1-1-2019

Is a Picture Worth a Thousand Numbers? An Exploration of Understanding and Recall of Data Presentations

Holly L. Roof
University of Denver

Is a Picture Worth a Thousand Numbers? An Exploration of Understanding and Recall of Data Presentations

Abstract

This dissertation reports results of a study with a quasi-randomized experimental component and a protocol analysis, or think aloud, component. The experimental component was designed to determine if people with no statistical training and people with some statistical training differed in their understanding and recollection of statistical information with varying degrees of complexity. Information was presented using data visualization techniques based on cognitive theory and compared to presentations using APA-style numerical tables of statistical output. The focus was on using empirically-supported graphical displays in PowerPoint presentations such as one might see at a research conference. Classroom groups of beginning and more experienced statistics students (n = 194) were randomly assigned to watch one of two scripted PowerPoint presentations; one presentation predominantly utilized graphs while the other depended on tables to present the same information. Participants were tested for understanding immediately after viewing the presentations and two weeks post viewing to test their recall of the material. Protocol analysis was used to illuminate the thought processes of individuals with advanced statistical training as they interpreted either the graphs or tables.

Experimental results indicate large effects for complexity and time, and a small positive effect for the graphs treatment. Significant interactions in favor of the graphs treatment were found with novices on easy items in round 1 and for advanced beginners on difficult items for the advanced beginners in round 2. Protocol analysis found that advanced statisticians use the slide title to cue processing and interpretation of the slide content regardless of presentation type, however, they reached the interpretation stage more rapidly and directly when presented with graphs. Results support the use of graphs to enhance understanding and recall of empirical research presentations and present new findings to advance researchers', statisticians', and evaluators' impact, and enhance communication in the classroom and boardroom.

Document Type

Dissertation

Degree Name

Ph.D.

Department

Quantitative Research Methods

First Advisor

Kathy Green, Ph.D.

Keywords

Data visualization, Graphs, PowerPoint, Presentations, Protocol analysis, Tables

Subject Categories

Education | Educational Assessment, Evaluation, and Research

Is a Picture Worth a Thousand Numbers?

An Exploration of Understanding and Recall of Data Presentations

---

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Holly L. Roof

August 2019

Advisor: Kathy Green, Ph.D.

Author: Holly L. Roof
Title: Is a Picture Worth a Thousand Numbers? An Exploration of Understanding and Recall of Data Presentations
Advisor: Kathy Green, Ph.D.
Degree Date: August 2019

Abstract

This dissertation reports results of a study with a quasi-randomized experimental component and a protocol analysis, or think aloud, component. The experimental component was designed to determine if people with no statistical training and people with some statistical training differed in their understanding and recollection of statistical information with varying degrees of complexity. Information was presented using data visualization techniques based on cognitive theory and compared to presentations using APA-style numerical tables of statistical output. The focus was on using empirically-supported graphical displays in PowerPoint presentations such as one might see at a research conference. Classroom groups of beginning and more experienced statistics students (n = 194) were randomly assigned to watch one of two scripted PowerPoint presentations; one presentation predominantly utilized graphs while the other depended on tables to present the same information. Participants were tested for understanding immediately after viewing the presentations and two weeks post viewing to test their recall of the material. Protocol analysis was used to illuminate the thought processes of individuals with advanced statistical training as they interpreted either the graphs or tables.

Experimental results indicate large effects for complexity and time, and a small positive effect for the graphs treatment. Significant interactions in favor of the graphs treatment were found with novices on easy items in round 1 and for advanced beginners on difficult items for the advanced beginners in round 2. Protocol analysis found that advanced statisticians use the slide title to cue processing and interpretation of the slide content regardless of presentation type, however, they reached the interpretation stage more rapidly and directly when presented with graphs. Results support the use of graphs to enhance understanding and recall of empirical research presentations and present new findings to advance researchers', statisticians', and evaluators' impact, and enhance communication in the classroom and boardroom.

Acknowledgements

I would like to thank to the members of my dissertation committee for their willingness to help, and for their ideas, advice, and support. Thank you Dr. Olmos for setting me on this path that became this work. Thank you Dr. Asseffa for being there and offering your excellent ideas.

In particular, I would like to thank Dr. Kathy Green with my endless gratitude. Without her support I might not have finished. Her endless patience with my countless questions, ability to pull my head out of myriad rabbit holes and keep me focused, technical expertise, quest for perfection, and availability were all instrumental to this project.

To my friend, Kim, the only person in my real life to get what it means to earn a Ph.D. and who was there in the beginning when we said "where will we be in five years?"

To the many friends I have made at the University of Denver, who in their unique and individual ways helped me keep my sanity through this endeavor. Thank you, especially, to Priya, who from the other side of the world, provided support and advice, and to Lilian who I can always count on for a smile and to commiserate in the struggle between life and dissertation.

Finally, I would like to thank my family for not standing in my way as I pursued yet another degree from a university that is one hundred and thirty miles from home. In particular, I would like to thank my daughter, Bethany. She was my sounding board and biggest cheerleader.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

When I went to my first research conference, a conference for educational researchers, I was a doctoral student with advanced training in research methods and statistics. The conference attendees included a number of graduate students like me and researchers who are well established in their careers. In addition, a good portion of the attendees were professional educators, teachers from a variety of K-12 environments, who had come to learn what was on the cutting edge of their fields of practice.

The format of the conference allowed each presenter a maximum of twenty minutes, during which time the presenter needed to explain the research topic and get the audience to understand the results and their impact on the field. I sat through several presentations. One earnest young presenter explained her project, about computer games in the classroom, faithfully projecting her results tables. The last thing I wanted to do at that time was interpret unsorted tables of t-test statistics and p-values. As a trained researcher, I was able to interpret those results if I wanted to but what about all the classroom teachers in the room? Could they? Yet, they are the ones responsible for implementing the presenter's findings. Could it be that researchers can facilitate understanding of their results by changing their presentation strategies? Are some methods more effective than others at getting one's message across? To borrow a phrase from my K-12 teacher friends, what are the “best practices” for presenting research findings so that they are easily understood by a variety of audiences?

For researchers, whose studies might involve years of work, the expression ‘death by PowerPoint’ is more nightmare than joke. Significant personal and institutional capital are expended to learn something that could, potentially, change the world. For researchers, it is not enough to present results to interested audiences; for research to have lasting impact, the research results must be remembered, perhaps long after the presentation ends. Frequently, presentations include tables of results that require some amount of mental processing on the part of the viewer. Does this mental processing limit what viewers understand and remember from research presentations?

New research is frequently presented at professional conferences to interested audiences, so it is important to know if applying theories of visual processing and principles of graphic design to conference presentations support enhanced audience understanding. There are hundreds of professional conferences held each year with far-reaching impact. At the 2017 annual meeting of the American Educational Researchers Association (AERA), for example, 521 research papers were presented plus several hundred professional development sessions and symposia to an approximate 15,000 attendees (2017 AERA Annual Meeting Online Program Portal, 2017). For researchers and other presenters, it is important to know the best ways to present their work to audiences so that results are remembered after the conference ends.

Using PowerPoint, the presentation software produced by Microsoft, to project instructional material directly from a computer to a screen is common practice in university lecture halls (Mann & Robinson, 2009) as well as for conference presentations. Students prefer PowerPoint presentations over traditional (lecture and chalk board) presentations (Savoy, Proctor, & Salvendy, 2009). They think PowerPoints help them actively engage in learning, are the most effective teaching method, and enhance content understanding (Luse & Miller, 2011). However, evidence is mixed that PowerPoints actually lead to increased student learning (Savoy, Proctor, & Salvendy, 2009). “The research on PowerPoint® [sic] is not widely known and, as a consequence, is not reflected in classroom practices” (Berk, 2011, p. 24). In addition, little research exists to guide statistics teachers’ choice of graphs or tables when using PowerPoint as an instructional tool.

Presentations and lectures by researchers, educators, and others supported by PowerPoint, are ubiquitous (Kosslyn, 2007; Susskind, 2005). It is estimated that more than 30 million PowerPoint presentations are given daily (Lowenthal, 2009, p. 59). Knowing the best ways to reach one’s audience and maximize recall of the information presented is important for researchers to advance their fields of knowledge, for educators to have the broadest reach, and for all professionals to be heard and understood.

The Problem

Quantitative research rests on a foundation of numbers. Data are described using counts, means, standard deviations, and other numbers. Results of statistical tests are evaluated by comparing summaries of numbers. Statistical software packages present these numbers in tables making the comparisons relatively easy. Some presenters use similar tables to show their results. Others use graphs. Of course, a combination of the two is also seen. Is presenting research results in arrays of numbers the best presentation style to get audiences to remember the results? Or are there different types of data displays that are better to cement ideas in the memories of viewers? What is the best way to present empirical results so they are rapidly understood? Knowing the answers to these questions will help researchers, teachers, and other professionals influence comprehension and learning among their audiences.

Marketers and journalists have long recognized the importance of graphic imagery to deliver their messages. More than 40 years ago, Tukey (1977) developed a number of novel ways for statisticians to better explore data through visualizations such as box plots, and stem and leaf plots. Today, beginning statisticians are taught these visualizations as a matter of course (see for example Anderson, Sweeney, & Williams, 2015). Introductory business analytics texts include chapters on data visualization (for example Camm et al., 2016). There is a plethora of resources available on the internet from which people can learn about data visualization and tools for visualization are included in commonly used software such as Excel and PowerPoint as well as software dedicated to data visualization such as Tableau (www.tableau.com).

Nonetheless, little is known about how consumers of statistical research and reporting process and retain statistical information which they see presented at researchers' conferences, in evaluation reports, in the boardroom, and the classroom. Evergreen (2011) argues that better understanding of data visualization will position evaluators to "remedy communication-cognition gaps" (p. 2) in important ways resulting in increased the use of evaluations. Researchers, educators, and analysts in every field will benefit from greater understanding of effective communication strategies. This research presents new findings in the use of graphic visualization of statistical results to improve audience understanding and recall of quantitative information.

Purpose

Since PowerPoint slide presentations are a common communication method for presenting research findings, a problem for viewers is a gap between presentation style and their own visual processing realities, and issues for researchers are knowing what limits exist to bridge that gap and knowing how individuals approach the task of interpreting presented material. This dissertation focused on the role of graphs, visualizations of quantitative data typically using points, lines, or areas, in communicating statistical results to audiences with some statistical training through slide presentations. It does not address the use of infographics which utilize graphic design elements in representations of multiple forms of information to present information quickly and clearly (Infographic, 2019). Principles that facilitate understanding by drawing upon visual processing theory, graphic design best practices, and empirically tested graphic data displays were used in creation of PowerPoint slides. The purpose of this study was to test the hypothesis that people understand and retain statistical research results more readily when they are presented in graphical form as compared to tabular form. Further tested was the effect of information complexity on understanding and retention of results. The study was a 2 x 2 x (3 x 2) quasi-experiment—novice/advanced beginner x mode of presentation (graph, table) x item complexity (easy, moderate, difficult conceptual level) x time (immediate post-test, follow-up). Embedded within the study purpose was an examination of expert statisticians' thought processes as they interpreted graphical and tabular presentations.

The following sections provide a review of the literature. This literature review briefly summarizes visual processing theory, Cognitive Load Theory, and Gestalt principles of visual perception as frameworks for ways people process and retain visual stimuli. In addition, this review explores what is known about how people read and interpret different graph forms, and links graphic design principles with best-practices for effective PowerPoint presentations. It also presents information about understanding and recall by level of information complexity and about think aloud protocols.

Visual Processing Theory

According to visual cognition science, visual processing works in three stages. The first stage is called pre-attention. In this stage, the eyes continually scan the environment, noticing changes in basic attributes like color, motion, orientation, size, and contour (Malamed, 2009; Ware, 2013). In pre-attention, information is processed concurrently by neurons in the back of the eye without conscious thought. Occurring in large arrays, these neurons are specialized to extract basic features of the environment by receiving specific types of information such as color, orientation of lines and edges, or movement. Neurons extract features from the visual field simultaneously. For the viewer to understand information quickly, it must be presented so that it is easily detected by the neurons in the eyes. Pre-attention processing includes:

> rapid parallel processing; extraction of features, orientation, color, texture, and movement patterns; transitory nature of information, which is briefly held in an

iconic store; bottom-up, data-driven model of processing; serving as the basis for understanding the visual salience of element in displays (Ware, 2013, p. 21).

Our capacity to see these characteristics does not appear to be context-bound (Ware, 2013). When one of these characteristics is noticed, the eyes move to bring it into focus, where its features can be viewed by the fovea, an area of the retina that provides acute vision. It is called pre-attention because of the belief that it occurs without focused energy on the part of the viewer. In pre-attention, we might notice a movement in a bush, but would need to focus on the movement to distinguish a bird from a squirrel.

The second stage is called working memory. In this stage, active processes partition what we see into regions and simple patterns, such as continuous contours, areas of the same color, and fields of similar texture. This pattern finding stage is extremely flexible, influenced by the huge amount of information available from stage one and by the viewer directing attention to certain aspects of the visual field. Working memory is characterized by: serial processing which is slower than the parallel processing in stage one; attention that is guided by the viewer's visual interest; "a small number (one to three) patterns becoming 'bound' and held for a second or two under top-down attentional processes; different pathways for object recognition and visually guided hand motions (the perception and action channels)" (Ware, 2013, p. 22).

Working memory is defined as "the system for the temporary maintenance and manipulation of information, necessary for the performance of such complex cognitive activities as comprehension, learning, and reasoning" (Baddeley, 1992, p. 281). This is when the viewer attempts to make sense of what was noticed in the first phase. While something is in working memory, the viewer thinks about it, grapples with its message, and digests it (Baddeley, 1992). Patterns that were noticed before are now studied for meaning. People can hold a limited amount of information in working memory at one time; if the information is complex they hold less. Working memory does not retain chunks of information for long either (Cowan, 2000). If working memory is distracted or overloaded, some chunks of information will be dropped, possibly resulting in misunderstanding by or frustration in the viewer. As a result only the most pertinent or relevant information will be retained; the viewer will quickly glance over other chunks that are deemed unimportant (Woodman, Vecera, & Luck, 2003).

Visual working memory is the highest level of perception since the demands of active attention hold the objects in working memory. Only a few objects can be held in working memory at a time (Cowan, 2000); available patterns in the visual field merge into objects which, combined with information stored in long-term memory, may provide answers to the viewer's visual query. Our brains interface visual information with verbal information to connect words to images. Motor systems that control muscle movements are also interfaced with objects in working memory (Ware, 2013).

Working memory itself is comprised of three components: the Central Executive, Visuospatial (thought-processes that involve visual and spatial awareness) Sketchpad, and Phonological Loop. The Central Executive controls attention by choosing and organizing information from the environment. The Visuospatial Sketchpad and Phonological Loop work like two channels, one for visual/pictorial information and the other for auditory/verbal information (Mayer & Moreno, 2010), independently streaming information to the central executive. Since working memory capacity is limited to ± four bits of information (Cowan, 2000), and the channels have limited capacity (Mayer & Moreno, 2010), material that is presented in both visual and auditory modes capitalizes on these limitations (Baddeley, 1992).

The third phase of perceptual processing is called long-term memory. In long-term memory, new information is incorporated into existing mental schemas, or networks of information stored in the human brain. With the right balance of cognitive load in working memory at the most relevant time, objects, their attributes, and their message are encoded into long-term memory (Malamed, 2009; Ware, 2013). Occasionally the new information modifies an existing schema. An individual's culture and past experience impact how new information is received and adopted. When information is in long-term memory, individuals are able to recall and use it to make action-based choices. It is at this point in visual processing theory that comprehension is said to occur (Evergreen S. D., 2011).

While many aspects of working memory remain a mystery, it has been studied extensively. According to Plass, Moreno, and Brünken (2010), working memory capacity is divided into intrinsic load, extraneous load, and germane load. These three cognitive loads are the foundation of Cognitive Load Theory (CLT) which "links design characteristics of learning materials to principles of human information processing" (p. 1).

According to CLT, intrinsic load is generated by the inherent difficulty of the material and extraneous load is the cognitive burden caused by the design of the instruction and materials. Since intrinsic and extraneous loads are additive, germane load is the amount of mental effort that remains in working memory that can be invested by the learner towards processing and understanding the information presented. Easy material poses a low intrinsic load, while challenging material increases the intrinsic load (Sweller, 1994; Sweller, 2010; Sweller, Ayres, & Kalyuga, 2011). Extraneous load is caused by content or environmental elements that are unnecessary to learning the material. Examples of extraneous load in a presentation include slides with too much text, random colors, or chart junk (Tufte, 2001).

Finally, germane load is “the working memory resources that are devoted to information that is relevant or germane to learning” (Sweller, Ayres, & Kalyuga, 2011, p. 57). It is the effort required to attend to the material, mentally organize it, and form preliminary mental schemas. Because the difficulty of the material (or the element interactivity) cannot be altered, the assumption is that intrinsic load is relatively fixed by the content of the presentation even though its burden might vary from person to person. Since the remaining amount of working memory is left for extraneous and germane loads, instructional delivery has the potential to ease or compound extraneous load, thereby decreasing or increasing the amount of working memory available for learning (Sweller, 2010B)..

Several instructional effects that focus on reducing extraneous load have been identified by researchers (Sweller, 2010A). Two that are applicable to learning from presentations are the redundancy effect and the split-attention principle. The redundancy effect is caused by including the same information multiple times. This adds to the extraneous load because the viewer must use cognitive resources to sort necessary elements from unnecessary ones. An example of a redundancy effect that adds to cognitive load in a presentation is when the narrator reads the text from a slide. This requires the viewer to mentally compare the spoken words with the written to determine if they are the same or different. If the presentation does not require the learner to add steps to integrate the material, the redundancy effect is reduced and cognitive resources can be focused on learning.

The split attention principle of CLT is where the learner's attention must be split between multiple sources of information to mentally integrate the material. For example, when the legend for a graph is off to the side, the viewer is forced to alternately focus on the legend and then the graph while trying to remember the information contained in either. According to the split-attention principle, separating text and visual information increases extraneous load because it forces learners to use mental efforts to integrate the information. Physically integrating textual and visual information in instructional deliveries reduces extraneous load (Sweller, 2010B). If the elements are physically integrated, there is no need to mentally integrate the material (Sweller, Ayres, & Kalyuga, 2011; Sweller, 2016). Thoughtful design of instructional materials can reduce extraneous load, freeing working memory resources for germane mental processing.

Visual Perception

"Vision is by far our most powerful sense. Seeing and thinking are intimately connected" (Few, 2006, p. 78). Humans possess a vast memory for pictures (Standing, Conezio, & Haber, 1970; Vogt, 2007). The picture superiority effect (Paivio & Csapo, 1973), the apparent advantage that pictures have over words for object recognition, association, and memory recall tasks, is well established (see for example: Hockley W. E., 2008; Hockley & Bancroft, 2011;; Larkin & Simon, 1987; Seifert, 1997; Stenberg, 2007). A possible exception to the picture superiority effect might be for verbs (Hung, Edmonds, & Reilly, 2016). If pictures are more readily recalled than words, then the more visual material an instructional or informational message contains, the more likely it is to be recalled.

> When visuals are used effectively, they serve to help people understand abstract, complicated, and complex information, especially when people are unfamiliar with the concept and do not have a pre-existing mental model to assist with the comprehension of new information (Dunlap & Lowenthal, 2016, p. 44).

Because visual representations incorporate multiple parameters, they can tell a richer story of cause and effect or any other relationship than data points alone. Furthermore, the amount of information that can be held in working memory can be increased by chunking multiple individual elements into a single element (Sweller, 1994) which is an advantage of graphic displays.

Gestalt Principles of Visual Perception.

Current theory and practice in graphic design incorporates visual processing theory as a central feature of the way design communicates with an audience. Gestalt principles help us understand how we perceive pattern, form, and organization. Applying Gestalt principles to intentionally tie data together, separate data, or distinguish aspects of the display (Few, 2006) can guide the viewer’s attention and understanding. Six principles are particularly relevant to graphic design. These principles are: proximity, closure, similarity, continuity, enclosure, and connection.

The principle of proximity means that humans tend to group items that are physically close. Designers can use the principle of proximity to guide viewers' attention in a direction, for example top to bottom or left to right, simply by structuring the visual elements into a horizontal or vertical pattern (Few, 2006).

The principle of similarity indicates that people group items that are visually similar whether that similarity is color, shape, size, or orientation (Few, 2006). The principle of enclosure can also be used to visually group items by bounding a group of objects with a line or including them in a shaded region (Few, 2006). The tendency to complete outlines in order to perceive whole structures even when parts are missing is called the Principle of Closure. This principle can be used by graph designers to eliminate visual clutter from unnecessary graphic elements such as borders (Few, 2006). The Principle of Continuity encapsulates the idea that people perceive a continuous whole if the objects appear to align with one another or if they appear to be a continuation of each other (Few, 2006). This is easy to see in those 3-piece lawn ornaments that include a serpent's head, a body loop, and a tail. If the pieces are placed in a line with a little distance between each piece, people readily perceive a complete serpent swimming through the grass. The principle of connection indicates that people perceive objects as belonging together if they are connected in some way. Linking dots in a plot with a line creates such a visual connection.

Coupling pre-attentive attributes of visual perception and Gestalt principles provides a useful set of tools for meeting the challenges of making important data stand out and linking distinct evidence “in a way that makes sense, gives it meaning and supports its efficient perception” (Few, 2006, p. 95).

