cdot \ell}\right)^{-1}$$

for $v = 1 + f'(x_0)$. This gives the results as shown in equations (2.7) and (2.8)

Adjusting the top and bottom segments of the vertical lines separately so that the extant width is constant breaks the illusion, but slightly distorts the sinusoidal shape of the peaks.

Figure 2.13 shows the correction factor based on a quadratic approximation compared to the untransformed data. Unlike the linear solution, the half-segments here are not necessarily of the same length, and thus there are separate correction factors for each half-segment.

2.2.3.4 Mathematical Properties of the Y Transformation

The quadratic correction breaks whenever the expression in the square root of eqn. (2.7) becomes negative, i.e. whenever $v^2 \pm \ell \cdot f''(x) \cdot f'(x)^2 < 0$. This happens for combinations of large values of ℓ, which signify a large vertical extent, or large conditional variability $E[Y|X]$, and simultaneous large changes in the slope of the main trend, i.e. large values of the curvature $f''(x)$. In the linear approximation of f the same situation leads to a massive overcorrection of the vertical lines, changing the shape of the 'corrected' function beyond recognition.

Similar to the correction of the x-axis, we can use a weighted approach to find a balance between counteracting the illusion and representating the original data:

$$\ell_{new_w}(x) = (1 - w) \cdot \ell_{old} + w \cdot \ell_{new}(x) \tag{2.10}$$

2.3 Transformations in Practice – a User Study

In order to more fully understand the sine illusion and test the proposed corrections, we created an applet to allow users to investigate the illusion's prominence with respect to its parameters. Users can examine the sine illusion by changing line length, the function's amplitude, and compare corrections in x-axis and y-values to uncorrected data. All corrections proposed in this paper are implemented in the applet located at `http://glimmer.rstudio.com/srvanderplas/SineIllusion/`.

We employed a second applet to collect data on users' preferences on the amount of correction used, i.e. we are interested in identifying a range of 'optimal weights' in each of the corrections. This applet presents users with a graph that is the result of a correction in x or y with a randomly selected starting weight value. Users are asked to adjust the graph until the illusion (a) is no longer apparent (adjustment of weights from the bottom) or (b) becomes visible (adjustments of weights from the top).

Both applets are implemented in `shiny` (RStudio Inc., 2013).

The graphs in the data-collection applet are adjusted using a plus/minus button to either increase or decrease the amount of correction used. Underlying this adjustment is the value of

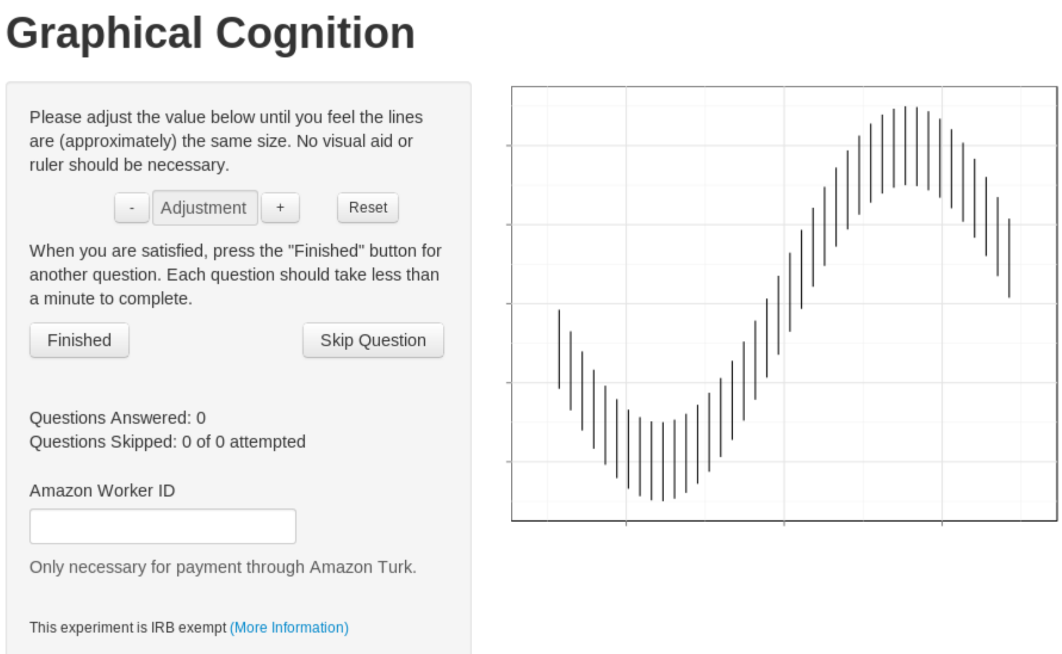


Figure 2.14: Screenshot of the shiny application used to collect information of observers' preference with respect to an optimal correction for the illusion under each of the transformations discussed in the previous section.

the weight w as defined in eqns. (2.3) and (2.10). The numerical value of w was hidden from the user to prevent anchoring to a specific numerical value.