Data Visualization

Visual processing theory can help us appreciate how people understand graphs and graphic displays. Understanding graphs and other visual-spatial displays involves three processes (Shah, Mayer, & Hegarty, 1999). First, in the pre-attention stage, the visual system grabs hold of major patterns by recognizing the features of the graph such as straight or jagged lines, parallel or converging lines, color and shape, and encodes these features in a mental representation. What gets encoded depends on the viewer’s attention which in turn depends on the viewer’s goals and expectations, and what aspects of the display are most salient (Ratwani & Trafton, 2008).

Next, working memory translates the patterns into conceptual or quantitative representations. This process involves identifying the representations that are referred to from labels and titles. The viewer also applies existing knowledge of display conventions. Display conventions, or graph schema (Ratwani & Trafton, 2008) include the meaning of axes, knowing what type of data are typically displayed, what is typically omitted, what is literal, and what is not. When interpreting a graph, viewers activate their graph schema enabling prior knowledge and learned processes to be applied to the new problem (Ratwani & Trafton, 2008). Finally, the display’s patterns are interpreted for qualitative and quantitative meanings with domain knowledge affecting understanding (Pirrung, 2015).

In the pattern-recognition process, people first encode graphic patterns, then they incrementally interpret the patterns to retrieve or build qualitative and quantitative meanings, and finally they integrate these meanings with the referents identified by labels and titles. Graph interpretation is an iterative process whereby the viewer scans the graph for interpretive clues and chunks the clues together (Carpenter & Shah, 1998). For example, the viewer perceives the pattern of the lines on the graph, then looks at the legend, then back at the lines to chunk the information presented in the legend with the line. Information-dense graphs take longer to process than those with less information (Carpenter & Shah, 1998). The time it takes to interpret a graph is closely related to the number of unique quantitative relations and/or functions that must be individually interpreted and integrated. Naturally, there is an influence of individual differences in graphic knowledge on the interpretive process (Shah & Carpenter, 1995; Carpenter & Shah, 1998).

Many factors, including the format of the display, influence what knowledge viewers are readily able to construct from a graph. It is not enough that graphs be technically correct to be readily understood. Graph construction plays an important role in how they are understood. An advantage of properly designed graphs over tables is that values can be combined into chunks of information. For example, values can be combined into lines on a graph. Giving values a simple visual shape helps people hold more information at one time in memory (Few, 2007) because the pattern of lines becomes a chunk that people can hold in short-term memory. Visual representations such as graphs, diagrams, and schematic pictures chunk information that, when used appropriately, can facilitate reasoning about abstract higher order relations (Gattis & Holyoak, 1996).

Shah and Carpenter explored what constitutes a visual chunk with college students. They found that distinctive two-variable *x-y* functions such as those commonly used in scatter plots or line graphs are units of encoding but *z-y* functions are not (1995). In some situations, viewers lack the knowledge to associate the visual chunk to the quantitative referent so for them interpretation takes longer because their chunks are inherently smaller (Carpenter & Shah, 1998). In other situations, individual visual chunks may not be associated with the relevant data so viewers must rely on complex inferential processes. Such processes involve quantitatively transforming the information in the display. Graph designers can guide a viewer's cognitive processing of the graphs so that he or she is more likely to represent the data as the author intended (Shah, Mayer, & Hegarty, 1999).

Today graphic representations of data are commonplace, however cognitive scientists argue that representations that contain identical information are not necessarily computationally equivalent (Larkin & Simon, 1987). To be computationally equivalent, a viewer must be able to make the same inferences from either representation with the same amount of cognitive energy. Evidence that task performance differs with different visual displays of the same information (Breslow, Trafton, & Ratwani, 2009; Shah & Carpenter, 1995) argues for the importance of visual display design and, perhaps, supports CLT.

External visual-spatial representations are symbols for objects, events, or other data (Hegarty, 2011) and can be used to represent abstract relationships. For example, a scatterplot shows the relationship between two variables. The variables can represent things and properties that are not necessarily visible or tangible. Color, shape, and location represent dimensions of the display but these dimensions can be any category or quantity. Graphs are external representations that can store information thereby freeing working memory for other thinking (Scaife & Rogers, 1996).

Representations based on abstract information can make problem-solving easier by reducing the cognitive effort required to solve the problem. In addition, elements in a graphical representation constrain the kinds of inferences that can be made about the underlying information. The more closely the elements in the visual display match what is represented, the easier it is to control conclusions highlighting the cognitive benefits of good design (Scaife & Rogers, 1996).

One advantage of visual displays is that they can allow offloading of intellectual processes onto perceptual processes (Hegarty, 2011; Scaife & Rogers, 1996). When quantitative data are converted into visual variables, hidden patterns sometimes emerge that can be easily picked up by the visual system. This enables complex computations to be replaced by simple pattern recognition processes, thereby reducing intrinsic load. Also, offloading cognition on perception occurs when proper representation of a problem limits plausible conclusions (Scaife & Rogers, 1996).

Another advantage of visual-spatial displays over sentential (language-like) representation is they organize information within space. Similar entities are visually grouped. By placing entities on x- and y-axes, they are visualized as close together (Hegarty, 2011). Moreno and Mayer demonstrate that students performed better on a transfer test when on-screen text was placed next to the corresponding element in an animation compared to when it was placed at the bottom of the screen (Moreno & Mayer, 1999, cited in Mayer & Moreno, 2010). They refer to this concept as the "spatial contiguity principle – placing on-screen text near corresponding elements in the screen" (Mayer & Moreno, 2010, p. 142). In CLT this is known as the split-attention principle.

Graphs

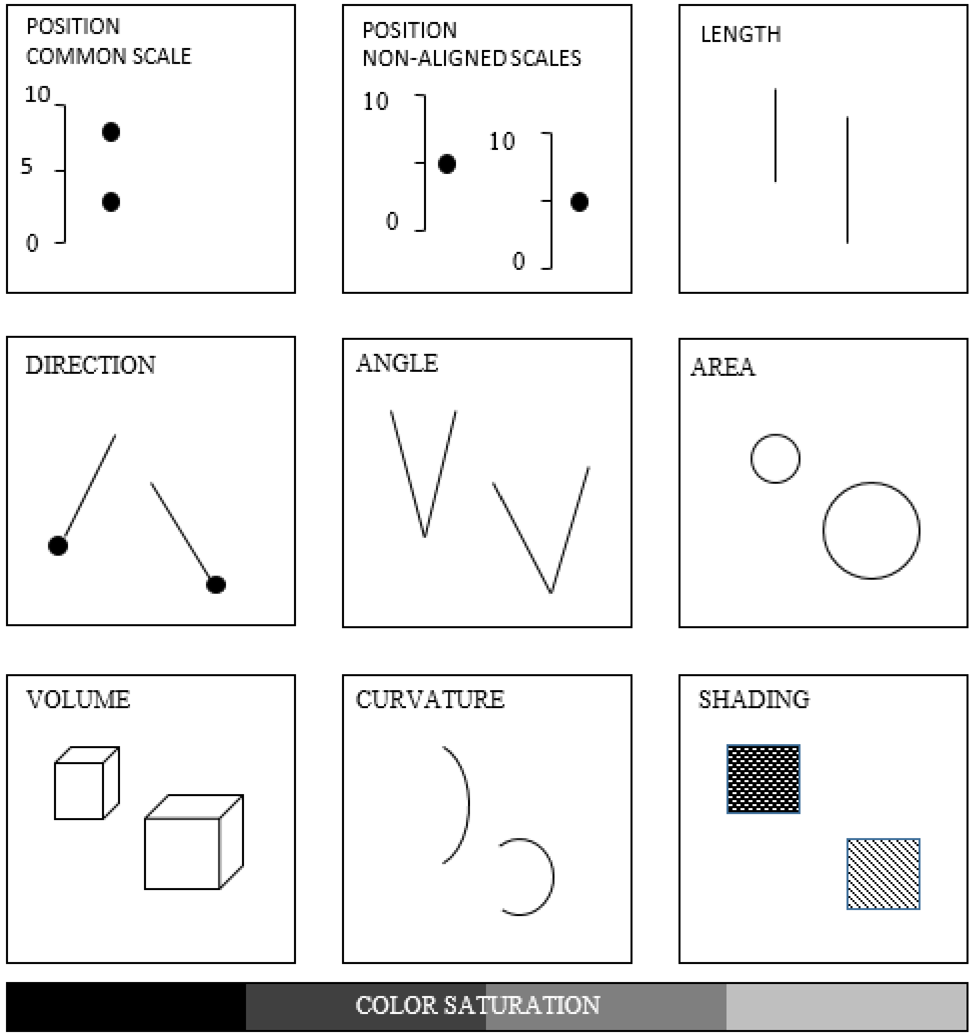
Graphic visualizations have been helping people understand data since William Playfair first proposed using graphics to convey "quantitative phenomena" (Wainer, 1990, p. 343). Charles Minard's mapping of Napoleon's disastrous 1812 campaign against Russia (see, for example, Kosslyn, 2007; Tufte, 2001, 2006) and John Snow's mapping of the water pumps in London that pinpointed the source of a cholera outbreak in 1855 (Tufte, 2001) are early examples of effective visualizations that illuminate what was previously unseen in the data.

Tukey (1977) developed novel ways to explore data through visualization, advising researchers to look at the big picture of their data by plotting "as in the large, so in the small" (p. 125). His work is foundational to many graphing procedures that are now included in most spreadsheet and statistics software programs. Today, box plots, histograms, scatter plots, and other visualizations are recommended in the early stages of data exploration by introductory statistics texts without attribution (Anderson, Sweeney, & Williams, 2015; Camm et al., 2016).

While Tukey focused on graphical exploration of data by statisticians, William Cleveland's concern was using graphs to transmit understanding to others. Cleveland says, "Visualization is an approach to data analysis that stresses a penetrating look at the structure of data. No other approach conveys as much information" (1993, p. 5). Advantages of data visualization include: providing a way to rapidly understand a large amount of data while facilitating understanding of both large-scale and small-scale features of the data, allowing users to see unanticipated properties and enabling identification of problems in the data, and helping with hypothesis formation (Ware, 2013).

People extract quantitative information from graphs only if decoding is effective. According to Cleveland (1984, p. 3) "there are many special considerations that arise when a graph is made to present data to others." Visual decoding starts with the "instantaneous perception of the visual field that comes without apparent mental effort. …what distinguishes them from tables—comes from the ability of our retentive visual system to detect geometric patterns and assess magnitudes" (Cleveland & McGill, 1985, p. 828).

Researchers have compared viewer understanding of specific types of graphs. In a series of experiments, Cleveland and McGill identified viewer comprehension of some graph types as superior to others and developed a paradigm for graphical perception based on the isolation of elementary codes of graphs. The elementary codes, geometric patterns detected by the preattentive visual system, are fundamental geometric, color, and textural aspects that encode the quantitative information of a graph. The "elementary perceptual tasks” (Cleveland & McGill, 1985, p. 828) ordered from most accurately perceived to least accurate are: “1) position along a common scale; 2) position along identical, non-aligned scales; 3) lengths; 4) angles” (Cleveland & McGill, 1987, p. 197) and slopes (aspect ratio of approximately 45 degrees); 5) areas; 6) volumes; densities; color saturations; and 7) color hues (Cleveland & McGill, 1985, 1987). See *Figure 1* for visualizations of the elementary perceptual tasks. The viewer performs one or more of these elementary perceptual tasks to extract the values of the variables represented by the graphs.



Adapted from Cleveland & McGill, Graphical perception: Theory, experimentation, and application to the development of graphical methods, 1984, p. 532.

Figure 1. Cleveland and McGill's hierarchy of graph types ordered from most accurate to least.

Graphic elements that are higher in the above list elicit judgments that are more accurate than elements lower in the list. Cleveland and McGill found viewers judge position more accurately than length by factors ranging from 1.4 to 2.5 and they judge position 1.96 times more accurately than angles. Consequently, they recommend using bar graphs over pie graphs and grouped dot graphs over divided bar graphs. See Figure 2 for examples of different graph types. Because position is judged more accurately than length, dot graphs replace bar graphs. Dot graphs help us visually summarize the distribution of the data (Cleveland, 1984).

Bar graphs outperform pie graphs when direct estimates of magnitude are required, and both bar and pie graphs are superior to tables as display devices (Spence & Lewandowsky, 1991). Bar graphs emphasize comparisons when variables are grouped together on the display while line graphs help viewers quickly understand quantitative trends by using the line to create a visual chunk. Line graphs stress the relationship between the variables while bar graphs emphasize differences in equally important independent variables (Shah, Mayer, & Hegarty, 1999). Graph processing is more accurate with line graphs in two dimensions than in three dimensions (Shah & Carpenter, 1995).

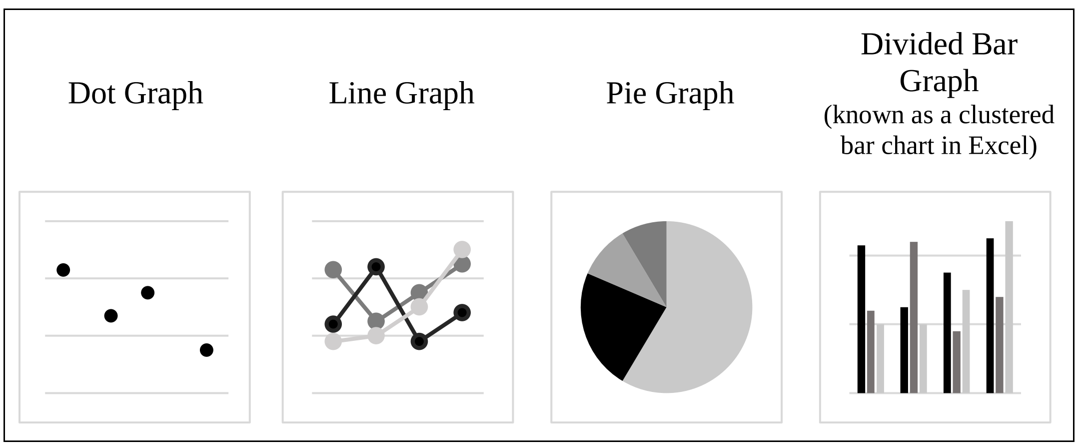


Figure 2. Examples of different graph types.

Cleveland said “the objective is to use an encoding scheme that provides high visual contrast so that we can focus on all of the values of one type of item, mentally filtering out the rest of the values” (p. 5). To that end, he introduced dot graphs as superior to bar graphs because the bars have length and width. The length and width of bars are visually encoded such that the area of the bar holds meaning in addition to the relative position of the end of the bar along a common scale. Since humans perform judgements of position along a common scale more accurately than judgments of length and area (Cleveland & McGill, 1984), bar graphs may lead to erroneous judgments with certain types of data.

To take advantage of the propensity to judge position along a common scale most accurately, designers should use dotted lines in a dot chart that has a meaningful baseline that ends at the data dots to make judging the position of the data dots or the lengths of the dotted lines easy to visually decode (Figure 3). If no meaningful baseline exists, dotted lines should go all the way across so line length does not signify an aspect of the graph that would hold meaning (Cleveland & McGill, 1985). See Figure 3.

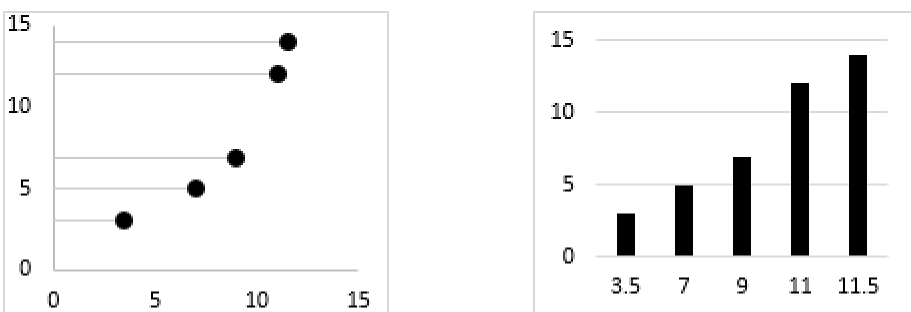


Figure 3. Examples of dot and bar graphs with meaningful x- and y-axes. Both graphs display the same data.

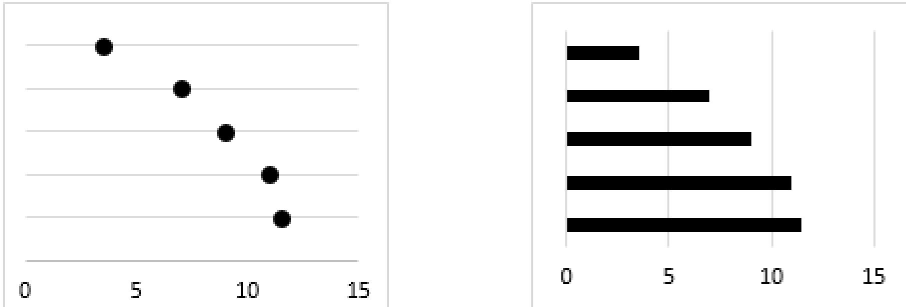


Figure 4. Examples of dot and bar graph emanating from the y-axis. Both graphs display the same data.

In an experiment, Cleveland, McGill, and McGill, (1988) showed that people judge slope ratios most accurately when they have a mid-angle of ± 45 degrees. Slope judgments are important for graphs that show how variable *y* depends on variable *x* because they give the best rough visual estimate of the rate of change. Inaccurate slope judgements can lead to inappropriate description of data and models. A graph's aspect ratio is the ratio of the height of the data rectangle to its width (Robbins, 2005). Because of the relative accuracy with perception to slope, designers should manipulate the aspect ratio to achieve slope ratios of approximately 45 degrees. If accurate judgment of the rate of change is important, designers should graph the rate of change directly rather than forcing the viewer to distill change from two trend lines. This way change is decoded by the more accurate judgments of position along a common scale (Cleveland, McGill, & McGill, 1988). Displaying differences on their own graph is also recommended to guide conclusions (Cleveland & McGill, 1984).

In a series of experiments, Ratwani and Trafton (2008) compared response times for different graphical patterns to measure comprehension of different graph types. They found response time was significantly faster for bar graphs than for line graphs and line graphs significantly faster than pie graphs. Response time for reading horizontal bar graphs was faster than line graphs. They also found response time for all three, line graphs, bar graphs, and pie graphs to be faster than for doughnut graphs (pie graphs with a hole in the center). They did not evaluate response time for dot graphs. They conclude that response time is fastest for bar graphs suggesting that bar graphs are best for extracting discrete values.

Other research indicates that logarithmic transformations of data are often more effective at showing variation in the data than original values, and that full scale breaks provide a clearer visual indication of the change in scale than partial scale breaks (Cleveland, 1984). Audiences react positively to the use of percentages in graphs (Brown & Newman, 1982). Multicolored scales allow faster and more accurate absolute-value identification than brightness scales but brightness scales are faster and more accurate on relative comparison tasks (Breslow, Trafton, & Ratwani, 2009). In addition, horizontally formatted pictographs are perceived faster and more accurately than vertically formatted pictographs and that shaded and one-graph pictographs are preferred (Price, Cameron, & Butow, 2007).

Tufte published a number of authoritative texts on design principles for visual displays of data. While it is unclear if his recommendations were empirically grounded or based on his own observations, he was concerned with excellence in using statistical graphics to transmit complex ideas with clarity, precision, and efficiency. He addresses many design elements such as the data-ink ratio, the grid system of a graph; “chartjunk” ”

(Tufte, 2006, p. 152); and other features with a focus on creating data displays that maintain integrity (Tufte, 2001) and beauty (Tufte, 2001, 2006). Data-ink is the non-erasable core of a graphic, the ink that is essential to the variation in the numbers represented; while chart junk is useless or optically active grids, boxes and frames, redundant representations of data, unrelated graphics (Bartsch & Cobern, 2003), and decoration. Maximizing the data-ink ratio reduces unnecessary mental processing required to separate the essential aspects of the message from the noise (Tufte, 2001). A main theme, which he demonstrates with numerous examples, is that the credibility of the data can be lost with poor design (Tufte, 2001).

Hegarty, Canham, and Fabrikant (2010) demonstrate that good displays should make task-relevant information salient. For example, color and line orientation can represent different variables in the data, and the display can be ordered so that important themes are clear. These display variables can affect the accuracy of task performance. They also provide evidence that domain knowledge, not just graphics conventions affect graphics comprehension.

The persuasiveness of data presented through bar graphs and line graphs compared with data presented through tables was explored by Pandey, Manivannan, Nov, Satterthwaite, and Bertini (2014). They found graphs to be more persuasive than tables for viewers whose initial attitude is not strongly polarized. The reverse was true with negatively polarized participants. Tables led to more participants with positive change than graphs, so presentation type may have an effect on persuasion and the effect may be moderated by initial attitude.

Breaking complicated information up into smaller pieces helps the reader integrate the information into existing understanding. Effective graphs do some of the heavy cognitive lifting for viewers by synthesizing, organizing, and grouping information into chunks allowing the quantitative information to be "absorbed more quickly than with other presentations" (Robbins, 2005, p. 225).

Graphic Design

Graphic design is the practice of combining text and images, or graphics, in visual media like advertisements, web pages, magazines and books, or PowerPoint presentations. Visual processing theory posits the way design is used to communicate with an audience. According to Evergreen (2011), "color, type, placement, and graphics comprise the basic aspects of design that have stemmed from the theories of visual processing. Thoughtful and strategic use of these aspects…support the reader's attempts to comprehend the material" (p. 28). When data are presented in certain ways, they form patterns that can be readily perceived. Following perception-based guidelines, data can be presented so that the important and informative patterns stand out. Disobeying the guidelines leads to data that are incomprehensible or misleading (Ware, 2013).

Knowledge of the three phases of visual processing is important for those creating slide shows and other visual media that has a likelihood of being retained in the long-term memories of viewers. For a visual display to be noticed in the pre-attentive stage, graphic designers support the most important information with such attributes as high contrast colors, movement, blinking, or large type size. To facilitate working memory, techniques need to support legibility and understanding. To maximize viewer attention in these two phases, graphic designers rely on an arsenal of variations in color, type, placement, and graphics.

Color.

Color is now a readily accessible tool to draw attention to selective elements of interest (Few, 2006). Background colors should generally be white or have very subdued colors, body text should be dark grey or black, and headings, short call out texts or other graphic elements can be highlighted with color to attract attention, guide understanding, and aid information processing (Ware, 2008). However, the use of color for emphasis can impede comprehension if too many colors are used indiscriminately; readers expect a change in color to indicate a change in meaning and they will spend time and effort trying to understand the meaning shift (Few, 2006; Malamed, 2009; Ware, 2008). In addition, colors that are too bright can distract the reader from the rest of the text, or be difficult to read if they do not sufficiently contrast with the background (Malamed, 2009; Ware, 2008). Red-green or blue-yellow color combinations should be avoided because they are difficult for people with color blindness to distinguish (Few, 2007; Ware, 2008).

Type.

Effective use of typeface, the specific pattern of letters, and font, the general appearance of the typeface such as italics or bold compared to normal weight letters, have been well researched as it applies to long passages of text (see for example Perea, 2013; Arditi & Cho, 2005). Anyone who has seen the work of a third-grader knows that dramatic or unusual fonts can be used to attract attention, but font influences more than attraction. Fonts must be consistent with the message of the presentation. For example, researchers can imply a level of professionalism with the choice of font used in presentations. Mackiewicz (2007) investigated qualitative properties of ten fonts, used in PowerPoint presentations by asking participants to judge fonts on "'professional,' 'interesting,' and attractive'" (p. 296). The fonts studied were Garamond, Times New Roman, Bookman Old Style, Arial, Verdana, Tahoma, and four that must be purchased separately. Results indicate that viewers perceived Times New Roman to be slightly more professional than Tahoma among the standard PowerPoint fonts but Tahoma was perceived as being slightly easier to read and more attractive than Times New Roman.

Fonts must also be legible and readable. Legibility refers to how easily readers can identify letter forms, while readability refers to the functional properties of the typeface. While studies of comprehension have shown serif fonts to be easier to read in long passages (Arditi & Cho, 2005), Mackiewicz also found that in PowerPoints there is no significant difference between serif and sans serif fonts in terms of comfortable to read, attractiveness, and interesting in presentations (2007). Additional specific recommendations for using fonts effectively in PowerPoint presentations can be found in Appendix A.

Placement.

Practical graphic design has adopted elements of the theory of Gestalt to predict how specific arrangements of information on a page will influence interpretation by the brain. Tourangeau, Couper, and Conrad conducted a number of experiments with surveys that identify how respondents use visual clues in interpreting questions. They developed a set of heuristics based on gestalt principles which assign meaning to spatial or visual clues. The five main heuristics for visual interpretation are: middle means typical, left and top mean first, near means related, up means good, and like (in appearance) means close (in meaning) (2004, p. 370; 2007, p. 94; 2013). For example, like means close means that viewers interpret items that appear close together as being connected, regardless of whether the closeness is in color, font, size, or physical proximity (Malamed, 2009; Tourangeau, Couper, & Conrad, 2004; 2007; Ware, 2008). Interpretations that perceptually grouped items belong together support the ability to comprehend graphs and other graphic elements (Shah, Mayer, & Hegarty, 1999).

Position on the page or screen determines what gets noticed first and enhances comprehension (Few, 2006; Malamed, 2009; Ware, 2013). The "up means good" heuristic means that viewers infer value of an item by its position on the screen (Tourangeau, Couper, & Conrad, 2013, p. 71). Viewers give more attention to elements located in key positions which are the top half and left side of a page. Size, color, orientation, and motion also emphasize key positions, which make manipulation of these elements another tool to capture attention and support comprehension (Malamed, 2009). Secondary, supportive, and explanatory information can be emphasized with smaller size, less contrast, or a position in the bottom half or right side of the page or screen. Differences may exist across cultural groups (Walton, Vukovic, & Marsden, 2002).

Viewers can also be directed to essential material with signals such as borders, headings, and highlights which decreases extraneous processing (Sweller, 2010A). Arrows, lines, numbers that rank order items, and compositional elements in a photo can also cue the viewer to pay attention to areas the designer wishes to emphasize. Such direction improves focus, facilitates processing, and increases understanding by providing visual cues to the audience. These steps activate more visual processing schemas, promoting retention and recall from long-term memory (Malamed, 2009).

Graphics.

Graphics include any non-textual elements and imagery in the two-dimensional space of the page, advertisement, or slide. Frequently artistic, photographs and illustrations create impact (Sherin, 2013) while arrows, stars, or other shapes can be used to signal important information (Shah, Mayer, & Hegarty, 1999; Sweller, 2010A). Graphs of quantitative data are regularly used by news media and others to help tell their stories (see for example: Rattner, 2017; Keneally & Diehm, 2015).

Graphs, charts, and to some extent, tables, rely on principles of color, type, and placement to get their messages across. In addition, graphic designers can support understanding by removing extraneous elements from the graphic (Few, 2006; Malamed, 2009; Tufte, 2001). Extraneous elements include three-dimensional displays, unnecessary gridlines, and color gradation, all of which Tufte refers to as “visual noise” (2001, p. 105).

Color, type, placement, and graphics are elements in the graphic design arsenal that have stemmed from the theories of visual processing. Thoughtful and strategic use of these elements works to attract the attention of the viewer, aids in decoding the information, and supports the viewer’s attempts to understand the material. PowerPoint is a program that supports graphic design projects with its easily manipulated text, lines, shapes, and images (Craig, 2017).

PowerPoint

PowerPoint slide presentations are now ubiquitous in lecture halls, seminars and webinars, on web pages, and elsewhere. PowerPoints are a staple in classrooms, conference rooms, and computer-based training (Savoy, Proctor, & Salvendy, 2009). In fact, PowerPoint presentations have been found to be preferred by students over traditional lectures (Apperson, Laws, & Scepansky, 2006; Susskind, 2005) and lectures with over-head transparencies (Bartsch & Cobern, 2003).

Durso, Pop, Burnett, and Stearman (2011) offer guidelines for creating effective slide presentations based on perceptual and cognitive principles relevant to PowerPoint slides. These guidelines cover such design aspects as font, color, layout, and tips to aid comprehension of textual and graphic slides. These guidelines are listed in Appendix A. Others have applied some of the principles of CLT to PowerPoints to improve their use as a learning tool.

The redundancy principle of CLT implies that on-screen text should not be used if it is repeated with audio or instructor narration (Mayer & Moreno, 2010) in PowerPoint presentations because the redundant nature of using duplicate modalities forces the viewer to consider both. Other studies have verified that the redundancy principle of CLT can be applied to PowerPoint presentations with strong results by eliminating text when narration is also a part of instructional delivery (Betancourt Lopez, 2014; Savoy, Proctor, & Salvendy, 2009). On the other hand, Mayer and Johnson (2008) found that guiding learners' attention by placing two- or three-word descriptions next to the appropriate visual information facilitated their learning. When small amounts of text are used as an "attention-guiding mechanism" (Mayer & Moreno, 2010, p. 11) the redundancy principle is not violated. In CLT, placing short descriptions next to the visual or diagram also satisfies the split attention principle and spatial contiguity principle. As such, extraneous processes are decreased. Betancourt Lopez (2014) found in experimental conditions that introductory statistics students reported a decrease in mental effort and improved retention and transfer scores after watching a PowerPoint that applied the split-attention principle or the redundancy principle compared to a control group. These studies suggest that instructional delivery can lessen the load on the visual channel by adhering to these principles.

Kosslyn (2007) takes a somewhat different approach and offers advice for creating effective PowerPoint presentations based on the use of eight psychological principles in support of three typical goals. The three typical goals of a presenter are: 1) connect with one's audience, 2) focus and hold attention, and 3) promote understanding and memory. Kosslyn elaborates that, in general, PowerPoints should be relevant to the audience's needs and the presenter's message. They should include audience appropriate language, concepts, and displays. Differences should be salient and discriminable. Presentations and individual slides should be organized to facilitate perception because people automatically group elements into units to help them pay attention to and remember the message. Additionally, the form of a message should be compatible with its meaning with changes in properties, such as color or font, signaling changes in meaning. Lastly, presenters should remember that people have a limited capacity to retain and process information, so they will not understand a message if they are overloaded with information (Kosslyn, 2007). Kosslyn's specific recommendations are included in Appendix A.

If the role of graphic design is considered useful in enhancing understanding in fields such as weather mapping, business analytics, and survey design, then researchers also have the potential to gain utility and increase audience understanding by considering elements of graphic design when disseminating results. Research indicates what types of graphs are more readily understood for which types of data, and what strategies work best to attract and hold the viewer's attention; however, this research considers graphs in isolation, comparing one graph with another without considering the broader context of an overall message that the presenter hopes to convey. To some extent, presentation best practices have been researched targeting the effectiveness of PowerPoint as a teaching tool. Limited research indicates that application of principles of CLT facilitates learning a statistics lesson. No studies comparing the efficacy of tables and graphs in PowerPoint presentations on understanding or recall were found. There is a gap in the research in regard to the effectiveness of graphs compared to tables in slide format (PowerPoint) for understanding of data and research results and if data visualization techniques support greater understanding and recall of presentations in their entirety.

Recall and retention by level of information complexity

Copious research exists indicating that recall and retention vary depending on level of difficulty of the information presented (for an example related to aging and executive function see Angel et al., 2016; for physiologic response to test taking see Kuhlman, 2014; for incorporating retrieval as part of learning tasks see Roelle and

Berthold, 2017; for learner age and multi-media instruction see Sierra, Fisk, and Rogers, 2002).

Research shows that children's working memory performance as measured with reading comprehension declines with more complex sentences (Magimairaj & Montgomery, 2012) and that children given an easy filler task during a retention interval performed better than those given a difficult task (Mahy & Moses, 2015). Furthermore, immediate and delayed recall falls as the cognitive load of the task increases (Camos & Portrat, 2015).

In contrast, correlations between memory span and comprehension were higher with moderate difficulty reading- or math-related background tasks compared to when the tasks were simple or difficult (Turner, 1989) suggesting that recall declines when an individual's memory span capacity is reached. However, the number of familiar features, not the complexity of features, in letter shapes affects the speed and capacity for encoding into visual working memory (Ngiam, Khaw, Holcombe, & Goodburn, 2018).

The number of errors made is also a function of information complexity. Subjects made fewer errors with easy memory span tasks compared to difficult tasks while spending more time on difficult memory span tasks compared to moderate span tasks (Conway & Engle, 1996). Subjects made more errors recalling a string of answers with more difficult mental arithmetic problems compared to easier problems (Conlin, Gathercole, & Adams, 2005). In addition, Irrazabal, Saux, and Burin (2016) demonstrated that subjects made more errors with a complex task compared to an easier task. These results suggest that information complexity plays a role in immediate and delayed recall as well as accuracy of recall.

However, multimedia instruction (PowerPoint with animations) improves comprehension and reduces the negative impact of information complexity compared to text-only and overhead instruction (Andres & Petersen, 2002). A meta-analysis of the use of cues (titles, labels, arrows, and other devices) to reduce cognitive load in multi-media presentations found that perceived difficulty decreased with cues and that retention and the ability to apply information to different contexts was improved (Xie et al., 2017) suggesting that PowerPoint presentations may be manipulated to enhance understanding and recall.

While the research on understanding and recall of information by level of difficulty in a wide variety of contexts is extensive, none was found investigating the effects of graphs compared to tables in presentations or the level of complexity of statistical analysis on understanding and recall.

Think Aloud Protocols

Verbal protocol analysis and, particularly, the think aloud method, were developed by Ericcson and Simon (1984). The think aloud method of protocol analysis involves asking people to think out loud while solving a problem, then analyzing the resulting verbal protocols to develop or test a model of the problem solving process (van Someren, Barnard, & Sandberg, 1994). “Verbal data is recordable behavior, which should be observed and analyzed like any other behavior” (Ericsson & Simon, 1984, p. 9). While the think aloud method was designed to explain cognitive structures, it has been shown to be a rigorous and methodologically reliable tool with applicability to different fields (Cansino, 2011; Taylor & Dionne, 2000; Yang, 2003). The verbal protocol "is a unique source of information on cognitive processes." (van Someren, Barnard, & Sandberg, 1994, p. xi)

"Adults have been observed to speak aloud spontaneously with intent to communicate" (Ericsson & Simon, 1984, p. 63) so verbal methods can be used to examine subjects' internal states. Verbal protocols are created by asking people to solve one or more problems while saying what they are thinking. These verbal reports are fundamental data which "require substantial interpretation and analysis to see their implications" (van Someren, Barnard, & Sandberg, 1994, p. 8) for development of a problem-solving theory. Think aloud protocol analysis provides a way to validate or create theories of problem-solving or other thought processes.

From the perspective of information-processing theory, thinking aloud requires the verbalization of the dialog in one's head while engaged in problem solving. The requirement for direct reporting of inner speech from short-term memory, without elucidation, limits the demand on mental resources and is essential to produce valid and reliable data (Ericsson & Simon, 1984; Newell & Simon, 1972). The information available in short term memory for verbalization is limited by what one is actually thinking during the problem-solving process (Ericsson & Simon, 1984)), however, verbalization does not capture all mental processes (White, 1980).

One reason why verbal reports of one's inner thoughts might not be complete is that only the fragments of thinking that are recognized by the thinker can be verbalized, so highly practiced thinking or parallel processing cannot be reliably reported (Ericsson & Simon, 1984). Similarly, information that is in working memory for a brief period, that is too complex to verbalize, or that can be characterized as non-verbal in character may be incomplete (van Someren, Barnard, & Sandberg, 1994).

Another reason verbal reports might be incomplete is that verbalizations are limited by the individual's capacity to simultaneously think and report thinking. So it is possible that additional information will not be reported because of competing demands for processing resources (Ericsson & Simon, 1992). Third, some types of information, such as goals and the steps taken to reach those goals, are more likely to be reported than other types of information (Pressley & Afflerbach, 1995 as cited in Taylor & Dionne, 2000). As a result, think aloud data do not produce a complete record of the thinking process, but rather produces a guide to thinking that permits the systematic tracing of the problem-solving process (Anderson J. R., 1987, Ericsson & Simon, 1992, Taylor & Dionne, 2000). Nevertheless, "instructions to think aloud do not alter the sequence of cognitive processes significantly" (Ericsson & Simon, 1984, p. 62).

In the think aloud method, subjects are asked to verbalize their thinking while performing some task. Occasionally, additional prompting might be needed during the problem-solving process to encourage them to keep talking. Subjects' verbalizations are recorded and transcribed later. The transcriptions are segmented based on phrasing and pauses. The segments are then coded into categories (van Someren, Barnard, & Sandberg, 1994). These codes become the data that allow for interpretation of the thoughts and behaviors employed in the problem solving process (Cansino, 2011).

By relying on this simple verbalization process, the think aloud method avoids interpretation by the subject. Since the output is available to anyone with an audio recorder, think aloud constitutes an objective method for analyzing problem solving and creating verbal protocols (van Someren, Barnard, & Sandberg, 1994).

Definitions

Axes – horizontal and vertical scales on which data are plotted. Typically, the *x*-axis is horizontal, the *y*-axis is vertical, and the *z*-axis represents a third axis if there is one. If there is a relationship between *x* and *y*, the *x*-axis represents the independent variable while the *y*-axis represents the dependent variable.

Advanced Beginners – Advanced beginners are individuals who have limited statistics training consistent with satisfactory completion of a college-level introductory statistics course. Advanced beginners in this experiment are students enrolled in the third term of a three-term sequence of statistics and business analytics courses.

Chart – The words graph and chart are frequently used synonymously. Since the word chart is also used to refer to such things as nautical charts, weather maps, and other visualizations, the word graph is used here to avoid confusion. Note: Microsoft Excel uses chart to refer to graph types.

Density – Shading or amount of black.

Expert – Experts have advanced training in statistics consistent with a doctoral-level education. Expert participants in this study have completed or are nearing completion of Doctor of Philosophy degrees in Research Methods and Statistics.

Graph – A graphical display of numerical information.

Glyph – A symbolic figure or character (Merriam-Webster Inc., 2017).

Graphical framework – The structural components of a graph.

Graphical pattern – The pictorial object of the graph (e.g., the actual lines in a line graph, bars in a bar graph, or slices in a pie graph.)

Hue – Color

Legend – Information that identifies coding of colors or symbols used in a graph.

Marker – The point on the graph that refers to a specific value.

Novice – Novices are individuals whose experience with statistics is consistent with the mathematics education of a high-school graduate. Novice participants in this study are in their first few weeks of statistical training and are enrolled in the first term of a three-term sequence of statistics and business analytics courses.

Plot – The area of a graph that contains the axis and corresponding numerical information.

Recall – Delayed memory of information.

Referent – The specific value represented by the data marker.

Scale – Units of measure used on an axis.

Scale break – A visual separation in an axis that represents a jump along the scale.

Saturation – Intensity (as saturation decreases the color becomes grayer).

Understanding – Immediate recall of information.

Chapter Two: Method

Research Questions

Is there an effect of statistical training (advanced beginners/novices) on understanding and recall of information provided?

Is there an effect of presentation (tables/graphs) on understanding and recall of information?

Does the degree of complexity of presented material impact understanding and recall of information provided?

Are there interactions between level of statistical training, mode of presentation, level of complexity, on understanding and recall of information provided in presentations of research results?

Do participants' personal characteristics such as interest in research topic, attitude toward empirical research, and level of statistical training predict assessment test scores?

How do people with advanced statistical training interpret graphical and tabular presentations of research results?

Null Hypotheses

H_{01}: There is no statistically significant main effect of level of training in statistics (Novice or Advanced Beginner).

H_{02}: There is no statistically significant main effect of treatment (graphs or tables).

H_{03}: There is no statistically significant main effect of time (Round 1 or Round 2).

H_{04}: There is no statistically significant main effect of level of slide complexity (easy, moderate, and difficult).

H_{05}: There is no statistically significant interaction between time (Round 1 and Round 2) and treatment (graphs or tables).

H_{06}: There is no statistically significant interaction between time (Round 1 and Round 2) and training (INFO1010 and INFO2020)

H_{07}: There is no statistically significant interaction between level of complexity and treatment.

H_{08}: There is no statistically significant interaction between level of complexity and training.

H_{09}: There is no statistically significant interaction between time and level of complexity.

H_{010}: There is no statistically significant between-subjects effect between level of statistical training and treatment.

H_{011}: There is no statistically significant interaction between time, treatment, and training.

H_{012}: There is no statistically significant interaction between level of complexity, treatment, and training.

H_{013}: There is no statistically significant interaction between time, level of complexity, and treatment.

H_{014}: There is no statistically significant interaction between time, level of complexity, and training.

H_{015}: There is no statistically significant interaction between time, level of complexity, treatment, and training.

H_{016}: Sex, major, number of quarters towards degree, number of statistics classes, US secondary education, research attitude, Colorado resident, issue interest, number of PowerPoints (PPTs) viewed in last month, number of PPTs created, treatment, and training do not significantly predict test score.

Quantitative Design

The quantitative portion of this project was a strong quasi-experiment with a cluster random assignment design (Gilner, Morgan, & Leech, 2009) of classroom groups to two treatments with three levels of information complexity as a repeated factor. The goal was to determine if there was a statistically significant difference in immediate understanding and delayed recall of research results presented with a slide presentation with results tables formatted in APA-style or presented using data visualization techniques based on cognitive theory and graphic design best practices. Tables are "characterized by a row/column structure" (Concise rules of APA style, 2011, p. 105), whereas, graphs (or plots) are characterized by a visual representation that typically shows relationships between numbers using lines, dots, bars, or other symbols (Graph, 2017). Randomization of clusters (classroom groups) is necessary to reduce the possibility of bias and confounding variables within the sample population and is the foundation for the assumption of independence between groups (Suresh, 2011).

Potential biasing factors include systematic differences within the make-up of each class such as preference for the teacher or class schedule. Because there was no reason to suspect differing ability from one class to another and because class membership was unrelated to the experiment, the potential for bias was negligible. A relatively large number of sections was available for participation which mitigated this threat as well. Additional threats to internal validity and strategies to manage them are noted in the following sections.

Participants.

Participants were undergraduates at the University of Denver enrolled in Analytics I or Analytics III. Analytics is a required sequence of statistics courses for all undergraduate business majors. The first in the series is Analytics I. Students enrolled in Analytics I generally have no prior coursework in statistics. They are typically in their first or second year of college, between the ages of 18 and 20. A prerequisite for students to enroll in Analytics III is satisfactory completion of Analytics I and II or an introductory statistics course such as Advanced Placement Statistics, so all Analytics III participants will have had some statistical training. Students in this course are typically second- through fourth-year college students between the ages of nineteen and twenty-two with majors in any department (finance, marketing, accounting, etc.) within the college of business. In both groups, approximately ninety-two percent were graduates of American high schools; the balance were international students. The number of male and female students was approximately equal. A summary of participants' demographic background is provided in

Table 1. Instructors of record were asked for permission to present to their classrooms. The experiment including introduction with informed consent paperwork, slide show presentation, and assessment took less than twenty-five minutes of classroom time. The second round assessment took about ten minutes of class time.