A low initial weight (w_0 close to 0) indicates that the amount of correction is low and the response from a trial like this will give us an idea of the minimal amount of weight necessary to break the illusion, while a high initial weight (w_0 close to 1) indicates that the data is fully corrected. We asked participants to change the amount of adjustment until the lines appear to be the same length assumes that the correction is overcorrecting in practice, and a response from this type of trial gives us an upper boundary for the amount of weighting preferred. Generally, responses from the two different types of trials do not result in the same threshold weight, but rather lead to a range of acceptable weights.

It is of additional interest to determine whether and how much these optimal weights are subject-specific or population-based, whether they depend on the initial weight, and how much within-subject variability we find compared to between-subject variability.

Figure 2.14 shows a screenshot of the applet used to collect user data. This applet is available online at http://glimmer.rstudio.com/srvanderplas/SineIllusionShiny/. Line length and function are controlled in this app, and we used the linear transformation for adjusting y values; the transformation does not break under any combination of parameters tested in this experiment.

We deployed the applet to participants recruited online, collecting their responses and other metadata. The results of the analysis suggest that the correction factors in X and Y are both preferrable to uncorrected data, but that a full correction is not necessary to break the illusion.

2.3.1 Study Design

The study aims to determine the range of "optimal" transformation weights for each transformation type. Psychophysics methodology typically approaches threshold estimation by using the method of adjustment (Goldstein, 2009), where stimuli are provided showing states both above and below the hypothesized optimal value and participants adjust the stimuli until the stated goal is met (in this case, until the lines appear to have equal length). It is expected that there will be a difference in user-reported values from below and from above, and these values are typically averaged to produce a single threshold value. Beyond averaging these values, we use a mixed model to compare user responses for different starting points in a more continuous fashion, incorporating some of the advantages of the method of constant stimuli to more robustly estimate the range of optimal transformation weights. For a review of general psychophysics methodology, the method of adjustment, and the method of constant stimuli, see Goldstein (2009).

The study is set up as a fractional factorial design of correction type (x or y correction) and starting weight w_0. Each participant is asked to evaluate a total of twelve situations, six of each correction type. Starting weights were chosen as follows: each user was given a trial of each type starting at 0 and 1. The remaining four trials of each type had starting weights chosen with equal probability from 0.25 to 0.75 (see figure 2.15). We decided to have a higher coverage density for starting weights around 0.6 after a pilot study indicated a preference for that value. Using a distribution with a wide coverage allows us to more fully explore the space of plausible weights w while focusing on the $(0, 1)$ interval and enabling precise estimation of the optimal weight in the region indicated by the pilot study.

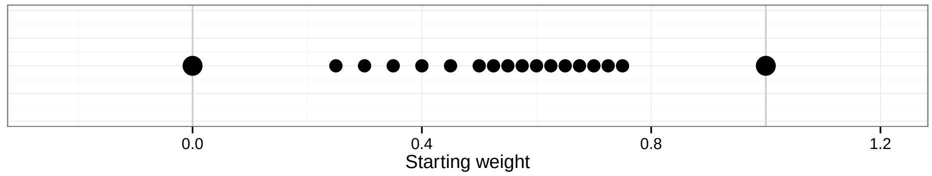


Figure 2.15: Overview of possible starting weights. Weight values are discrete, but staggered so as to provide fine-grained adjustments around 0.6 and more coarse discriminatory information toward the outside.

A trial begins with the presentation of a graph at the chosen starting weight w_0. Participants adjust the graph using increment and decrement buttons. A trial ends with the participant clicking the 'submit' button, at which point the weight for the final adjustment is recorded. This provides a clear starting value and ending value, allowing us to assess the range of optimal values for each participant. In addition to starting weight, correction type, and anonymized user-specific data (partial IP address, hashed IP address, and hashed browser characteristics), each incremental user chosen weight is recorded with a corresponding timestamp. The user-specific browser data is sufficient to provide a 'fingerprint' to distinguish and recognize individual users (or rather their computer settings) in an anonymous fashion.

Each participant is provided with two initial "training" trials in which the graph of the underlying mean function is superimposed on the line segments to give participants some idea of the function the lines represent. This approach was taken to reduce incidences of extremely high correction values under the X transformation, as large adjustment values do not change the impression of same line length, but the resulting function bears little resemblance to a sine function, see figure 2.16 for examples of overcorrection.

2.3.2 Results

Participants were recruited from Amazon Mechanical Turk and the reddit community.

As this study was conducted outside a laboratory setting, we can not gauge a participant's willingness to follow the guidelines and put in their best effort. This, besides potential technical issues (server outage, speed of response) make a careful selection of data going into the analysis unavoidable. The following exclusion criteria were used:

- Participants did not interact with the applet: we required participants to use the adjust-

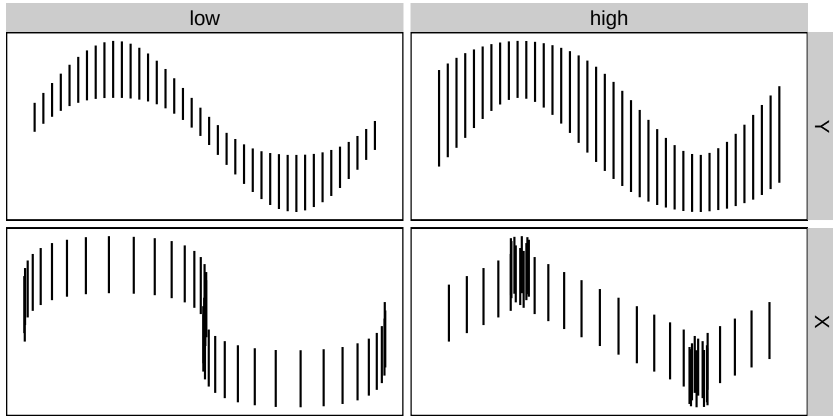


Figure 2.16: Transformation weights outside of the intervals $[-2.5, 3.5]$ for y and $[-2, 2]$ for x produce figures which do not maintain the underlying function shape (in x) or which are composed of extremely uneven length lines (in y). Trials with final results that were more extreme than these examples were excluded from the analysis.

ment at least once in order to include data for this trial (592 trials removed).

- Participants finished fewer than four trials: while participants were asked to complete twelve trials, some did not finish all of those. In order to stabilize predictions of random effects, participants' data were excluded if there were fewer than four trials (78 out of a total of 203 participants).

- Out-of-bounds results: weights leading to severely over- or undercorrected results were excluded from the analysis. For trials to adjust Y-values, weights outside of $[-2.5,\ 3.5]$ show dramatically unequal line lengths; weights from X-transformations outside the range of $[-2,\ 2]$ do not preserve the underlying function shape and concavity. Figure 2.16 shows results at the threshold of acceptability. Only more severely distorted results were excluded from the analysis (12 of the X and 5 of the Y trials out of 1227 trials remaining after application of other criteria).

The following analysis is based on the cleaned data, consisting of 125 participants with 1210 valid trial results. The psychophysics model shown in figure 2.17 is based on weighted averages (by adjustment type) of all trials with starting weights $w_0 = 0$ and 1.

According to this analysis, the optimum transformation value for x is 0.35, and the optimum transformation value for y is 0.45. Figure 2.17 shows the estimates and 95% Wald intervals for

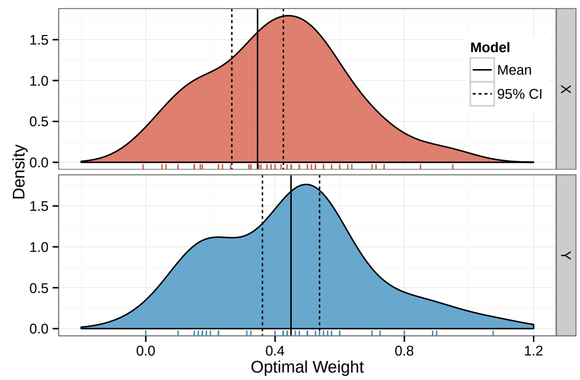


Figure 2.17: Estimated density of participant-level means using the standard psychophysics method of limits analysis. The overall means are both near 0.4, however, there is quite a bit of user-level variability.

the mean, as well as estimated density of participant-level responses.