Table 1

Demographic Characteristics of Experimental Participants

Characteristic	Category	Count	*M*	*SD*	Skew	Kurtosis
Novice						
Sex (n = 96)	Female	42				
	Male	54				
Program (n = 93)	Business Major	81				
	Other Major	12				
Nation of Secondary Education (n = 93)	United States	89				
	Other	4				
Colorado Resident (n = 97)	Yes	40				
	No	57				
Age (n = 96)			19.04	1.70	3.22	13.05
Number of PPTs Viewed (n = 96)			2.06	0.89	0.31	-1.20
Number of PPTs Created (n = 95)			1.53	1.17	0.06	-1.20
Number of Quarters Toward Degree (n = 96)			1.71	2.28	1.28	0.87
Number of Statistics Classes Taken (n = 96)			0.31	0.53	1.46	1.23
Attitude toward Statistics (n = 95)			3.32	1.06	-0.78	1.42
Attitude toward Empirical Research (n = 95)			3.31	0.95	-0.27	1.13
Issue Interest (n = 96)			3.11	1.20	-0.93	0.83
Advanced Beginner						
Sex (n = 94)	Female	54				
	Male	40				
Program (n = 86)	Business Major	78				
	Other Major	8				
Nation of Secondary Education (n = 94)	United States	83				
	Other	11				
Colorado Resident (n = 94)	Yes	43				
	No	51				
Age (n = 94)			19.76	2.24	7.10	60.00
Number of PPTs Viewed (n = 92)			2.15	0.80	-0.55	-0.46
Number of PPTs Created (*n* = 92)			1.91	1.11	-0.32	-1.43
Number of Quarters Toward Degree (*n* = 91)			3.93	2.00	2.58	7.12
Number of Statistics Classes Taken (*n* = 93)			2.12	0.55	0.07	3.32
Attitude towards Statistics (*n* = 94)			3.32	1.21	-0.90	0.71
Attitude toward Empirical Research (*n* = 94)			3.39	1.02	-0.42	0.47
Issue Interest (*n* = 94)			3.39	0.87	-0.77	0.37

Demographic variables were compared across class groups to test group similarity. Extraneous experiences of participants were controlled to some extent with the selection of the Analytics I and III courses because Analytics I is the first course in the series and Analytics III is taken by undergraduates soon after completing their introductory statistics courses. Environmental variables were controlled within the classrooms as all sections were held in lecture halls that are indistinguishable from one another from the inside of the room.

Materials.

Two PPT presentations were created using descriptive statistics and analysis of data from 2016 Colorado Health Rankings (Colorado, 2016). See Appendix B for images of all PPT slides. One presentation displayed results using graphs and imagery designed using graphic design best practices and the other presented results using tables based on guidelines established in the Publication Manual of the American Psychological Association, Sixth Edition. Both PowerPoint presentations utilized elements of CLT to minimize confounding results from presentation techniques. Both presentations used the same font and slide titles as well. The title of every slide signaled the conclusion or important information viewers should glean from that slide. The presentations were recorded with scripted narration so that the presentations were identical except for the visual aspects of the slide content. See Appendix C for slide-by-slide script. The slides were ranked by complexity of information as easy, moderate, and difficult. See Appendix D for slide complexity and slide counts by complexity.

A panel of experts (see Table 2) reviewed the presentations to ensure that they adhere to the guidelines in Appendix A – Principles for Creating Effective PowerPoint Presentations, Appendix E – Principles for Creating Effective Graphs, Appendix F – APA Table Construction Guidelines. The panel judged that the slides were informationally equivalent and provided feedback for the slide rankings. Adjustments were made to slides based on expert reviewer comments to ensure informational equivalence; slides were then returned to the expert reviewers for final judgment.

Table 2

Expert Review Panelist Characteristics

Reviewer	Position/Education	Contributions
1	Lecturer/PhD	Slide and assessment review
2	Doctoral Student	Slide and assessment review
3	Doctoral Student	Slide and assessment review

Slides.

Slides for each of the two presentations are provided in Appendix B and the corresponding script is in Appendix C. As an example of graph/table pairing, slide 6 in Appendix B show cluster membership of Colorado's counties based on five health outcome variables. Slide 6 presents the cluster membership in a table which is formatted according to APA guidelines (Concise rules of APA style, 2011). The table has rules on the top and bottom, with a rule separating the table title from the column labels. White space is used to visually separate the columns and different clusters. Per APA guidelines, the word "NOTE" below the table is used to call attention to explanatory information Slide 6 shows the same cluster membership of Colorado's counties using a map of the state with each county color coded to show its cluster membership. A number of design principles have been applied to this slide. According to visual processing theory, color attracts attention in the pre-attentive phase so using color should attract attention to the slide. The gestalt principle that Like Means Close can be seen in the choice of colors. Since the Worst and Poor clusters include counties with poorer health outcomes than the Good and Best clusters, Worst and Poor clusters are color coded in different shades of blue while Good and Best clusters are colored in different shades of yellow. Since readers infer that upper left locations signal importance, the note that appears under the table in Slide 6 is located in the upper left of this slide. The reference to clusters "Worst," "Poor," "Good," and "Best" in the note are colored to match the color coding of the clusters, highlighting important information while doubling as a legend which reduces clutter on the slide. It is also possible that the recognizable shapes of the state and counties evoke an emotional response in some viewers which might help cement information contained in this image for those viewers. These two slides are informationally equivalent.

Assessment.

A twenty one-item assessment with multiple choice and true/false questions was created to gauge understanding and recall of the content in the presentations. The same panel of experts that reviewed the slide content (see Table 2) reviewed the assessment to provide support for content validity—the assessment logically measures understanding and recall of the material. Reliability was assessed with Cronbach's alpha. For the entire assessment $\alpha = .48$, for easy items $\alpha = .42$, for moderate items $\alpha = .41$, and for difficult items $\alpha = 0.13$ which indicates a low level of internal consistency and suggests a high degree of error variance for the assessment as a whole and at each level of complexity

(DeVellis, 2012). The correlation (r = .50) between the assessment given immediately following the slide shows and the same assessment with re-ordered questions administered after two weeks was statistically significant which indicates reasonable test-retest reliability.

Experimental Procedure.

Approval of the Institutional Review Board (IRB) was acquired before proceeding with the experiment. IRB assigned the project #1006318-1. Instructors of Analytics I (novice) and Analytics III (advanced beginner) classes were recruited via email. See Appendix G for recruitment materials. Intact sections of Analytics I and Analytics III were randomly assigned to view one of the two presentations. Scheduling was coordinated with each professor. The researcher entered the classroom according to the professors' directions. The professor introduced the researcher. The researcher then proceeded to introduce the study using the script approved by the IRB. See Appendix H for introductory materials. After the introduction, the researcher distributed a packet that included the Informed Consent form for the experiment and the assessment. Students were not required to participate and did not receive any compensation or course credit if they did. All sections except one were required by their teacher to remain in the classroom regardless of their choice to participate. The cover page of the assessment instructed participants not to open it until after the presentation.

After a few minutes pause to allow students to read the informed consent form the presentation was started. The presentations were displayed using the classroom's overhead projection system with the recorded audio broadcast over the sound system.

Slides advanced automatically. Each presentation lasted on average 9 minutes 51 seconds. Immediately following the presentation, participants completed the assessment. The assessment was administered using paper and pencil. The assessment questions are provided in Appendix I. This phase of the experiment took approximately 20 minutes of class time. The assessment was administered a second time at two weeks post-presentation to gauge recall of the material. Questions were re-ordered for the second test in an effort to minimize any practice effect. This phase of the experiment took approximately 10 minutes.

Analysis.

All analyses were completed using SPSS version 24. The research hypotheses were evaluated using ANOVA to estimate the significance of differences in means for each main effect and interaction of factors. A mixed-design analysis of variance (ANOVA) was used to compare means for the two groups on the immediate post- and follow-up tests. The mixed ANOVA model has two between-subjects factors with two levels each and two within-subjects factor with three and two levels. The between-subjects factors were treatment with 2 levels; level 1 is the PPTt presentation with graphs and level 2 is the PPT with tables; the second between-subjects factor was training level (novice, advanced beginner). The within-subject factors were the immediate post-test (R1) and the follow-up test (R2) and level of slide information complexity with three levels (easy, moderate, difficult). See Table 3 for an overview of the experimental design.

Table 3

Mixed ANOVA Design

Between Subjects Factors	Within Subjects Factors					
	Initial Understanding (R1)			Recall (R2)		
	Easy	Moderate	Difficult	Easy	Moderate	Difficult
Novice Group 1 - Graphs	$Score_{11}$	$Score_{12}$	$Score_{13}$	$Score_{21}$	$Score_{22}$	$Score_{23}$
Novice Group 2 - Tables	$Score_{11}$	$Score_{12}$	$Score_{13}$	$Score_{21}$	$Score_{22}$	$Score_{23}$
Advanced beginner Group 3 - Graphs	$Score_{11}$	$Score_{12}$	$Score_{13}$	$Score_{21}$	$Score_{22}$	$Score_{23}$
Advanced beginner Group 4 - Tables	$Score_{11}$	$Score_{12}$	$Score_{13}$	$Score_{21}$	$Score_{22}$	$Score_{23}$

A priori sample size for repeated measures, between factors ANOVA with four factors and six repetitions was calculated for a medium effect size (f = .25) and power of 0.70 (α = .05) using G*Power 3.1.9.2. Sample size was estimated as n = 88. Most Analytics I and III classes have 30 or more students so approximately four classes (two of each) were needed to get a sufficient sample allowing that some students would choose not to participate or be absent during the presentation or follow-up test. The final sample size was 193. Regression analysis was used to determine if demographic variables such as sex and level of statistical training predicted level of recall and understanding with either the graphic presentation or the tabular presentation.

Qualitative Design

The embedded qualitative portion of this project used the think aloud method (van Someren, Barnard, & Sandberg, 1994) of protocol analysis to reveal stages of the process of interpreting tables and graphs employed by people with advanced statistical training. In the think aloud method, participants are asked to articulate their thought processes while attempting to solve a problem with the goal of obtaining a “unique source of information on cognitive processes” (p. xi) involved in interpreting graphs and tables.

Participants.

Participants were six advanced graduate students in the Research Methods and Statistics Program at the University of Denver who had completed many quantitative analysis classes including analysis of variance, correlation and regression, multivariate statistics, hierarchical linear modeling, and structural equation modeling.

Materials and Procedure.

Participants were randomly assigned to view either the slide set with graphs or the set with tables. The same slide sets from the experimental portion were presented without the accompanying audio. After a brief introduction, and reading and signing the Informed Consent form, participants were given instructions to verbalize their thought processes while interpreting the slides. They were provided a brief warm-up slide show to familiarize themselves with the procedure after which they watched the experimental slide show. One participant saw the presentation on a mounted television screen, the other five saw it on the researcher’s laptop computer screen. Participants viewed each slide for the same duration as participants in the experiment during which they articulated their thinking. The researcher recorded the verbalizations with a hand-held recorder. No input from the researcher was needed other than an occasional reminder to keep talking. Participants completed the same demographic questionnaire and first-round assessment as participants in the experimental component. The time required of each participant was about twenty minutes.

Data Handling

Experimental Data Handling.

Immediately upon leaving a classroom, assessments were labeled with course and section. They were scored by assigning one point for a correct answer and zero points for a wrong answer. See Appendix I for question and answer content of the assessment and Appendix J for demographic questions.

During scoring for the first class that participated in the experiment, it was discovered that assessment Question 13, a multiple-choice question, did not include the correct answer. The assessment was changed for all other groups. Since the first group was unable to choose the correct answer, its response to this question is Missing Not at Random. Only one person in this first group got the answer correct on the second assessment, that person was given correct on the first and everyone else was given incorrect on the first.

Question 12: Which Colorado County had the best health outcome rank in 2016? In the first group two responses indicated the county with the cities of Parker or Castle Rock instead of the county's name which is Douglas. The city information was provided in the audio. The decision was made to count answers that correctly indicated the county by any information provided in the presentations as correct.

Assessment questions that were left blank were treated as incorrect. Questions that were left blank in the demographic questionnaire were treated as missing at random since no discernable pattern exists among the unanswered questions.

Sixty-three people who participated in the first round did not participate in the second round. These participants were removed from the data set.

Protocol Analysis Data Handling.

Each participant’s think aloud protocol was audio recorded by the researcher. Audio recordings were later transcribed. Transcriptions were divided into phrases based on pauses or transitions in the recordings. Since the verbalizations corresponded to specific slides in the presentations, the phrases were organized and, ultimately, analyzed by slide. A system of codes was developed and the codes were applied to every phrase. See table 4 for codes used.

Table 4

Codes used in Protocol Analysis

Code	**Meaning**
"Interesting"	Says "interesting" or uses the word in a phrase
"Makes Sense"	Says "makes sense" or uses the word 'sense'
"Okay" - U	Says "okay" to indicate understanding of slide content
"Okay" - T	Says "okay" to indicate transition in slides
Critique-Neg	Offers negative criticism of slide content or format
Critique-Pos	Offers positive criticism of slide content or format
Curious	Verbalization indicating curiosity for extension of information provided
Interpret	Verbalization indicating an attempt to interpret slide content
NA	Not coded (nothing said or banter with researcher)
Personal	Looks to apply slide content to own experience/life

Preparation	Phrase or statement indicating anticipation of data display
Process	Phrase or statement indicating mental processing of slide content
Question	Phrase or statement looking for missing/additional information to aid interpretation
Read	Reads slide text verbatim
Re-read	Reads part of slide during interpretations
Understand	Verbalization indicating understanding of slide content
Schema	Uses specific statistical vocabulary indicating schema in long-term memory

Once coding was complete, flow charts were created for each participant and each slide in an effort to visualize similarities and differences in the protocols. Analysis followed.

Chapter Three: Results

In this chapter, results from the quantitative phase and qualitative phase are reported in separate sections. The chapter begins with a section for the experiment which is followed by a section for the protocol analysis. Each section starts with a description of the participants in that phase. The quantitative phase includes a subsection with assumptions testing of the linear model, and finally, results are presented for both sections.

Quantitative Results

Description of participants.

A total of six novice and five advanced-beginner classes participated in the experiment with final sample size n = 193. Of the six novice classes, three watched the graphs PowerPoint (n = 49) and three watched the tables PowerPoint (n = 47). Three advanced beginner sections watched the tables PowerPoint (n = 60) and two watched the graphs PowerPoint (n = 37). The Round 1 (R1) assessment was administered immediately post treatment and the Round 2 (R2) assessment was administered to all groups exactly two weeks after R1.

Participants completed a demographic questionnaire. Complete demographic information is provided in Table 1. All participants were undergraduates. Participants were asked to list their degree program. Twenty different programs were listed. For analysis, program was collapsed into business majors (n = 159) and other majors (n = 20). There were fifteen non-responses to this question.

The novice students had completed an average of 1.71 quarters towards their degrees and had taken M = .31 statistics courses, while advanced beginners had completed 3.93 quarters towards their degrees with an average of 2.12 statistics classes.

Participants were asked to list the nation where they received their secondary school education. In addition to the United States (n = 172), ten countries were listed. These ten countries were collapsed into one category (n = 15) representing only 8% of the sample.

Participants were asked to rate their attitude towards statistics on a five-point rating scale ranging from "very unfavorable" to "very favorable" with 1 being "very unfavorable" and 5 representing "very favorable." The mean rating for participants in both courses was M = 3.32. The same rating scale was used to assess participants' attitude toward empirically based research with INFO1010 students expressing M = 3.31 and INFO2020 students M = 3.39.

Participants were asked to rate their interest in Colorado health issues on a five-point rating scale ranging from "very uninterested" to "very interested" and categorized as 1 – 5 with 5 being "very interested." Means scores were 3.13 for INFO1010 and M = 3.39 for INFO2020.

Quantitative Analysis Assumptions.

A general linear model with repeated measures was run to understand the effects of treatment (graphs or tables), statistics training level (beginner or novice), time (R1 or

R2), and slide complexity level (easy, moderate, difficult) on score. Statistical significance was evaluated at $\alpha = .05$ with Bonferroni correction applied. There was one extreme value in the scores as defined by SPSS Version 24 which was three box lengths outside the box in a box and whiskers plot. That record was removed from analysis. Effects of univariate outliers were analyzed at $Z > |2|$ ($n = 143$) and $Z > |3|$, ($n = 191$). Results of the $Z > |3|$ model were nearly identical to results with the full data set. Result of the $Z > |2|$ model were similar to results from the full data set. In the full data set, no record had a leverage value greater than the critical value of .0368 (Karadimitriou & Marshall, 2015) indicating that no individual record exerted undue influence on the model. Subsequently, only results based on the complete data set are reported here. The interested reader may contact the author for more detail.

Normality was established for all six dependent variables, that is easy, moderate, and difficult scores for each round, with skewness values < |1|. There was homogeneity of variance for R1 easy ($p = .458$), moderate ($p = .490$), and difficult scores ($p =.264$), and R2 easy ($p =.513$), moderate ($p = .063$), and difficult scores ($p = .439$), as assessed by Levene's test for equality of variances.

Sphericity could be assumed (Greenhouse-Geisser Epsilon >.70) for all repeated factors and interactions of repeated factors.

Effects.

Experimental results are presented in this section. The assessment was comprised of twenty-one questions, with five questions pertaining to the easy slides, nine for the moderate slides, and seven for the difficult slides. Correct answers earned one point, incorrect answers earned zero points. Because of the unequal number of questions at each difficulty level, scores for each level are the percent of total questions at that level that the participant answered correctly. See Table 5 for means and standard deviations of scores for each round and difficulty level.

Table 5
Means and Standard Deviations for Dependent Variables

Round & Complexity	Treatment	Training	*M*	*SD*	*N*
R1 Easy Slides	Tables	Novice	43.33	16.67	48
		Adv. Beginner	46.67	15.91	60
		Total	45.19	16.26	108
	Graphs	Novice	54.69	21.90	49
		Adv. Beginner	54.05	22.04	37
		Total	54.42	21.83	86
	Total	Novice	49.07	20.21	97
		Adv. Beginner	49.48	18.73	97
		Total	49.28	19.43	194
R1 Moderate Slides	Tables	Novice	57.18	18.48	48
		Adv. Beginner	60.56	19.56	60
		Total	59.05	19.07	108
	Graphs	Novice	60.54	20.42	49
		Adv. Beginner	58.56	15.42	37
		Total	59.69	18.36	86
	Total	Novice	58.88	19.46	97
		Adv. Beginner	59.79	18.03	97
		Total	59.34	18.71	194
R1 Difficult Slides	Tables	Novice	40.48	15.41	48
		Adv. Beginner	41.19	18.43	60
		Total	40.87	17.08	108
	Graphs	Novice	44.90	16.24	49
		Adv. Beginner	41.70	15.93	37
		Total	43.52	16.09	86
	Total	Novice	42.71	15.90	97
		Adv. Beginner	41.38	17.43	97
		Total	42.05	16.66	194

Means and Standard Deviations for Dependent Variables

Round & Complexity	Treatment	Training	*M*	*SD*	*N*
R2 Easy Slides	Tables	Novice	36.67	18.60	48
		Adv. Beginner	38.67	17.22	60
		Total	37.78	17.79	108
	Graphs	Novice	42.04	20.10	49
		Adv. Beginner	40.00	19.44	37
		Total	41.16	19.73	86
	Total	Novice	39.38	19.46	97
		Adv. Beginner	39.18	18.01	97
		Total	39.28	18.70	194
R2 Moderate Slides	Tables	Novice	47.69	12.75	48
		Adv. Beginner	47.04	16.79	60
		Total	47.33	15.07	108
	Graphs	Novice	45.35	19.22	49
		Adv. Beginner	49.25	14.47	37
		Total	47.03	17.35	86
	Total	Novice	46.51	16.30	97
		Adv. Beginner	47.88	15.91	97
		Total	47.19	16.08	194
R2 Difficult Slides	Tables	Novice	37.74	17.05	48
		Adv. Beginner	27.62	16.61	60
		Total	32.12	17.47	108
	Graphs	Novice	37.32	17.93	49
		Adv. Beginner	35.14	16.36	37
		Total	36.38	17.21	86
	Total	Novice	37.53	17.41	97
		Adv. Beginner	30.49	16.83	97
		Total	34.01	17.44	194

NOTE: All scores are reported as percent of total questions that are correct at each level of slide difficulty. R1 = immediate posttest, R2 = follow-up test at two weeks.

Level of training was not statistically significant, $F(1,190) = .161$, $p = .689$, H_{01} was retained. The main effects of treatment, $F(1,190) = 4.29$, $p = .040$, partial $\eta^2 = 0.022$; time, $F(1,190) = 125.60$, $p < .001$, partial $\eta^2 = 0.40$; and complexity, $F(2, 380) = 71.42$, p

$< .001$, partial $\eta^2 = 0.27$, were statistically significant. H_{02}, H_{03} and H_{04} were rejected. The graphs treatment mean was 3.2% higher than the tables treatment mean and R2 score mean was 9.9% lower than R1. Scores for the moderate slide questions were higher than either easy slide questions (mean difference = 8.8%) and for difficult slide questions (mean difference = 15.0%). See Figure 5 for a display of the main effect means.