While these results suggest that the transformation is useful and that complete transformation is not necessary, we can get more precise bounds on the range of acceptable transformation weights using a linear model that can incorporate starting points other than 0 and 1, and at the same time allow for user-specific variability.

In order to account for user-level variability, we fit a random effects model for the adjusted weight value as a function of starting weight and trial type, with a random intercept for each participant.

Let W_{ij} denote the final adjustment to weight by participant i, $1 \leq i \leq 125$, on trial j, $1 \leq j \leq n_i$. We model the final weight W_{ij} as a function of the correction type $T(i, j)$ (where $T(i, j) \in \{X, Y\}$), and starting weight X_{ij}, with a random intercept for participant to account for subject-specific ability:

$$\begin{aligned} W_{ij} &= \alpha_{T(i,j)} + \beta X_{ij} + \gamma_{i,T(i,j)} + \epsilon_{ij} \\ \gamma_{iX} &\overset{\text{i.i.d.}}{\sim} N(0, \eta_X^2),\ \ \gamma_{iY} \overset{\text{i.i.d.}}{\sim} N(0, \eta_Y^2), \\ \epsilon_{ij} &\overset{\text{i.i.d.}}{\sim} N(0, \sigma^2) \text{ and } \mathrm{Cov}(\gamma, \epsilon) = 0 \end{aligned} \tag{2.11}$$

$\alpha_{T(i,j)}$ is either α_X or α_Y, describing the lower threshold of the acceptable range for each of the types of correction, while $\alpha_X + \beta$ and $\alpha_Y + \beta$ describe the upper thresholds for the respective correction.

We can therefore interpret β as the length of the interval of plausible weights. Additionally, this allows the interpretation of the quantity $(\alpha_* + \beta/2)$ as equivalent to the estimate of the optimal weight based on the psychophysics methodology.

The fitted model parameters are shown in tables 2.1 and 2.2.

Transformation	Threshold	Parameter	Estimate	95% C.I.
X	Lower	α_X	0.097	(0.045, 0.150)
	Upper	$\alpha_X + \beta$	0.625	(0.570, 0.682)
Y	Lower	α_Y	0.143	(0.097, 0.188)
	Upper	$\alpha_Y + \beta$	0.671	(0.626, 0.718)

Table 2.1: Fixed effect estimates of model (2.11) for the boundaries for reasonable weights. In parentheses, 95% parametric bootstrap confidence intervals are given based on model (2.11) (N=1000).

Groups	Correction	Parameter	Estimate	95% C.I.
Participant	X	η_X	0.171	(0.167, 0.247)
Participant	Y	η_Y	0.145	(0.107, 0.179)
Residual		σ	0.304	(0.290, 0.317)

Table 2.2: Overview of random effects for model (2.11), including 95% confidence intervals based on parameteric boostrap results (N=1000).

Table 2.2 gives an overview of the variance estimates. 95% confidence intervals are, based on 1000-fold paramteric bootstrap of model 2.11. All variance components are significant and relevant; variability within a single individual's trials is about half the size of variability across participants.

We use parametric bootstrap to generate responses for each correction type and each participant from the model, which we use to both create user-level densities, population-level densities, and bootstrap intervals for model parameters.

The variability of the random effects for each trial type is similar; but the model benefits significantly from allowing separate random effects for individual's variability by correction type (0.1452 and 0.1705 for Y and X transformations, respectively, as opposed to 0.3044 for the overall variability). The interaction between starting weight and trial type was not significant, however, and was thus removed from the model (p-value = 0.901).

Figure 2.18 gives an overview of the relationship between starting weights and user-preferred