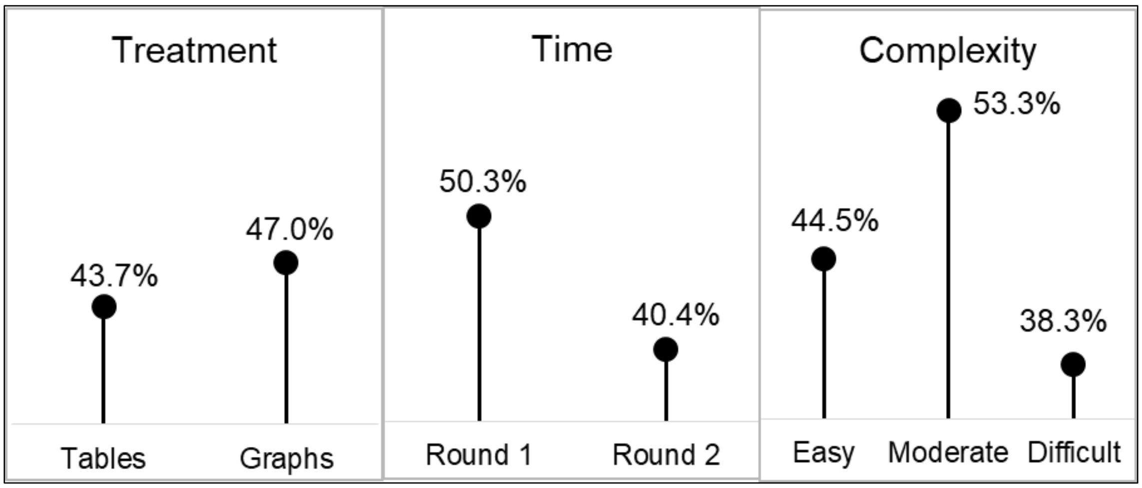


Figure 5. Statistically significant main effects.

Table for complete details of the repeated measures ANOVA results. There were no statistically significant two-way interactions. H_{05}, H_{06}, H_{07}, H_{08}, H_{09}, and H_{010} were retained. The interaction between Time, Treatment, and Training, $F(1, 190) = 4.15$, $p = .043$, partial $\eta^2 = 0.02$ was statistically significant. $H_{011}$was rejected. The simple effects of this interaction were investigated and are reported below. There were no other statistically significant three-way interactions. H_{012}, H_{013}, and H_{014} were retained. The four-way interaction between Time, Complexity, Treatment, and Training was not statistically significant. H_{015} was retained.

Table 6

ANOVA Summary Table for the Effects of Treatment, Time, Complexity, and Training

Source	*df*	*F*	*p*	Partial η^2
Treatment	**1**	**4.29**	**0.040**	**0.02**
Training	1	0.16	0.689	0.00
Treatment * Training	1	0.07	0.797	0.00
Error	190			
Time	**1**	**125.60**	**<0.001**	**0.40**
Time * Treatment	1	1.14	0.287	0.01
Time * Training	1	1.01	0.316	0.01
Time * Treatment * Training	**1**	**4.15**	**0.043**	**0.02**
Error (Time)	190			
Complexity	**2**	**71.42**	**<0.001**	**0.27**
Complexity * Treatment	2	2.89	0.057	0.01
Complexity * Training	2	2.24	0.108	0.01
Complexity * Treatment * Training	2	0.72	0.487	0.00
Error (Complexity)	380			
Time * Complexity	2	2.27	0.105	0.01
Time * Complexity * Treatment	2	1.66	0.192	0.01
Time * Complexity * Training	2	1.05	0.350	0.01
Time * Complexity * Treatment * Training	2	1.24	0.289	0.01
Error (Time*Complexity)	380			

The simple effects of time, treatment, and training were evaluated using independent-samples t-tests because of their statistically significant interaction. The data set was split by level of training. Statistical significance was evaluated at $\alpha = .05$. See Figure 6 for a graphical representation of the interaction.

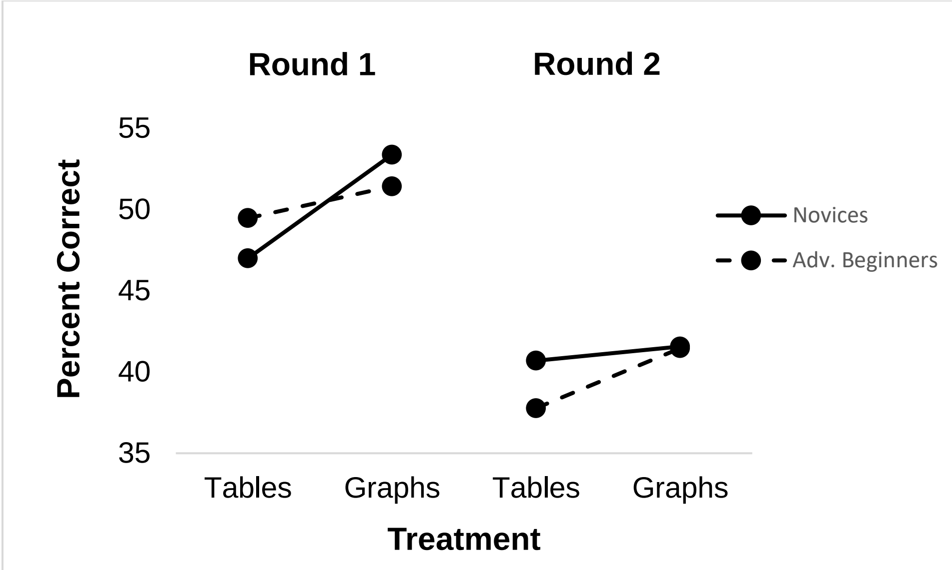


Figure 6. Interaction of time, treatment, and training.

Round 1 (R1) easy item scores for novices were statistically significantly higher for the graphs treatment (*M* = 54.7%, *SD* = 21.9%) than for the tables treatment (*M* = 43.3%, *SD* = 16.6%), $t(95) = -2.87$, $p = 0.005$. This is a medium effect (Cohen's $d = 0.58$) (Cohen, 1977, p. 25) indicating that the novices in the graphs treatment scored 0.58 standard deviations higher on easy items than novices in the tables treatment. In addition, Round 2 (R2) difficult item scores for advanced beginners were statistically significantly higher for the graphs treatment (*M* = 35.1%, *SD* = 16.4%) than for the tables treatment (*M* = 27.6%, *SD* 16.6%); $t(95) = -2.18$, $p = 0.032$. The effect size approaches a medium effect with $d = 0.46$ indicating that advanced beginners in the graphs treatment scored an average of 0.46 standard deviations higher on difficult items than those in the tables treatment. All other simple effects are not statistically significant. All means and standard deviations are listed in Table 5. Complete *t*-test results are listed in Table 7.

Table 7

*T-test for Equality of Means**

Training	Score	*t*	*df*	*p***	Difference *M*	Difference *SE*	95% CI of the Difference LL	95% CI of the Difference UL
Novice								
	R1 Easy	**-2.87**	**95**	**0.005**	**-0.11**	**0.04**	**-0.19**	**-0.04**
	R1 Moderate	-0.85	95	0.397	-0.03	0.04	-0.11	0.04
	R1 Difficult	-1.38	95	0.172	-0.04	0.03	-0.11	0.02
	R2 Easy	-1.37	95	0.175	-0.05	0.04	-0.13	0.02
	R2 Moderate***	0.70	95	0.484	0.02	0.03	-0.04	0.09
	R2 Difficult	0.12	95	0.906	0.00	0.04	-0.07	0.07
Adv. Beginner								
	R1 Easy	-1.91	95	0.059	-0.07	0.04	-0.15	0.00
	R1 Moderate	0.53	95	0.599	0.02	0.04	-0.06	0.10
	R1 Difficult	-0.14	95	0.890	-0.01	0.04	-0.08	0.07
	R2 Easy	-0.35	95	0.725	-0.01	0.04	-0.09	0.06
	R2 Moderate	-0.66	95	0.509	-0.02	0.03	-0.09	0.04
	R2 Difficult	**-2.18**	**95**	**0.032**	**-0.08**	**0.03**	**-0.14**	**-0.01**

*Equality of Variance assumed

**Two-Tailed

***Equality of Variance not assumed (Levene's Test, $F = 8.14$, $p = .005$).

Analysis by demographic characteristics.

To determine if scores could be predicted by demographic variables, multiple regression was performed with the demographic variables sex, business major, number of quarters towards degree, number of statistics classes, research attitude, Colorado resident, issue interest, number of PPTs viewed in last month, number of PPTs created, and the group variables of treatment and training. US secondary education was not included in the analysis because only 8% of participants reported receiving their secondary school education outside the United States. Rummel (1970, as cited in Tabachnick & Fidell, 2013) recommends removing dichotomous variables with extreme splits because scores for “the small category are more influential” (p. 73) than those in the large category. Age was not included because of high skewness. At $\alpha = .05$, the only statistically significant model with significant predictors was R1 for percent-correct scores for easy items $F(12, 157) = 4.30$, $p < .001$, $R^2 = .248$. Relationships between statistically significant variables were investigated using partial and semi-partial correlations. All correlations were low indicating that any spurious relationships between independent variables are negligible. No other regression models in R1 were statistically significant. In R2, the model for percent-correct scores for difficult slides was statistically significant $F(12, 157) = 2.07$, $p = .022$, $R^2 = .137$. However, there were no significant predictors in this model. No other regression models in R2 were statistically significant.

Significant predictors for the R1 easy items model are displayed in. This model explains 24.8% of the variation in scores for the first round easy items. Participants with more favorable attitudes towards empirical research, who have viewed higher numbers of PPT presentations, or who are Colorado residents were predicted to score higher compared to their peers. More favorable attitudes towards empirical research may indicate a greater willingness to attend to a research presentation based in quantitative analysis. Individuals who have seen greater numbers of PPTs may be more familiar with the treatment delivery which could lead to higher scores by reducing extraneous load. Scores for participants who are Colorado residents are significantly correlated with scores for easy items ($r = .19$, $p = .009$, $n = 191$) probably due to greater familiarity with place names and geography of the state. In this model the experimental treatment of graphs predicts a higher score by 10.59% compared to the tables treatment suggesting that graphical presentation leads to greater understanding of easy complexity research results.

Table 8

Statistically Significant Regression Coefficients for R1 Easy Item Model

	Unstandardized Coefficient		Standardized Coefficient		
Model & Predictors	*B*	*SE*	*β*	*t*	*p*
R1 Easy					
Treatment	10.59	2.78	0.28	3.82	<0.001
Attitude Toward Empirical Research[a]	4.63	1.65	0.24	4.11	0.006
Colorado Resident[b]	11.67	2.84	0.30	4.11	<0.001
Number of PPTs Viewed in Past Month	3.85	1.67	0.17	2.31	0.022

[a] Scale: 5 (Very interested) to 1 (Very uninterested)

[b] Binary: 1 = yes, 0 = no

Qualitative Results

Description of participants.

There were six participants for the think aloud protocol analysis randomly assigned to either the tables treatment or the graphs treatment, with three participants assigned to each condition. Participants who saw the tables treatment are referred to as P1 through P3 and those who saw the graphs treatment are referred to as P4 through P6. All participants were advanced doctoral students or had completed their programs. Complete demographic information is found in Table 9.

Table 9
Characteristics of Think Aloud Participants

Characteristic	Category	Count	*M*	*SD*
Sex	Female	6		
Status	Graduate Student	4		
	PhD	2		
Colorado Resident	Yes	5		
	No	1		
Age			35.5	5.32
Number of PPTs Viewed in Last Month			1.5	0.84
Number of PPTs Created			2.5	0.84
Number of Quarters Toward RMS PhD				
In progress, quarters completed		2	7.5	2.12
Completed PhD or ABD		4		
Number of Statistics Classes Taken			10.67	5.16
Attitude Towards Statistics[a]			4.50	1.22
Attitude Toward Empirical Research[a]			4.83	0.41
Issue Interest[a]			4	0.63

[a]Scale 1 (low) to 5 (high)

Protocol Analysis.

In general, advanced statisticians followed the same protocol when interpreting research results regardless of whether the results were presented in tabular form or in graphical form. As stated previously, the title of every slide signaled the important information or conclusion that viewers should gather from that slide. In this analysis, participants read the title of the slide 63 out of 72 times or 87.5% of the time. Three participants made no statement about the title slide (the first slide). Disregarding the title slide, participants read the title of the other slides 91% of the time. After reading the slide title, the statisticians proceeded to process or verify the statements contained within the title.

When a participant did not fully understand the statement expressed in the title she searched the slide content for information to facilitate interpretation. The code "process" was used when there was some indication that the viewer was unsure of the meaning expressed by the title and used supporting information in the slide to infer meaning. For example: "let's see what else" (P2), "ah how do … are we comparing the mean with the, ah, membership, okay" (P3), "Trying to understand what the numbers in the blue circles mean" (P4), "Okay, what does that, what does that mean? Is that a large amount a small amount? Yeah" (P1).

If the participant understood the slide title, the process stage was skipped and the search focused on finding evidence to support the conclusion of the title. The code "interpret" was used when the verbalization indicated that the viewer understood the slide title and was looking to affirm its veracity. For example: "so I'm seeing that Colorado is lower on average than the US as far as premature death" (P3), "Wow that's very distinct. This is all over the place" (P5), "okay as one rises one goes down okay, county health outcome rank falls, okay, so health okay health outcome rank" (P2).

Participants who watched the tables presentation made 37 process expressions and 20 interpret expressions. Those who watched the graphs presentation made 19 process expressions and 30 interpret expressions. A chi-square test of independence was conducted between treatment and process/interpret for this stage of the protocol analysis. All expected cell frequencies were greater than five. There was a statistically significant association between treatment and process/interpret expressions, $\chi^2(1) = 7.22$, $p = .007$

indicating that participants viewing graphs moved directly to the interpretation stage more readily than those viewing the tables.

In the presentations, slide content became increasingly complex, either in terms of the number of data points being compared or in the statistical analysis being presented or both. As slide content became more complex, both groups relied on pre-existing knowledge or *schema* to validate the slide. For example, participants looked for evidence of statistical significance when presented with correlations. At one point, P3 said, "so the significant correlations are existing yeah," and P4 "correlation between health outcomes are significant." When looking at slide 6 which presented regression results, P2 wondered "What are those three health outcomes rank ANOVA predictors okay," P3 noted "the R-square is .75" and P1 observed "this sample size doesn't look very big."

People who saw the graphs presentation applied a schema 17 times compared to 28 times for the people who saw the tables. A one-tailed t-test was performed to determine if the tables participants applied schema more frequently than graphs participants. A statistically significant difference (M = -3.67, 95% CI [-8.39, 1.05], $t(4)$ = -2.16, p = .048) indicates tables participants applied schema (M = 9.33, SD = 2.52) more frequently than graphs participants (M = 5.67, SD = 1.53). There was homogeneity of variances, assessed by Levene's test for equality of variances (p = .47).

In addition, participants who saw graphical comparisons on the easier slides, which tended to compare means, were more likely to accept the conclusions stated in the titles whereas participants who saw tables were more likely to question the metric used in the comparison. For example P3 said about slide 5 “So we don’t know if those unhealthy days are above or below the average.”

After processing and/or interpreting the slide content with or without the application of schema from long-term memory, participants moved on to understanding (or not). Examples of phrases indicating understanding include: “so that’s pretty clear” (P6), “so we got three factors, three significant factors in this model,” (P3), “Heurfano is pretty low” (P4), “So those three things can determine your health rank. Okay.” (P1). In addition, the graphs participants were more likely to say the slide “makes sense” with ten uses of the phase compared to six for the tables group. A one-tail independent sample t-tests was performed to see if these differences are statistically significant. Results are not significant.

Sometimes the understanding phase of the protocol included an application of the content to the participant’s own life. In particular, four out of six participants identified in some way with the slide that showed counties (slide 6) clustered by outcome. P6 said, “I live in one of the better outcome counties so I’m happy about that. Ah, that was the first thing I looked at is where is my county and what color is it.” Participants also made content personal, for example, concluding “so living in Colorado might not be a bad idea alright” (P5) upon learning Colorado has a lower premature death rate than the country as a whole.

While the protocol used by participants in processing the material of the presentations is essentially the same, differences in responses exist between the two treatments. One difference is that those who saw the graphs presentation were more likely to say that the content was “interesting” using the word eighteen times compared to only five times for the tables group. An independent-samples t-test indicates the difference between groups was non-significant.

The graphs group uttered more positive critiques as well, with ten positive critiques compared to zero for the tables group. Examples of positive critiques include: “So I can see the Colorado flag clearly indicates hundred” (P3, slide 3), and “the color coding is very clear” (P6, slide 4). On the other hand, the tables group expressed more negative criticism of the slides. They made twelve negative critiques compared to five from the graphs group. “This is too much to look at” (P1, slide 6) and “it’s kind of throwing [in] a lot of different things” (P2, slide 7). A chi-square test of independence was conducted between treatment and critique-positive or -negative for this stage of the protocol analysis. All expected cell frequencies were greater than five except the expected count for positive critique from the tables group which was 4.4. There was a statistically significant association between treatment and positive or negative critiques, $\chi^2(1) = 12.71$, $p < .001$ indicating that participants viewing graphs uttered more positive critiques than those viewing the tables. Results for each slide can be found in Appendix K.

Slide Processing Protocol.

Participants followed the same general series of steps to understand slide content regardless of presentation. First they read the slide title, then in a somewhat iterative manner they processed and/or interpreted the slide title using content in the slide display to either inform or confirm their understanding of the title. Once they were sure of their interpretation of the slide, they signaled understanding at which point we can assume the material entered short-term memory. At this step they sometimes applied the slide content to their personal experience. See Figure 7 for a summary of table/graph viewing processes.

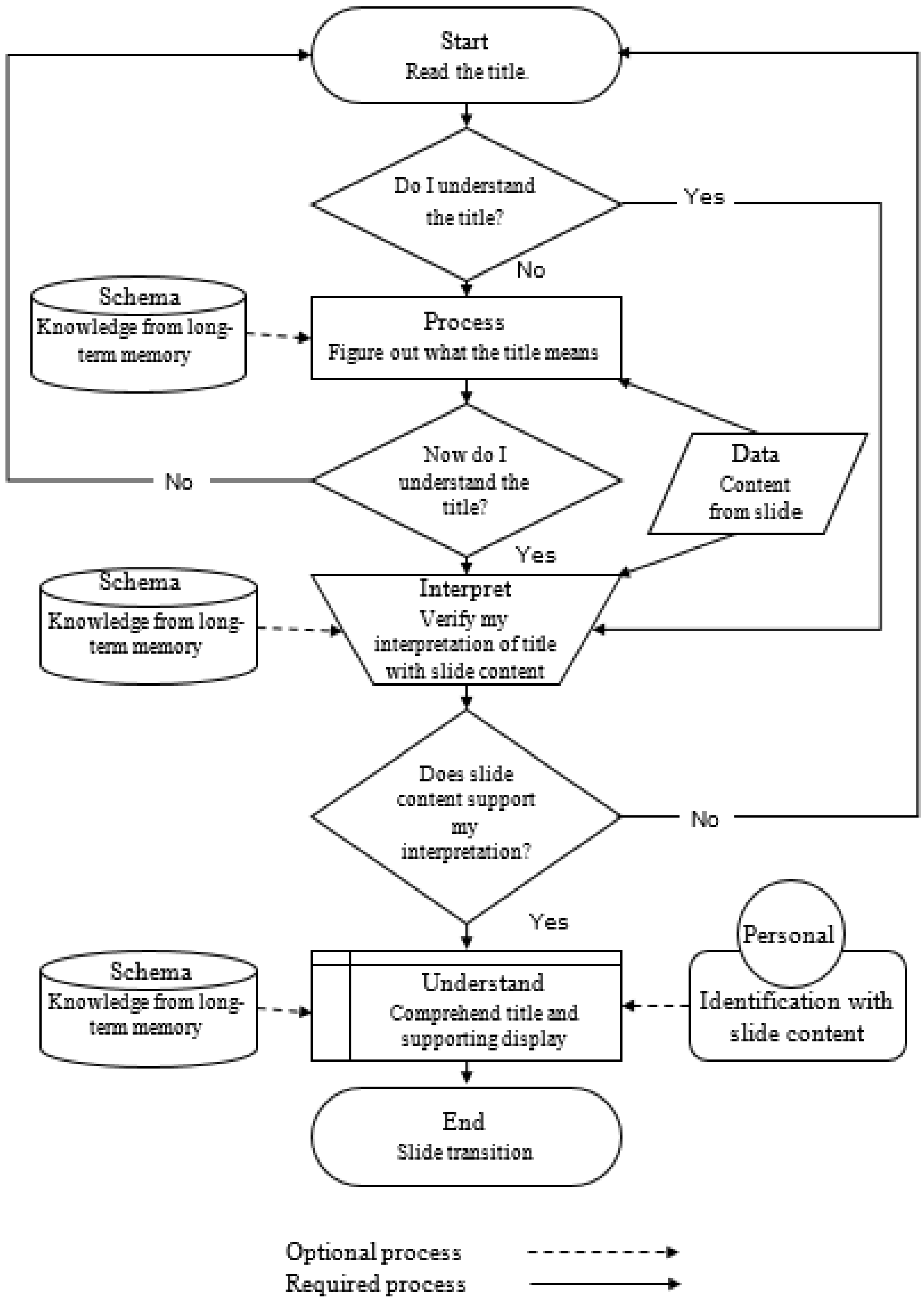


Figure 7. Flow chart of slide processing protocol.

Chapter Four: Discussion

Introduction

Knowing the best ways to reach one's audience and maximize what is remembered is important to presenters in all fields. Presentations of research results are founded in quantitative analysis in an environment where attitudes that foster math anxiety in American culture are pervasive (Ashcraft, 2002). For researchers, this makes reaching a broad audience with a presentation based on statistical inquiry a challenge.

This chapter summarizes the findings of this study which explored the efficacy of presentations using graphic or tabular displays. The quantitative phase, a quasi-experiment, and the qualitative phase, a think aloud protocol analysis, were conducted in parallel and are reported separately. This section merges the two phases and integrates the results of this study with the literature. It addresses the limitations of the study; provides recommendations for researchers, statisticians, presenters, and educators; and offers ideas for future study.

The overarching question this study addressed was: are graphs or tables more effective for researchers, evaluators, educators, and others to present research results grounded in complex statistical analysis so that results are readily understood and remembered by audiences with varied levels statistical training? To that end, it explored the relative efficacy of presentations relying on graphs compared to those relying on tables as the primary means to communicate research results and identified a protocol used by advanced researchers to interpret slides presenting results in either format.

The design of the study addresses gaps in the scholarship relating to research presentation format by qualitatively exploring the process by which individuals with advanced statistical training interpret slides presenting research results, while quantitatively exploring the effects of slide presentations with graphs or tables on understanding and recall for slides of varying levels of difficulty for viewers with no, or limited, statistical training.

The research questions that were addressed quantitatively are: Are there effects of statistical training, degree of complexity of presented material, and presentation format on understanding and recall of information? Are there interactions between level of statistical training, mode of presentation, and level of complexity, on understanding and recall of information provided in presentations of research results?

The qualitative portion addressed this research question: How do people with advanced statistical training interpret graphical and tabular presentations of research results?

Major findings

As presented in Chapter Three, the statistically significant main effects for time and complexity had large effect sizes (Haase, Waechter, & Solomon, 1982). There was a main effect for treatment with a small effect size with the graph treatment leading to superior understanding and recall for both novices and advanced beginners. The large main effect for time comes as no surprise. The passage of time explains 40% of the variance in scores. Participants understood more of the material immediately after the presentation than they recalled two weeks later.

There was a large main effect for complexity which explains 27% of the variance in scores. Interestingly, scores for moderate slides were higher than for easy slides and scores for easy slides were higher than for difficult slides. This might be because higher scores for easy slides were statistically significantly correlated with Colorado residency ($r = .19$, $n = 191$). Two out of five easy-slide questions depended on having some sense of the geography of Colorado (graphs treatment) or knowledge of names and locations of Colorado's sixty-four counties (tables treatment). Future research might clarify this difference.

Overall, these results indicate that there was a meaningful, though small, effect of using graphs to present research results on understanding and recall. Level of complexity and the passage of time also impacted results. Scores went down after two weeks compared to the immediate post-test. Scores were lowest for difficult items and highest for items of moderate difficulty. In addition to these main effects, specific research questions were answered with interactions. A discussion of specific questions follows.

Major findings by research question

Impact of graphs on understanding and recall of research results for people with some statistical training.

Advanced beginners who saw the graphs presentation earned scores that were, on average, 7.52% higher for the difficult slides in Round 2 indicating they recalled more of the presentation after a period of two weeks. Regression results did not reveal any significant predictor variables. While results for easy and moderate slides in both rounds and for difficult slides in Round 1 were not statistically significant, scores for the graphs treatment were higher for all but moderate slides in Round 1. The advanced beginners had completed the equivalent of an introductory statistics course. They were selected as a convenient group thought to have a similar level of statistics training as beginning researchers and Master's-degree holding professionals who might attend a research conference. Graphs helped them remember more content with a statistically significant impact for the more difficult slides.

Impact of graphs on understanding and recall of research results for people with no statistical training.

For novice participants, the graphs treatment yielded higher scores by an average difference of 11.36% for easy slides compared to the tables treatment on the first round scores, indicating that graphs lead to greater understanding for people with no statistical training for easy content. No statistically significant differences were found for moderate or difficult slides in Round 1 or for any complexity level in Round 2, however, mean scores were higher for the graphs treatment except for moderate item complexity in the second round.

Impact of slide complexity on understanding and recall using graphs and tables.

Slide complexity, easy, moderate, and difficult, was found to have a large effect on understanding and recall explaining 26.7% of the variance in scores ($\eta^2 = .267$) (Levine & Hullett, 2002). Scores for moderate items were higher than for easier items and scores for easier items were higher than for difficult items. Perhaps this indicates that items of moderate difficulty pose an optimal challenge level for learning (Guadagnoli & Lee, 2004) in presentations of research results to inexperienced researchers.

Protocol used by statisticians to interpret presentations of research results

A protocol was established. Regardless of presentation type, experts read the slide title. If they did not understand the title completely, they used slide content to process the title, returning to the title for clarification. If they did understand the title, they then used slide content to verify and interpret the title. Finally, viewers moved on to understanding the display unless the slide changed before reaching understanding. Viewers pulled from existing schema at any step of the process.

Viewers of graphs were more likely to skip the processing stage and move directly to interpreting the slide. They were less likely to pull schema from long-term memory in their efforts to interpret slides. They were more likely to find the content interesting and the display pleasing.

Effects of demographic variables.

Some demographic characteristics of the experimental participants had statistically significant effects on scores. In particular, in Round 1 for easy content, a more favorable opinion toward empirical research predicted higher scores, and the greater number of PPTs viewed in the month prior to the experiment predicted higher scores. Also, as previously mentioned, Colorado residents earned higher scores compared to non-residents. While this research does not attempt to explain these differences, it is important to note that presenters have no control over individual audience member characteristics, reiterating the necessity of striving to reach a broad audience possessing various levels of skills and understanding.

Conclusions

Significance and implications of the study

The experimental portion of this multiple-methods study was designed to mimic presentations of research results as typically seen at professional conferences. Both the tables presentation and the graphs presentation were built using strategies and techniques that support cognition as presented in Appendixes A, E, and F. In particular, both presentations included easy-to-read, adequately sized, *san-serif* fonts in high-contrast colors. Neither presentation included extraneous text or graphical elements. Both presentations used titles as headlines to inform the audience of information, results, or conclusions presented on each slide. In both treatments, viewers' attention was guided with top to bottom, left to right structuring of elements taking advantage of spatial proximity (Few, 2006). Consistent use of these design elements across experimental treatments may be one reason for the limited number of significant interactions between treatment, difficulty level, training, and time as presented in Chapter 3.

On the other hand, significant results for the graphs treatment indicate that individuals with no statistical training (novices) understood more easy content than with the tables treatment. This is consistent with Dunlap and Lowenthal (2016) who found effective visuals help people understand unfamiliar concepts when they do not possess pre-existing mental models supporting comprehension of the material.

Significant results also indicate that those with some statistical training (advanced beginners) recalled more of the difficult content two weeks after the presentations. Furthermore, experts frequently and significantly skipped the processing stage of slide interpretation when viewing graphs, arriving at the interpretation stage more rapidly. These results indicate that researchers should capitalize on the pictorial superiority effect (Stenberg, 2007) using eye-catching visual elements in graphs such as color, alignment, and size (Evergreen S. D., 2018) to tap into audiences' preattentive mental processing stage (Ware, 2013). By visually organizing chunks of information (Few, 2006) and emphasizing important features of the display, graphs can reduce the viewer's cognitive load (Shah, Mayer, & Hegarty, 1999) thus allowing more information into working memory at a time, easing interpretation and facilitating understanding (Stenberg, 2007). While these results do not indicate exclusive use of graphs over tables in presenting research results, they do suggest that graphs provide a slight advantage over tabular displays in facilitating understanding and recall.

In the experimental phase, graphs led to greater understanding of easy content among novices and to greater recall of difficult content among advanced beginners. Understanding and recall for participants viewing the tables presentation were never statistically significantly better compared to those viewing graphs. In the protocol analysis phase, both groups of participants increasingly called on pre-existing schema as slides became more complex, however, those who saw the graphs relied significantly less on information held in long-term memory than those who saw the tables. Perhaps by showing results graphically, participants had more confidence in the statistics, as suggested by Pandey, Manivannan, Nov, Satterthwaite, and Bertini (2014). These results also support the pictorial superiority effect (Stenberg, 2007), the idea that humans are wired to rely on vision as the dominant sense, since participants in both treatments heard the same audio. Think aloud participants echoed the conclusions of copious research about the visual appeal of graphs (see, for example, Ashton, 2013; Nussbaumer, 2011; Sweller, 1994) with more positive critiques of the graphs than of tables. Researchers and other presenters can paint a rich story of cause and effect or other numerical relationships (Sweller, 1994).

This study shows that advanced statisticians relied on the message of the slide title as one relies on the headline of a news article. In the same way that a headline should accurately reflect the contents of the article (Mann M. R., 2001), the slide title should clearly and concisely inform the viewer of the main message of the slide (Evergreen S. D., 2017). If the point of the slide is evident, advanced researchers use either graphs or tables to assess its merit. However, the researchers who saw the graphs frequently and statistically significantly skipped the process stage, arriving at the interpret stage immediately or more rapidly, and expressed more favorable impressions of the slides with graphs. The novice and advanced beginners may have relied on the slide title in the same way.

Practical Applications

This study demonstrates that the use of graphs to present research results leads to greater understanding and recall when compared to the use of tables. It also demonstrates that viewers read the title, then proceed to use slide content combined with information held in long-term memory to interpret and/or verify the title. Expert researchers conveyed more favorable impressions of the graphs treatment as compared to the tables treatment.

It is common advice for speakers to consider their audience when preparing a presentation. In this study, there were no statistically significant differences in scores for either understanding or recall between novice and advanced beginner researchers even though the advanced beginners had training equivalent to master’s degree-level practitioners who might attend a research conference. This result suggests that research conferences generally should consider their audience to have knowledge and experience of empirical research reporting consistent with that of lay people.

Furthermore, slides with more than three to five chunks (Cowan, 2000) of information might be too complicated to be understood in the time frame of a research presentation. Think aloud participants who saw the graphs presentation all expressed confusion about slide 7 (see Appendix K for details), a difficult slide which presented twenty different data points. On the other hand, tables participants used the same data points, which were shown in the table, to confirm slide conclusions without expressing any confusion. This suggests that if complex data cannot be broken up into simpler chunks for graphic representation, consistently formatted tables might beneficially be used to emphasize the most important numbers (Few, 2004).

Canonical forms of graphs (Tukey, 1977) that are accurately (Cleveland & McGill, 1984) and rapidly (Ratwani & Trafton, 2008) interpreted suggest reliance on dot plots and bar graphs when comparisons of discrete values are to be made. However, by following perception-based guidelines (Ware, 2013) presenters can use graphic elements to show other important and informative patterns in the data (Hegarty, Canham, & Fabrikant, 2010). The graph slide 9 (see Appendix B for slide images), in the experiment, uses arrows and lines to illustrate an inverse relationship between independent and dependent variables. All three of the think aloud participants who saw the graphs presentation said this slide was “interesting” while none of the tables participants made a similar claim (see Appendix K for slide-by-slide analysis).

For presenters, a thorough understanding of good design principles will help maximize viewers’ understanding and recall by focusing viewers’ attention on salient points. Existing research tells us what graphs to use for which data to facilitate audience understanding and what design techniques work to draw the viewer’s attention to the intended message (Cleveland, 1984, 1994; Cleveland & McGill, 1984, 1985, 1987; Evergreen, 2017; Few, 2004). This study extends what is known to research presentations. It indicates that presenters might increase understanding and recall of their presentations with the thoughtful use of graphs to display their results. Regardless of presentation style, this research introduces the importance of the slide title in communicating the meaning of each slide in a presentation.

Limitations

A limitation of this study is the choice of Business Analytics undergraduate students as a proxy for a research presentation audience. Since the sequence of courses is required of all business majors, these students might not have been motivated to engage with the presentations in the same way as those willingly attending a research presentation, thereby potentially affecting their understanding and recall of the content

(Pandey, Manivannan, Nov, Satterthwaite, & Bertini, 2014). Also, some participants may have had misgivings about their own math abilities (personal communications, various dates) which may have impacted their engagement with the presentations. Ashcraft claims, "Math anxiety disrupts cognitive processing by compromising ongoing activity in working memory" (2002). Math anxious students may not adequately substitute for conference participants. Results might have been more pronounced with undergraduates in a more research-focused field such as psychology or social work.

Another limitation of the experiment is the low reliability score of the assessment instrument which indicates a lack of precision in the estimate of participants' scores. The estimated attention span of early career college students and reluctance to inconvenience professors who were granting access to their classes led to the creation of a short instrument. More questions would have led to greater reliability (Rosenthal, 1994). Improved reliability of the assessment instrument, would lead to more trustworthy indications of the effect of treatment on understanding and recall.

Low internal consistency can also mean that the assessment measured multiple dimensions in addition to measuring experimental effects. For example, the significant correlation between Colorado residency and correct answers on geography-related questions ($r = .19$, $n = 191$) indicates that certain questions measured, at least in part, familiarity with Colorado geography and place names, not understanding and recall of the presentation content. Another example of a possible additional dimension is familiarity with canonical graph forms. A longer assessment would allow exploration of effects along different dimensions.

An assumption inherent in the design of the graphs presentation was that all participants had knowledge of canonical graph forms and, therefore, had pre-existing skills to interpret standard aspects of graphs such as the relationship of the x- and y-axes or the relative position of markers in a dot or bar graph. This assumption depends on what they learned in their secondary schools, especially for the novice participants most of whom were in their first term of college. Carpenter and Shah determined that viewers who lacked the knowledge to connect a graphic form to its underlying data took longer to interpret the graph than those who could use the form as a chunk of information (1998). If participants in this experiment did not have preexisting schema with which to interpret the graphs, the hypothesized advantage of the graphs treatment would not be expected to materialize since individual differences in graphic knowledge impact the interpretive process (Shah & Carpenter, 1995; Carpenter & Shah, 1998). This might explain the lack of statistical significance for both understanding and recall with moderate slides and for understanding for difficult slides since both levels present more complex statistical analyses than the easy slides.

Future research

As with most research, this study raised as many questions as it provided answers. At its heart, this was a study about teaching in that the goal of every research presentation is to teach the audience something about one's findings. Improved instrument reliability with better performing questions for each slide would help to distinguish which types of graphs are superior to tables for which data and detail the specific displays that work best for different types of data or analyses.

The result in this experiment of higher average scores for moderate difficulty slide content, regardless of presentation or level of training, compared to either easy and difficult slides, hints at intriguing aspects of optimal challenge (Guadagnoli & Lee, 2004) for viewers and presenters. According to Cowan (2000) people can only process three to five chunks of information in short-term memory at one time but research is not "very definite about what constitutes a chunk of information" (Miller, 1956, p. 93) especially considering individuals combine information held in long-term memory with new information to create larger chunks out of smaller ones (Cowan, 2000; Miller, 1956). Future research could explore what constitutes a chunk of information and its relationship to difficulty level of the presented material.

While the graphs treatment was hypothesized to accelerate cognition by chunking the information presented into visual units and every attempt was made to ensure the two presentations were informationally equivalent, it seems clear that tables allow a more dense presentation of highly detailed information. Future research could explore what informational aspects are gained or lost with both treatments and their effect on cognition.

This experiment explored the effects of presentation style on understanding and recall among viewers with limited statistical training. The results were enhanced by using experts for the protocol analysis phase. No attempt was made to explore presentation style on understanding and recall among expert statisticians. Future research could examine the impact of presentation style on understanding and recall of more statistically knowledgeable audiences. Since the experimental participants were selected as a proxy for a research conference audience, the effects of motivation and attention could be better evaluated by using participants at an actual research conference such as the annual conference held by AERA. Another option would be to run the experiment and the protocol analysis with the same participants or participants with the same level of statistical training. Protocol analysis participants could also pause the slide show to offer reflections before advancing to the next slide. This type of research might elucidate which elements of tables or graphs presentations people find most useful to interpreting the content.

Presentation strategies are another area ripe for research. As a teacher, I strive to structure every lesson around one important element which becomes a stated learning goal. In the context of a research presentation, evaluation, or business application, this goal could be called a take away or an actionable item. Would understanding and recall be greater if specific conclusions or actionable items were presented? Presumably providing handouts would improve understanding and/or recall. Do handouts affect understanding and/or recall at different rates depending on if the presentation content is tables or graphs? In addition, the hypotheses that were explored here could be explored in the context of the statistics classroom.

Future research could illuminate procedural aspects of using the think aloud protocol analysis for audio input as well. In this study, advanced researchers viewed the slide shows without the supporting voice recordings because of this researcher's assumption that participants could not successfully articulate their thoughts while listening to the audio. Furthermore, this researcher anticipated difficulty separating two different audio streams in the transcription phase of the analysis. Future research could explore the efficacy of having participants listen to recorded scripts on headphones while verbalizing their thoughts to learn to further develop the think aloud method and perhaps, determine if the protocol differed in the presence of audio?

In hindsight, there was no qualitative exploration of the impressions made by the two different treatments or the difficulty of the content or other aspects of the presentations on the experimental participants. Adding a simple open-ended question to the assessment in a forthcoming experiment could provide important insight into the thought processes of lay people exposed to research presentations. Using subjects without research training in a focus group exploration of graphical and tabular presentations or a think aloud protocol analysis would illuminate additional considerations for presenters to more effectively reach the consumers of their research.

An unanticipated result of the protocol analysis was learning the importance to the trained researcher of the slide title in the process of interpreting the meaning of either graphs or the tables. All of the slides in the presentation used strong titles to signal the meaning of content. Open questions about titles are: are titles equally important to all audiences in interpreting graph and tables, and are titles equally important in presentations with accompanying audio descriptions?

In short, this study has focused and corroborated the need to explore data visualization in presenting research results further. Is a picture worth a thousand numbers? The answer appears to be sometimes yes and sometimes no. Further inquiry into type, complexity, and quantity of data presented; how specific graphic displays relate to specific types of data; and presentation strategies will help determine the answer.

Concluding Remarks

As an undergraduate, I studied visual arts, math, and economics. Among many other lessons, I learned to paint, produce sculpture, and make prints using time-honored techniques. Popular wisdom would call that right-brained thinking. Many would call (and many have called) advanced study in research methods and statistics left-brain thinking, implying an undeniable and insurmountable dichotomy. This project completes a circle for me by demonstrating that the right supports the left, holistic balances analytic, seeing enhances thinking, and art buoys science. I have garnered a deeper, more holistic understanding of the application of visual processing theory, learned best-practices for displaying data and developed ideas for effectively presenting research results.

The lessons continue. With this project, the methodological learning has been immense. From the quasi-experimental set-up to the creation of the two treatments, from qualitative analysis with precious few guidelines for how to develop a protocol to quantitative analysis with some aspects limited to experience doing homework, lessons were learned and skills developed.

I have grown as a researcher, data analyst and visualizer, and presenter. These skills transfer directly to my work as an analyst and a teacher. In the world of statistics and data analytics, the effort to increase others' understanding of difficult concepts is endless. To me the fascinating aspect of quantitative and qualitative research, statistical analysis, and in-depth study is the hidden nuggets of truth they reveal. It is my hope that for myself and other researchers the results of this project enable the illumination of those truths for all viewers.

References

2017 AERA Annual Meeting Online Program Portal. (2017). Retrieved April 19, 2017, from AERA: https://convention2.allacademic.com/one/aera/aera17/index.php?click_key=2&obf_var=9967341&PHPSESSID=03uhfgh1uj2bm72k9uerm0d6c6

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2015). *Modern business statistics with Microsoft Excel* (5th ed.). Stamford, CT: Cengage Learning.

Anderson, J. R. (1987). Methodologies for studying human knowledge. *Behavioral and Brain Sciences, 10*, 467-505.

Andres, H., & Petersen, C. (2002). Presentation media, information complexity, and learning outcomes. *Journal of Educational Technology Systems, 30*(3), 225-246.

Angel, L., Bastin, C., Genon, S., Salmon, E., Fay, S., Balteau, E., . . . Collette, F. (2016). Neural correlates of successful memory retrieval in aging: Do executive functioning and task difficulty matter? *Brain Research, 1631*, 53-71. doi:http://dx.doi.org/10.1016/j.brainres.2015.10.009

Apperson, J. M., Laws, E. L., & Scepansky, J. A. (2006). The impact of presentation graphics on students' experience in the classroom. *Computers & Education, 47,* 116-126.

Arditi, A., & Cho, J. (2005). Serifs and font legibility. *Vision Research, 45*, 2926-2933.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181 - 185.

Ashton, X. (2013, June). *Thirteen reasons why your brain craves infographics.* Retrieved from NeoMam Studios: https://neomam.com/interactive/13reasons/

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience, 4*(3), 281-288.

Bartsch, R. A., & Cobern, K. M. (2003). Effectiveness of PowerPoint presentations in lectures. *Computers & Education, 41*(1), 77-86. doi:10.1016/S0360-1315(03)00027-7

Berk, R. A. (2011). Research on PowerPoint: From basic features to multimedia. *International Journal of Technology in Teaching and Learning, 7*(1), 24-35.