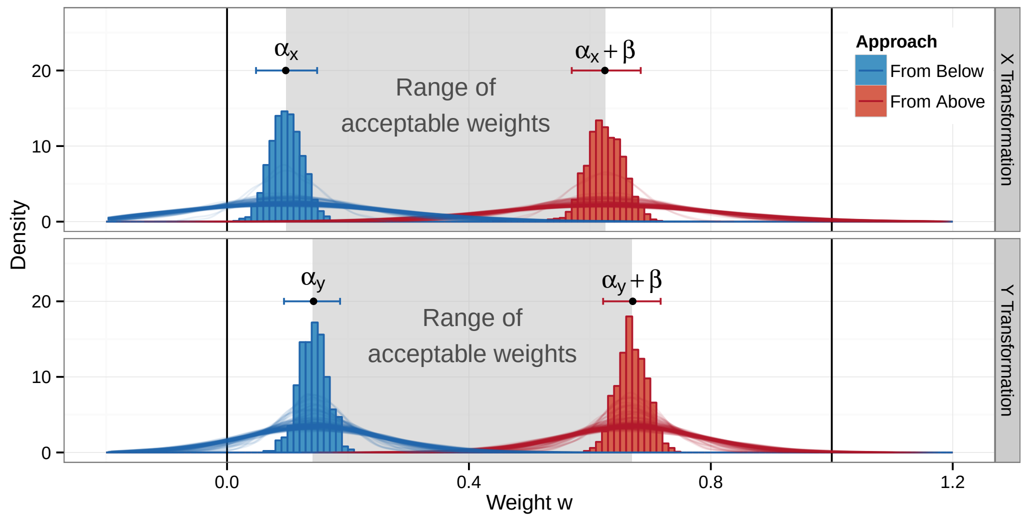


Figure 2.18: Simulation results from the fitted model, facetted by correction type. Fixed effects results are shown as histograms; the red values display the results when starting from an uncorrected plot and are concentrated around $w = 0.1$ for X and $w = 0.14$ for Y; the blue values represent user-chosen weights when starting from a fully corrected plot and are concentrated around $w = 0.63$ for X and $w = 0.67$ for Y. Additionally, 95% bootstrap intervals are shown as horizontal line segments above the histograms; these intervals are for the lower and upper bounds of the "preferred weight interval" tested in the experiment. User-level density curves show the individual variability around fixed effects α_* and $\alpha_* + \beta$.

weight values. Higher starting weights are associated with higher user-submitted values, and lower starting weights are associated with lower user-submitted values.

The ranges of optimal weights are similar under both transformations. Boundaries for the X transformation are slightly lower than boundaries for Y.

Bootstrap simulations for each of the coefficients suggest that the range of optimal w is between 0.098 and 0.625 for x and 0.142 and 0.67 for y, where the lower value is the estimate starting at $w = 0$ and moving up, and the upper value is the estimate starting at $w = 1$ and moving down. This suggests that either correction is preferrable to an uncorrected graph, and that a weighted correction is preferrable to the fully corrected graph, as neither 0 nor 1 is contained in any overall interval. In addition to showing the strength of the correction, this experiment also demonstrates the strength of the illusion itself: a correction appears more uniform than the uncorrected values, even though the corrected values are not uniform and the uncorrected values are completely uniform.

2.4 Application: US Gas Prices

Figure 2.19 shows daily gas prices for a time frame between 1995 to 2014 as published in the Energy Information Administration's historical database of gas prices (EIA, 2014b). This data includes prices for all three grades of gasoline as well as two chemical formulations which are sold in different geographic areas across the United States (for more information, see (EIA, 2014a)).

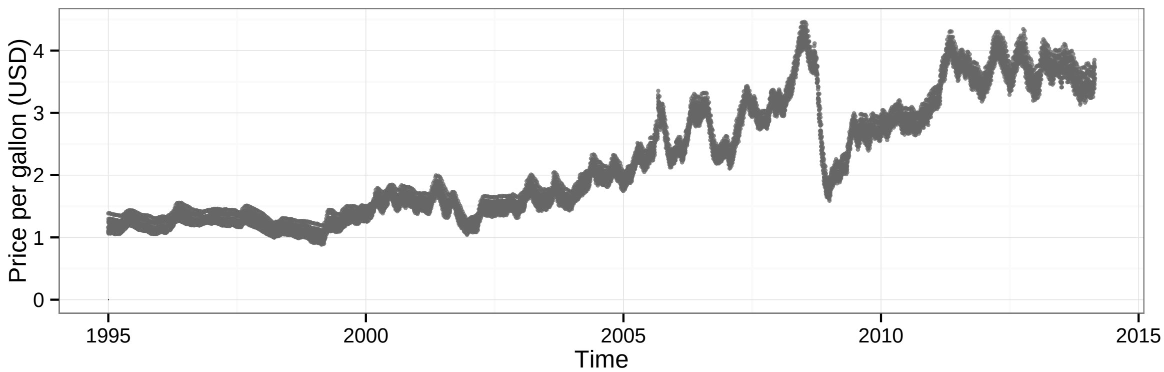


Figure 2.19: US Gas prices from 1995 to 2014. Gas prices steadily increase over the time frame, with some dramatic short-term developments. Peaks and troughs seem to exhibit more variability in daily prices than times of dramatic changes. This is an effect of the sine-illusion, which hides a fairly steady increase in variance in daily gas prices over time.

There is a clear increase in daily gas prices over time as well as several dramatic price changes. These developments mask the steady increase in variance shown in figure 2.20. Instead, we perceive an increase in variability in the frequent ups and downs along the overall trend. In particular, the strong decrease in gas prices at the end of 2008 seems to be associated

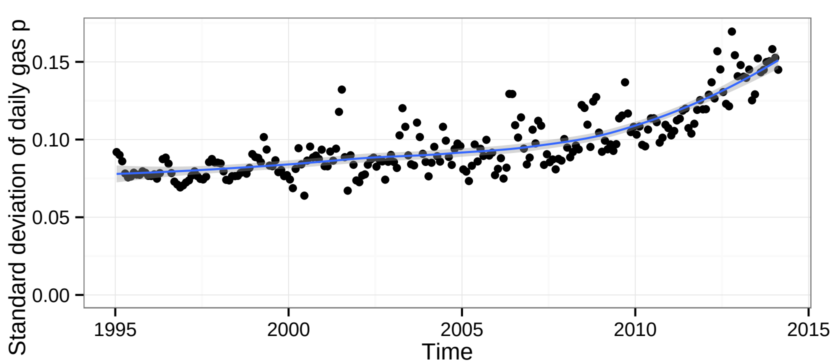


Figure 2.20: Standard deviation of daily gas prices between 1995 and 2014. The doubling of the standard deviation over the time frame is masked in figure 2.19.

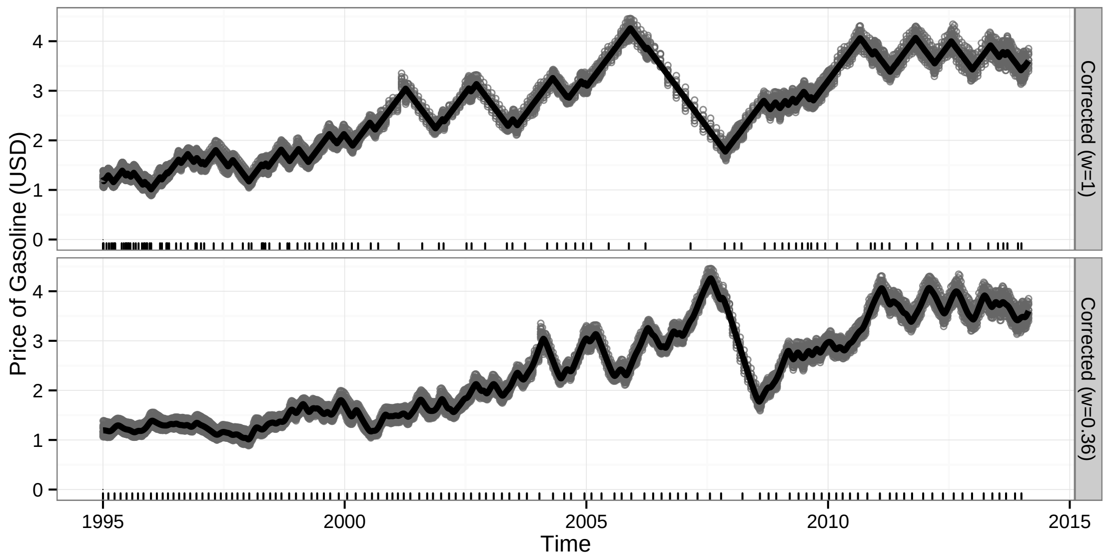


Figure 2.21: Gas price data corrected using the X transformation with $w = 1$ and with $w = 0.36$.

with a low variance. This is an effect of the sine illusion, and the actual variability in Oct 2008 is higher than previous months. In order to better judge variability along the trendline we applied the two different corrections to this data.

For either of the corrections we use a trendline fit based on smoothing splines, which provides the necessary first and second derivatives.

Figure 2.21 shows the results from the X transformation applied to the gas prices. The figure on top is a fully corrected version, while the one below only uses $w = 0.36$, the midpoint of the range of experimentally determined acceptable values, for the transformation. At $w = 1$, the transformation is severe, but it becomes clear that the variance between 1995 and 2000 is lower than it is between 2009 and 2014. When $w = 0.36$, the transformation is much less noticable but yields a near-constant absolute slope of the fitted line.