Betancourt Lopez, I. A. (2014). PowerPoint design based on Cognitive Load Theory and Cognitive Theory of Multimedia Learning for introduction to statistics. *Doctoral dissertation*. Retrieved from ProQuest UMI Number: 3628119.

Breslow, L. A., Trafton, J. G., & Ratwani, R. M. (2009). A perceptual process approach to selecting color scales for complex visualizations. *Journal of Experimental Psychology, 15*(1), 25-34. doi:10.1037/a001585

Brown, R. D., & Newman, D. L. (1982). An investigation of the effect of different data presentation formats and order of arguments in a simulated adversary evaluation. *Educational Evaluation and Policy Analysis, 4*(2), 197-203.

Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2016). *Essentials of business analytics* (2nd ed.). Boston, MA: Cengage Learning.

Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review, 22*(4), 1029-1034.

Cansino, P. P. (2011). Use of analysis of verbal protocols in the study of complex human behavior. *International Journal of Hispanic Psychology, 4*(2), 181-200. Retrieved from https://search-proquest-com.du.idm.oclc.org/docview/1713725178?accountid=14608

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75-100.

Cleveland, W. S. (1984). Graphical methods for data presentation: Full scale breaks, dot charts, and multi-based logging. 1-25.

Cleveland, W. S. (1993). *Visualizing data.* Summit: Hobart Press.

Cleveland, W. S. (1994). *The elements of graphing data* (Revised ed.). Summit: Hobart Press.

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

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 229*(4716), 828-833.

Cleveland, W. S., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society., 150*(3), 192-229. Retrieved November 18, 2016, from http://www.jstor.org/stable/2981473

Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 289 - 300. Retrieved from http://www.jstor.org/stable/2288843

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* Elsevier Inc. doi:https://doi.org/10.1016/C2013-0-10517-X

Colorado. (2016). Retrieved from County Health Rankings & Roadmaps: http://www.countyhealthrankings.org/app/colorado/2016/overview

Concise rules of APA style (6 ed.). (2011). Washington D.C.: American Psychological Association.

Conlin, J. A., Gathercole, S. E., & Adams, J. W. (2005). Children's working memory: Investigating performance limitations in complex span tasks. *Journal of Experimental Child Psychology, 90*(4), 303-317. doi:doi:10.1016/j.jecp.2004.12.001

Conway, A. R., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4,* 577-590.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Craig, J. (2017). *Basic graphic design for PowerPoint.* Retrieved July 25, 2017, from Udemy, Inc.: www.udemy.com/basic-graphic-design-for-powerpoint/

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks: SAGE Publications, Inc.

Dunlap, J. C., & Lowenthal, P. R. (2016). Getting graphic about infographics: design lessons learned from popular infographics. *Journal of Visual Literacy, 35*(1), 42-59. Retrieved from http://dx.doi.org/10.1080/1051144X.2016.1205832

Durso, F. T., Pop, V. L., Burnett, J. S., & Stearman, E. J. (2011, July). Laws & rules: Evidence-based human factors guidlines for PowerPoint presentations. *Ergonomics in Design, Downloaded from erg.sagepub.com*, 4-7. doi:10.1177/1064804611416583

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA, USA: The MIT Press.

Ericsson, K. A., & Simon, H. A. (1992). *Protocol analysis: Verbal reports as data* (Rev. ed ed.). Cambridge, MA: MIT Press.

Evergreen, S. D. (2011). Death by boredom: The role of visual processing theory in written evaluation communication. *Dissertation downloaded from ProQuest*. ProQuest LLC.

Evergreen, S. D. (2017). *Effective data visualization: The right chart for the right data.* Thousand Oaks, CA, USA: Sage Publications, Inc.

Evergreen, S. D. (2018). *Presenting data effectiviely: Communicating your findings for maximum impact* (2nd ed.). Thousand Oaks, CA, USA: SAGE Publications, Inc.

Few, S. (2004). *Show me the numbers: Designing tables and graphs to enlighten.* Oakland, CA, USA: Analytics Press.

Few, S. (2006). *Information dashboard design: The effective visual communication of data.* Sebastopol, CA: O'Reilly Media, Inc.

Few, S. (2007). *Three blind men and an elephant: The power of faceted analytical displays.* Retrieved November 2016, from Perceptual Edge: http://www.perceptualedge.com/articles/Whitepapers/Three_Blind_Men.pdf

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 231-239.

Gilner, J. A., Morgan, G. A., & Leech, N. L. (2009). *Research methods in applied settings: An integrated approach to design and analysis* (2 ed.). New York: Routledge.

Graph. (2017). (W. Inc., Producer) Retrieved April 2, 2017, from Business Dictionary: http://www.businessdictionary.com/definition/graph.html

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, *36*(2), 212-224.

Haase, R. F., Waechter, D. M., & Solomon, G. S. (1982). How significant is a significant difference? Average effect size of research in counseling psychology. *Journal of Counseling Psychology, 29*(1), 58-65.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*, 446-474. doi:10.111/j.1756-8765.2011.01150.x

Hegarty, M., Canham, M., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task.

Journal of Experimental Psychology: Leaning, Memory, and Cognition, 36(1), 37-53.

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & cognition, 36*(7), 1351-1359. doi:10.3758/MC.36.7.1351

Hockley, W. E., & Bancroft, T. (2011). Extensions of the picture superiority effect in associative recognition. *Canadian Journal of Experimental Psychology, 65*(4), 236-244.

Hung, J., Edmonds, L. A., & Reilly, J. (2016). Words speak louder than pictures for action concepts: An eyetracking investigation of the picture superiority effect in semantic categorisation. *Language, Cognition & Neuroscience, 31*(9), 1150-1166.

Infographic. (2019, June 14). Retrieved from Wikipedia: The Free Encyclopedia: https://en.wikipedia.org/wiki/Infographic

Irrazabal, N., Saux, G., & Burin, D. (2016). Procedural mulitmedia presentations: The effects of working memory and task complexity on instruction time and assembly accuracy. *Applied Cognitive Psychology, 30*(6), 1052-1060.

Karadimitriou, S. M., & Marshall, E. (2015, 12 3). *Outliers, Durbin-Watson and interactions for regression in SPSS*. Retrieved from University of Sheffield: https://www.sheffield.ac.uk/polopoly_fs/1.531431!/file/MASHRegression_Further_SPSS.pdf

Keneally, M., & Diehm, J. (2015, June 15). *Everything you need to know about shark attacks as told by graphs.* Retrieved from ABC News: http://abcnews.go.com/US/shark-attacks-told-graphs/story?id=31779076

Kosslyn, S. M. (2007). *Clear and to the point: 8 psychological principles for compelling PowerPoint presentations.* New York, NY: Oxford University Press.

Kuhlman, B. B. (2014, August). The test-taking pupil: Effects of depletion, difficulty, and threat on pupil responsivity. University of Utah.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612-625. Retrieved from https://msu.edu/~levinet/eta%20squared%20hcr.pdf

Lowenthal, P. R. (2009). Improving the design of PowerPoint presentations. In P. R. Lowenthal, D. Thomas, A. Thai, & B. Yuhnke (Eds.), *The CU Online Handbook:*

Teach differently: Create and collaborate (p. 122). Denver, Colorado: University of Colorado Denver. Retrieved August 12, 2017, from http://cuonline.ucdenver.edu/handbook/

Luse, D. W., & Miller, R. A. (2011). Business faculty and students' perceptions of the effectiveness of PowerPoint usage as a teaching and learning tool. *Franklin Business & Law Journal, 2*, 83-94.

Mackiewicz, J. (2007). Audience perceptions of fonts in projected PowerPoint text slides. *Technical Communication, 54*(3), 295-307. Retrieved from http://www.jstor.org/stable/43089505

Magimairaj, B. M., & Montgomery, J. W. (2012). Children's verbal working memory: Role of processing complexity in predicting spoken sentence comprehension. *Journal of Speech, Language, and Hearing Research, 55*, 669-682.

Mahy, C. E., & Moses, L. J. (2015). The effect of retention interval task difficulty on young children's prospective memory: Testing the intention monitoring hypothesis. *Journal of Cognition and Development, 16*(5), 742-758. doi:10.1080/15248372.2014.930742

Malamed, C. (2009). *Visual language for designers: Principles for creating graphics that people understand.* Beverly: Rockport Publishers, Inc.

Mann, M. R. (2001, September 8). *Headlines.* Retrieved March 18, 2019, from Columbia University: http://www.columbia.edu/itc/journalism/isaacs/client_edit/Headlines.html

Mann, S., & Robinson, A. (2009). Boredom in the lecture theatre: An investigation into the contributers, moderators and outcomes of boredom amongst university students. *British Educational Research Journal, 35*(2), 243-258.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*(2), 380-386. doi:http://dx.doi.org/10.1037/0022-0663.100.2.380

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken, *Cognitive Load Theory* (pp. 131-152). New York: Cambridge University Press.

Merriam-Webster Inc. (2017). Glyph. *Merriam-Webster Dictionary*. Retrieved July 2, 2017

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Ngiam, W., Khaw, K., Holcombe, A., & Goodburn, P. (2018). Visual working memory for letters varies with familiarity but not complexity. *Journal of Experimental Psychology: Learning*, *Memory and Cognition*, 1-15. doi:http://dx.doi.org/10.1037/xlm0000682

Nowak, M. K., Speakman, E., & Sayers, P. (2016). Evaluating PowerPoint presentations: A retrospective study examining educational barriers and strategies. *Nursing Education Practices, 37*(1), 28-31.

Nussbaumer, C. (2011, November). *Visual battle: Table vs graph.* Retrieved from www.stortellingwithdata.com: http://www.storytellingwithdata.com/blog/2011/11/visual-battle-table-vs-graph

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding. *Cognitive Psychology, 5*(2), 176-206. doi:10.1016/0010-285(73)90032-7

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M. L., & Bertini, E. (2014). The persuasive power of data visualization. *New York University Public Law and Legal Theory Working Papers*, Paper 474. Retrieved from http://lsr.nellco.org//nyu_plltwp/474

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M., & Bertini, E. (2014). The persuasive power of data visualization. *IEEE Transaction on Visualization and Computer Graphics, 20*(12), 2211-2220.

Perea, M. (2013). Why does the APA recommend the use of serif fonts? *Psicothema, 25*(1), 13-17.

Pirrung, M. (2015). Effective data visualization for communication and analysis of microbiome data. Ohio: Doctoral Dissertation.

Plass, J. L., Moreno, R., & Brünken, R. (2010). Introduction. In J. L. Plass, R. Moreno, & R. Brünken, *Cognitive Load Theory* (pp. 1-6). New York: Cambridge University Press.

Price, M., Cameron, R., & Butow, P. (2007). Communicating risk information: The influence of graphical display format on quantitative information perception--Accuracy, comprehension and preferences. *Patient Education and counseling, 69*, 121-128. doi:10.1016.j.pec.2007.08.006

Rattner, S. (2017, January 3). *2016 in charts. (And can Trump deliver in 2017?).* Retrieved from The New York Times:

www.nytimes.com/2017/01/03/opinion/2016-in-charts-and-can-trump-deliver-in-2017.html

Ratwani, R. M., & Trafton, G. J. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. *Psychonomic Bulletin & Review, 15*(4), 757-762. doi:10.3758/PBR.15.4.757

Robbins, N. B. (2005). *Creating more effective graphs.* Hoboken, New Jersey, USA: John Wiley & Sons.

Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: The complexity of the tasks matters. *Learning and Instruction, 49,* 142-156. doi:http://dx.doi.org/10.1016/j.learninstruc.2017.01.008

Rosenthal, J. A. (1994). Reliability and social work research. *Social Work Research, 18*(2), 115. Retrieved from https://doi-org.du.idm.oclc.org/10.1093/swr/18.2.115

Savoy, A., Proctor, R. W., & Salvendy, G. (2009). Information retention from PowerPoint and traditional lectures. *Computers & Education, 52,* 858-867.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *Int. J. Human--Computer Studies, 45,* 185-213. Academic Press Limited.

Seifert, L. S. (1997). Activating representations in permanent memory: Different benefits for pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1106-1121.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General, 124*(1), 43-61.

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

Sherin, A. (2013). *Design elements: Using images to create graphic impact: A graphic style manual for effective image solutions in graphic design.* Beverly, MA: Rockport Publishers.

Sierra, E. A., Fisk, A. D., & Rogers, W. A. (2002). Matching instructional media with instructional demands. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46,* p. 2089=2093. doi:https://doi-org.du.idm.oclc.org/10.1177%2F154193120204602520

Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology, 5,* 61-77.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory of pictures: single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*(2), 73-74.

Stenberg, G. (2007). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology, 18*(5), 813-847. doi:10.1080/09541440500412361

Suresh, K. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human reproductive Sciences, 4*(1), 8-11. doi:10.4103/0974-1208.82352

Susskind, J. E. (2005). PowerPoint's power in the classroom: Enhancing students' self-efficacy and attitudes. *Computers & Education, 45*, 203-215.

Sweller, J. (1994). Cognitive Load Theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

Sweller, J. (2010). Cognitive Load Theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 29-47). New York: Cambridge University Press.

Sweller, J. (2010A). Cognitive Load Theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brunken, *Cognitive Load Theory* (pp. 29-47). New York: Cambridge University Press.

Sweller, J. (2010B). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123-138.

Sweller, J. (2016). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition, 5*, 360-367.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer.

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics.* Upper Saddle River: Pearson Education, Inc.

Taylor, K. L., & Dionne, J.-P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology, 92*(3), 413 - 425.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly, 68*, 368-393.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly, 71*(1), 91-112.

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2013). "Up means good": The effect of screen position on evaluative ratings in web surveys. *Public Opinion Quarterly, 77*(Special Issue), 69-88.

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

Tufte, E. (2006). *Beautiful evidence.* Cheshire, Connecticut: Graphics Press LLC.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, Mass. : Addison-Wesley Pub. Co.

Turner, M. L. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154. doi:https://doi-org.du.idm.oclc.org/10.1016/0749-596X(89)90040-5

van Someren, M., Barnard, Y., & Sandberg, J. (1994). *The think aloud method: A practical approach to modelling cognitive processes.* London: Academic Press.

Vogt, S. &. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology, 54*(4), 298-303.

Wainer, H. (1990). Graphical visions from William Playfair to John Tukey. *Statistical Science, 5*(3), 340-346.

Walton, M., Vukovic, V., & Marsden, G. (2002). 'Visual literacy' as challenge to the internationalisation of interfaces: A study of South African student web users. *CHI EA '02 CHI '02 Extended Abstracts on Human Factors in Computing Systems. 20*, pp. 530-531. Minneapolis: ACM.

Ware, C. (2013). *Information visualization: Perception for design.* Waltham: Elsevier.

White, P. (1980). Limitations of verbal reports of internal events: A refutaion of Nisbett and Wilson and of Bem. *Psychological Review, 87*(1), 105-112.

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10*(1), 80-87.

Xie, H., Wang, F., Hao, Y., Chen, J., An, J., & Wang, Y. (2017). The more total cognitive load is reduced by cues, the better retention and transfer of multimedia learning: A meta-analysis and two meta-regression analyses. *PLOS One, 12*(8), 1-20. Retrieved from https://doi.org/10.1371/journal.pone.0183884

Yang, S. (2003). Reconceptualizing think-aloud methodology: Refining the encoding and categorizing techniques via contextualized perspectives. *Computers in Human Behavior, 19*(1), 95-115. doi:https://doi.org/10.1016/S0747-5632(02)00011-0

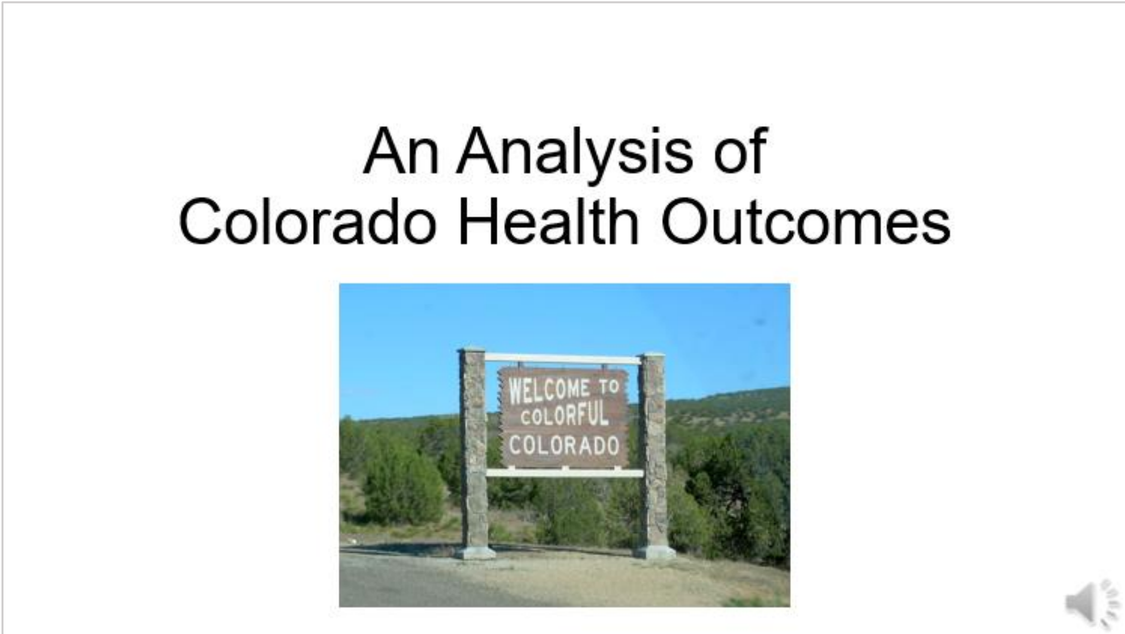
Appendices

Appendix A– Principles for creating effective slides

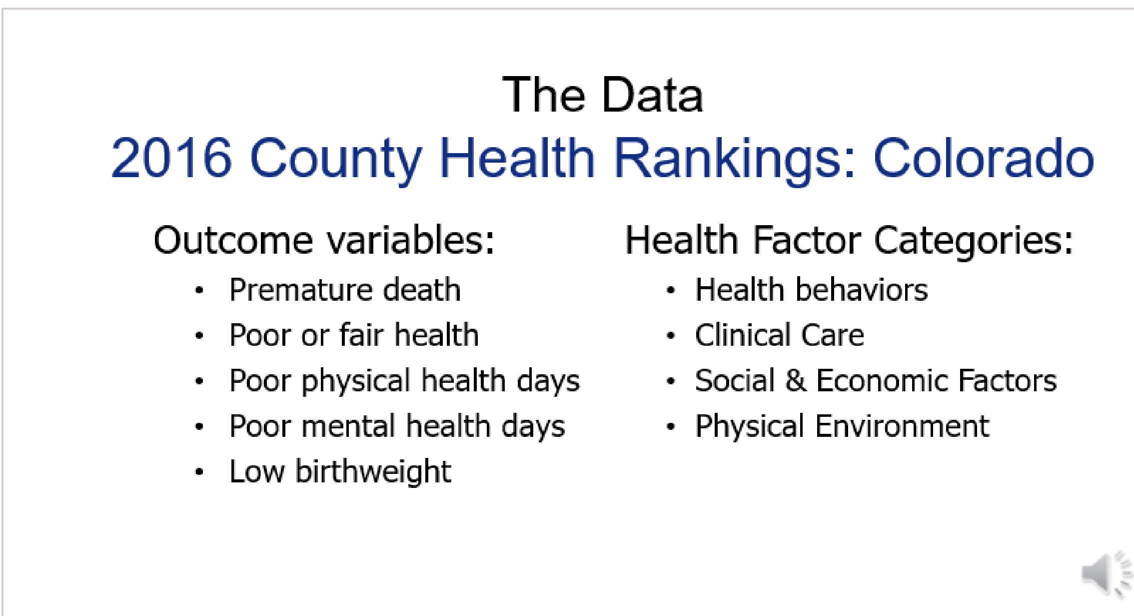
Principle	Detail	Author	Page
Speech	Spoken information, rather than text information, should accompany images.	Ware, 2013	333
	Use some form of pointer or timely highlighting to link spoken words and images.	Ware, 2013	335
	The relevant part of the visualization should be highlighted just before the start of the accompanying speech segment.	Ware, 2013	335
Font	Avoid all uppercase, all italics, or all bold	Kosslyn, 2007	62
	Don't underline	Kosslyn, 2007	63
	Use bold, italics or a change in color for emphasis	Kosslyn, 2007	64
	Minimize embellishments such as bold and italics. Avoid others such as highlighted or flashing text	Durso et al	5
	Use color for emphasis, to specify different classes of info, to group words together. Don't use vary text color for decoration	Kosslyn, 2007	65
	Use fonts that are easy to read, don't vary font for decoration, vary font for emphasis or to group words together	Kosslyn, 2007	66
	Use readable fonts	Berk, 2011 Nowak, Speakman, & Sayers, 2016	25 31
	Don't use visually complex fonts because it takes time and cognitive energy to read	Kosslyn, 2007	67
	Ensure that words are large enough to read easily	Kosslyn, 2007	68
	Ensure that viewers can easily discriminate text from the background.	Kosslyn, 2007	69
	Low contrast and even color background if used at all	Kosslyn, 2007	69
	San serif fonts better with small font because serifs get grouped together making letters hard to distinguish	Kosslyn, 2007	71