The minor effect of the weighted transformation on individual x-values contrasts with the effectiveness of the transformation in reducing the illusion; this is best seen in the fitted line, which is distinctly (piecewise) curved in the uncorrected data and appears to be much more piecewise linear in the corrected data, even at the reduced weighted value.

Similar to the X transformation, the Y transformation highlights local fluctuation in the variability of daily gas prices much more than the untransformed data. Figure 2.22 shows Y

transformations for the data. Again, we show a full transformation (top) and a transformation based on the midpoint of the previously determined acceptable region of $w = 0.40$. in the full transformation it is clear that the variance is nearly constant between 1995 and 2000 and then begins to increase with the price of gas. When $w = 0.40$, the transformation is much less noticable, and the resulting y-axis scale is much more similar to the uncorrected data.

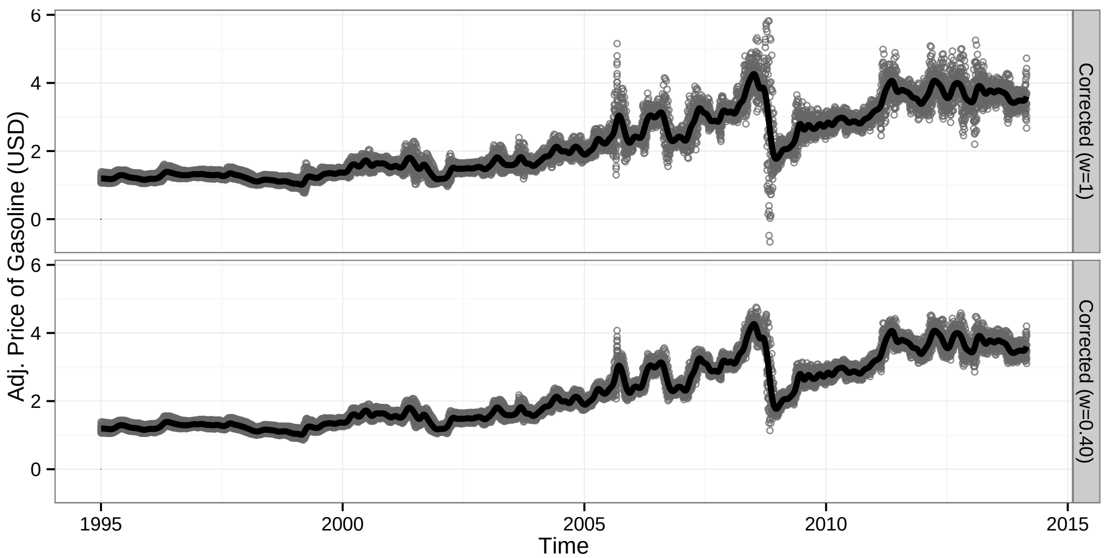


Figure 2.22: Gas price data corrected using the Y transformation with $w = 1$ and with $w = 0.40$.

2.5 Conclusions

The sine illusion is a persistent and powerful illusion that is very difficult to counteract without modifying the visual stimulus directly. While systematically modifying the data is uncommon in the statistical world, this approach is not out of place in the visual arts or architecture; as far back as 400 BC the builders of the Parthenon ensured a straight appearance of the columns from afar, by widening columns at the center, thereby counteracting the effects of the Hering illusion (Howe and Purves, 2005; Hering, 1861). Similarly, painters often exaggerate color hues used in shadows to account for color constancy in the brain. The systematic modifications we suggest here are also comparable to chloropleth maps, which scale a region's area based on some other variable.

We cannot counteract the illusion and represent the data visually without an intervention that is drastic enough to counteract the three-dimensional context the sine-illusion induces. The proposals in this paper for transformations in x and y provide the means to temporarily correct the data as a diagnostic measure, perhaps using an applet or R package for that purpose. These corrections are significant not only because of their implications for statistical graphics, but because previous attempts to resolve optical illusions using geometry have not met with success (Westheimer, 2008). These corrections are only a first step and could be improved upon; currently, the corrections break down for extreme (secant) values, but multiple iterations of the correction procedure will likely resolve some of these issues (though iteration removes the convenience of a functional form for the transformation). Similarly, the y corrections proposed here extend the line lengths (or for actual data, increase the deviation from the smooth line) – some normalization might make the necessary corrections less noticeable.

Our primary goal is to raise awareness of the illusion and its implications for statistics; the use of plots to guide the modeling process can leave us vulnerable to overlooking changes in the variance due to the illusion. While best practice has been to plot the residuals separately, this removes the context of the data and is not practical before there is a model. In addition, viewer attention spans may be limited if multiple graphs are presented. The proposed transformations require only a nonparametric smooth, maintain the context of the data, and are readily interpretable.

The data for this study was collected with approval from IRB-ID 13-257.

CHAPTER 3. VISUAL REASONING AND STATISTICAL GRAPHICS

Paper about lineup ability and army tests of visual reasoning

CHAPTER 4. STATISTICAL GRAPHICS AND FEATURE HIERARCHY

Paper about lineups and feature hierarchy

BIBLIOGRAPHY

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.

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.

Byron, L. and Wattenberg, M. (2008). Stacked graphs – geometry and aesthetics. IEEE Trans. Vis. Comput. Graph., 14(6):1245–1252.

C, C. M. and Wickens, C. D. (1987). Information integration and the object display an interaction of task demands and display superiority. Ergonomics, 30(3):511–527.

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.

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.

Coren, S. and Porac, C. (1983). Subjective contours and apparent depth: A direct test. Attention, Perception, & Psychophysics, 33(2):197–200.

Day, R. H. and Stecher, E. J. (1991). Sine of an illusion. Perception, 20:49–55.

EIA (2014a). Gasoline and diesel fuel update. http://www.eia.gov/petroleum/gasdiesel/reformulated_map.cfm. [Online; accessed 28-Feb-2014].

EIA (2014b). Weekly retail gasoline and diesel prices. http://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_w.htm. [Online; accessed 28-Feb-2014].

Environmental Protection Agency (2011). Air quality data. http://www.epa.gov/airdata/ad_data_daily.html. [Online; accessed 25-Oct-2013].

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. L. (1963). Distortion of visual space as inappropriate constancy scaling. Nature, 199(678-91):1.

Gregory, R. L. (1968). Perceptual illusions and brain models. In Proc. R. Soc. Lond. B, volume 171, pages 179–196. Citeseer.

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.

Hering, E. (1861). Beiträge zur Physiologie. Engelmann, Leipzig.

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.

Howe, C. Q. and Purves, D. (2005). Natural-scene geometry predicts the perception of angles and line orientation. Proceedings of the National Academy of Sciences of the United States of America, 102(4):1228–1233.

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.

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.

Morgan, M. (1999). The poggendorff 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.

National Climate Data Center (2011). Quality controlled local climatological data. http://cdo.ncdc.noaa.gov/qclcd_ascii/. [Online; accessed 25-Oct-2013].

Playfair, W. (1786). Commercial and Political Atlas. Cambridge, London.

Playfair, W., Wainer, H., and Spence, I. (2005). Playfair's Commercial and Political Atlas and Statistical Breviary. Cambridge University Press.

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.

Robbins, N. (2005). Creating More Effective Graphs. Wiley.

RStudio Inc. (2013). shiny: Web Application Framework for R. R package version 0.6.0.99.

Schonlau, M. (2003). Visualizing categorical data arising in the health sciences using hammock plots. In Proceedings of the Section on Statistical Graphics (JSM '03). American Statistical Association.

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.

Spence, I. (1990). Visual psychophysics of simple graphical elements. Journal of Experimental Psychology: Human Perception and Performance, 16(4):683–692.

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.

Tufte, E. (1991). The Visual Display of Quantitative Information. Graphics Press, USA, 2 edition.

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.

Varshney, L. R. and Sun, J. Z. (2013). Why do we perceive logarithmically? Significance, 10(1):28–31.

Wainer, H. (2000). Visual Revelations. Psychology Press.

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.

Weintraub, D. J., Krantz, D. H., and Olson, T. P. (1980). The poggendorf illusion: Consider all the angles. Journal of Experimental Psychology: Human Perception and Performance, 6:718–725.

Westheimer, G. (2007). Irradiation, border location, and the shifted-chessboard pattern. Perception, 36(4):483.

Westheimer, G. (2008). Illusions in the spatial sense of the eye: Geometrical-optical illusions and the neural representation of space. Vision Research, 48(20):2128–2142.

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., 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.

Wolfe, J., Kluender, K., and Levi, D. (2012). Sensation and Perception. Sinauer Associates, Incorporated, 3 edition.

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).