Principle	Detail	Author	Page
	Serif fonts better with low contrast because serifs give additional clues regarding identity of letter	Kosslyn, 2007	72
	Use standard fonts so it comes out the same on other computers	Kosslyn, 2007	73
	Use font size of at least 22 points for bullets, 16 point for figure legends and axes. Consistent 30 point font	Durso et al Nowak, Speakman, & Sayers, 2016	4 31
	Use sentence case for bullets	Durso et al	5
	Avoid compressed or extended texts	Durso et al	5
Color	Use high-contrast text-to-background combinations	Durso et al	5
Color	Use dark text on a light background (light gray may reduce eyestrain over stark white)	Durso et al	5
	High-contrast colors	Berk, 2011	25
	Consider split-complimentary colors e.g. dark blue on pale red-orange use one adjacent to direct contrast color and fade background color	Durso et al	5
	Avoid red/green contrasts because 5%-8% of males are colorblind and can't distinguish red-green combos.	Durso et al	5
Layout	Line space should be half a character height. Additional space between bullets	Durso et al	5
	Respect slide margins because can't control how much of slide projects	Durso et al	5
General	Be consistent with fonts and colors from slide to slide	Durso et al	5
	Avoid distracting elements such as animation	Durso et al	6
	3-D graph option appears to give two estimates for the y-axis. False illusion of volume. Don't use.	Durso et al	6
	Full-sentence headline, written as an assertion	Berk, 2011	25
	Less is more as long as it is enough. Search time and errors increase with graph complexity.	Durso et al	6

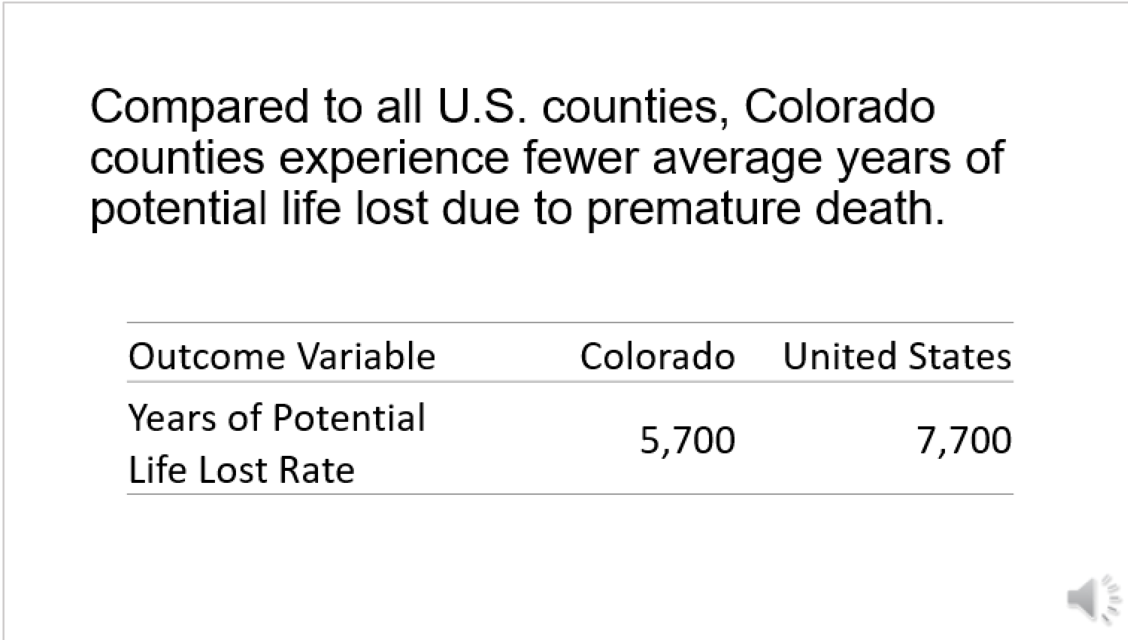
Appendix B – Images of Slides from PowerPoints



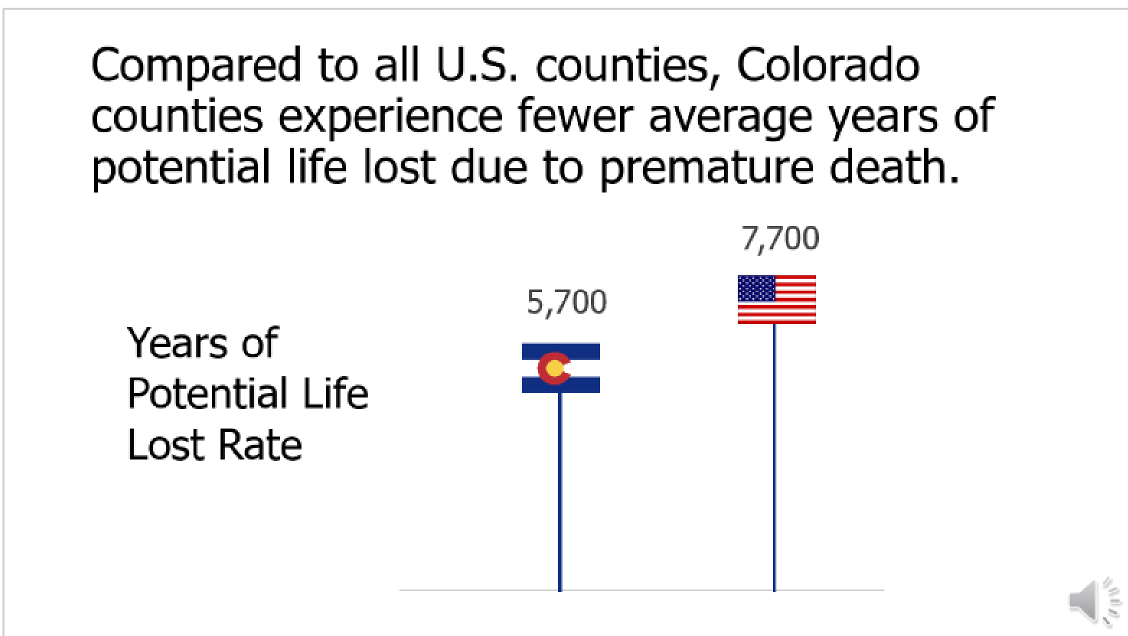
Slide 1- This slide is the same in both presentations



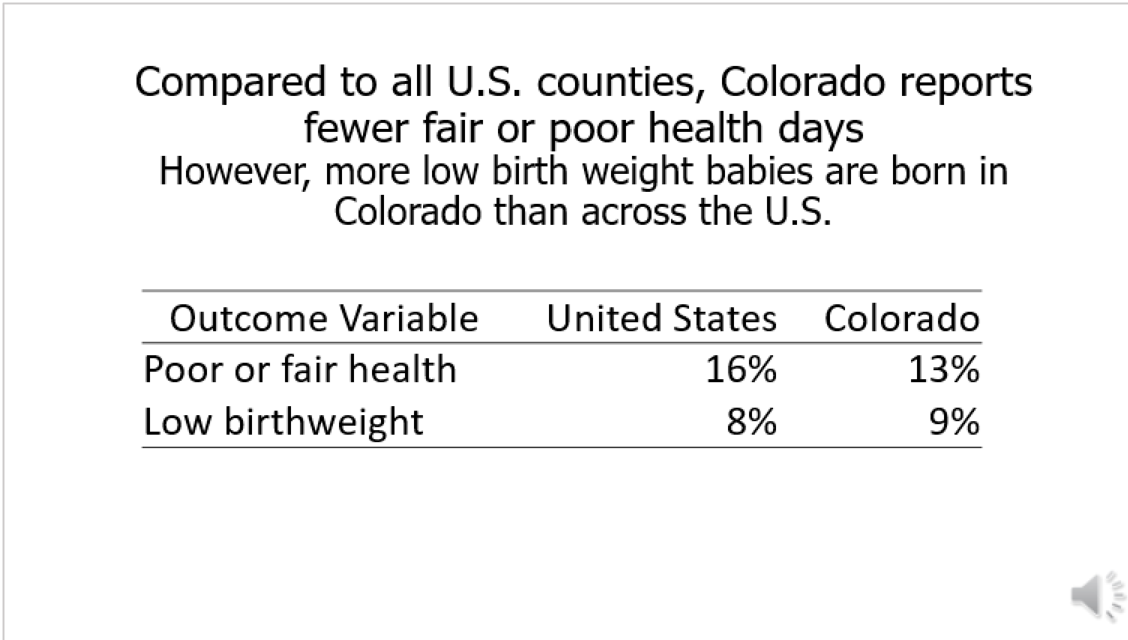
Slide 2 - This slide is the same in both presentations



Slide 3 - Tables presentation

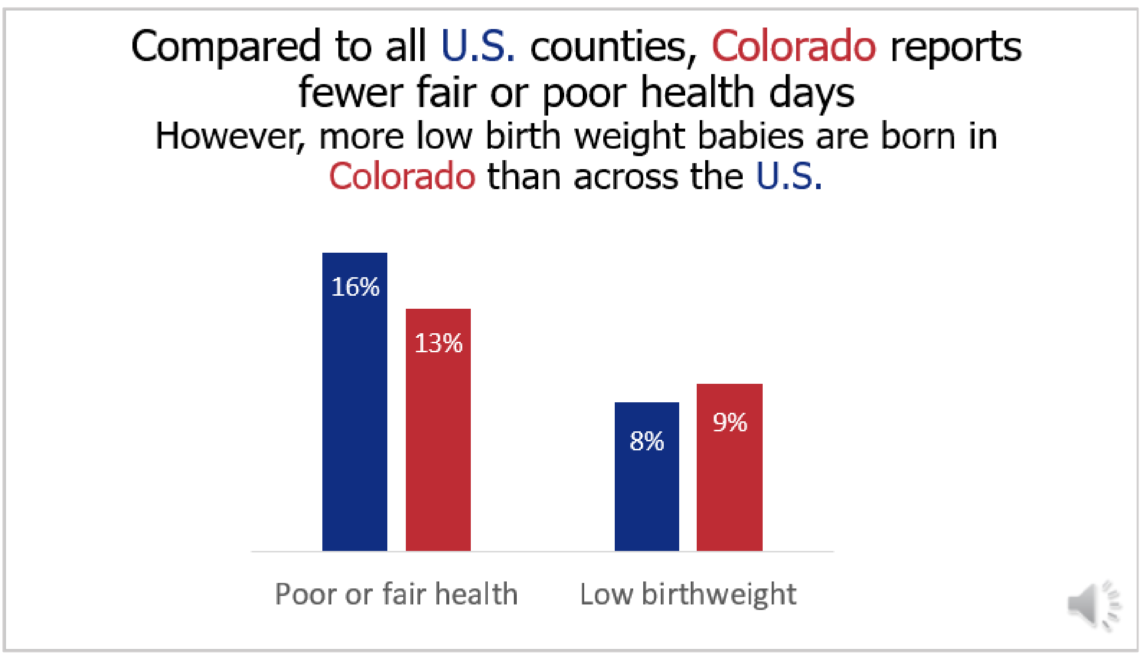


Slide 3 - Graphs presentation

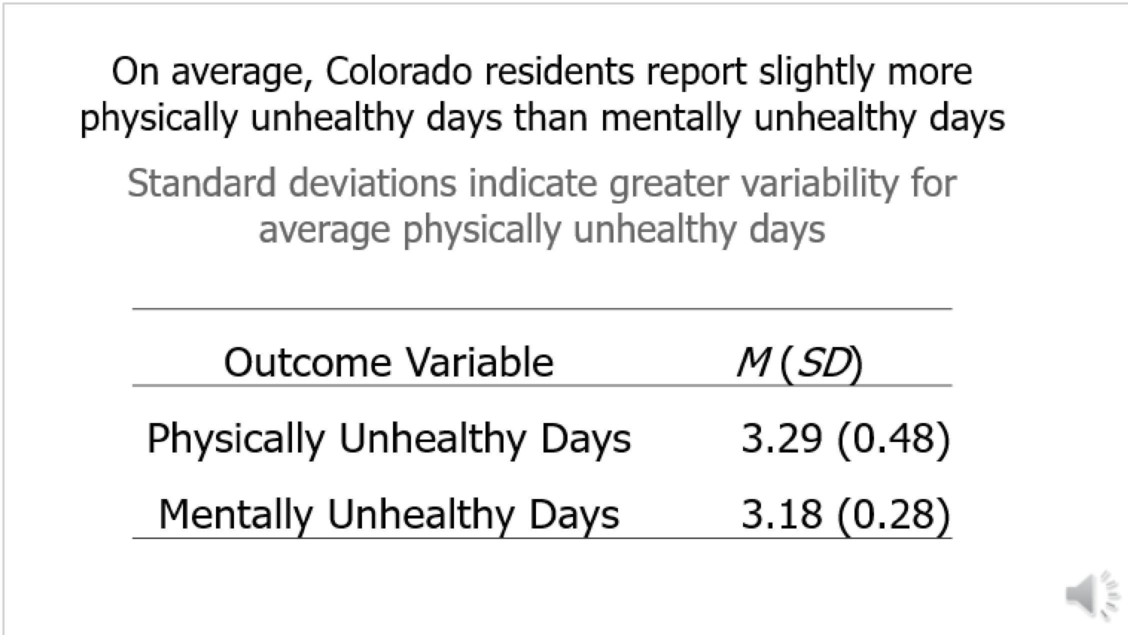

Compared to all U.S. counties, Colorado reports fewer fair or poor health days
However, more low birth weight babies are born in Colorado than across the U.S.

Outcome Variable	United States	Colorado
Poor or fair health	16%	13%
Low birthweight	8%	9%

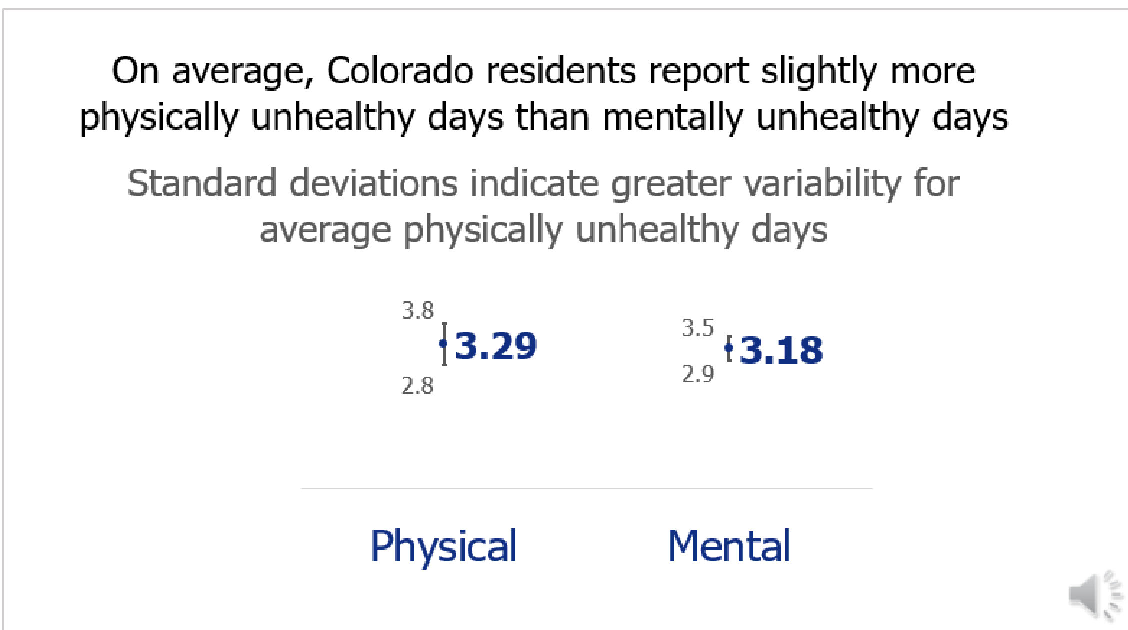
Slide 4 - Tables presentation



Slide 4 - Graphs presentation



Slide 5 - Tables presentation

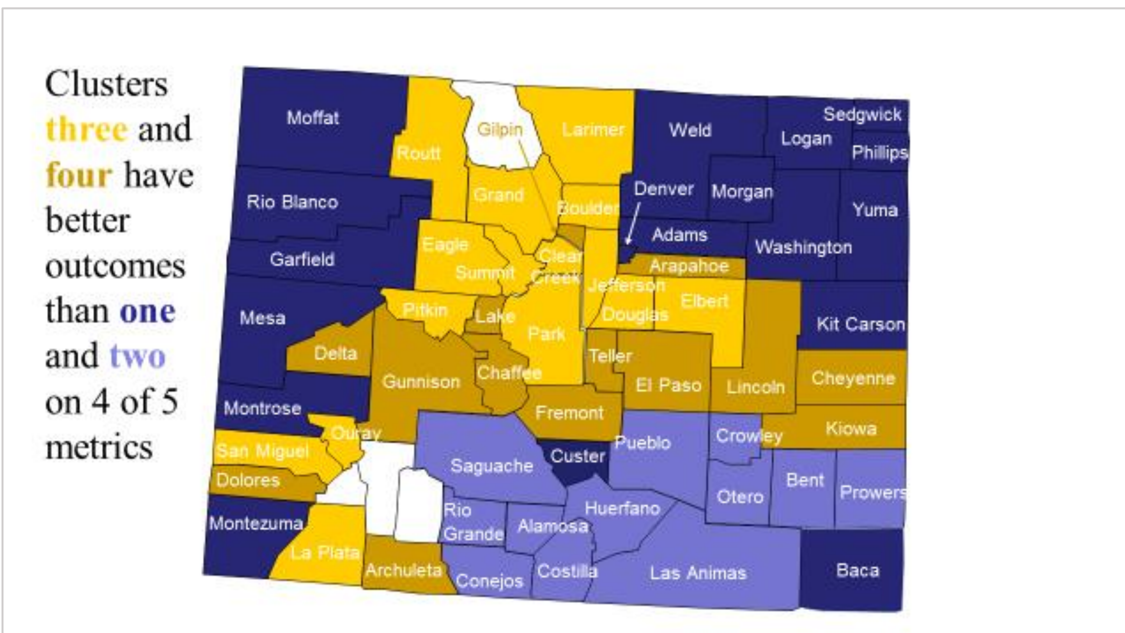


Slide 5 - Graphs presentation

Cluster Membership						
Cluster One	Adams	Denver	Logan	Montezuma	Phillips	Washington
	Baca	Garfield	Mesa	Montrose	Rio Blanco	Weld
	Custer	Kit Carson	Moffat	Morgan	Sedgwick	Yuma
Cluster Two	Alamosa	Conejos	Crowley	Las Animas	Prowers	Rio Grande
	Bent	Costilla	Huerfano	Otero	Pueblo	Saguache
Cluster Three	Boulder	Eagle	Jefferson	Ouray	Routt	Summit
	Broomfield	Elbert	La Plata	Park	San Miguel	Douglas
	Clear Creek	Grand	Larimer	Pitkin		
Cluster Four	Arapahoe	Cheyenne	El Paso	Gilpin	Kiowa	Lincoln
	Archuleta	Delta	Fremont	Gunnison	Lake	Teller
	Chaffee	Dolores				
No Data	San Juan	Hinsdale	Mineral	Jackson		

NOTE: Clusters three and four have better outcomes than one and two on 4 of 5 metrics

Slide 6 - Tables presentation



Slide 6 - Graphs presentation

On average, more poor health outcomes characterize the worst and poor clusters than the best and good clusters.

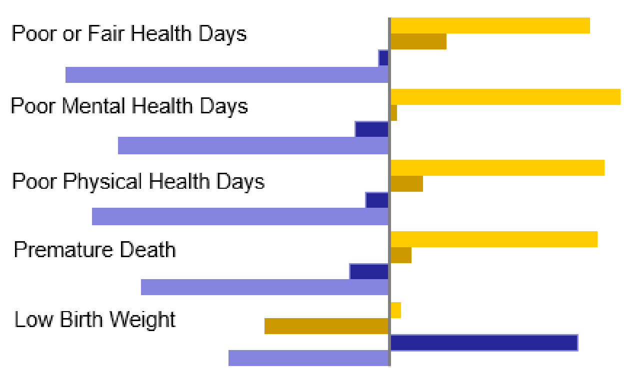
However, the poor and good clusters have more low birth weight babies on average than the worst and best clusters.

	Normalized Average Outcome				
Cluster	Low birthweight	Premature death	Poor or fair health	Poor mental health days	Poor physical health days
Poor	-0.96	0.20	0.05	0.16	0.11
Worst	0.80	1.25	1.63	1.37	1.50
Best	-0.07	-1.06	-1.02	-1.17	-1.09
Good	0.62	-0.12	-0.29	-0.04	-0.18

Slide 7 - Tables presentation

On average, more poor health outcomes characterize the **worst** & **poor** clusters than the **best** & **good** clusters.

However, the **worst** & **good** clusters have more low birth weight babies on average than the **poor** & **best** clusters.



Slide 7 - Graphs presentation

Health outcomes show positive correlations.

	Premature death	Poor or fair health	Poor physical health days	Poor mental health days	Low birthweight
Premature death	-				
Poor or fair health	.72**	-			
Poor physical health days	.70**	.96**	-		
Poor mental health days	.72**	.90**	.94**	-	
Low birthweight	.16	.28*	.23	.16	-

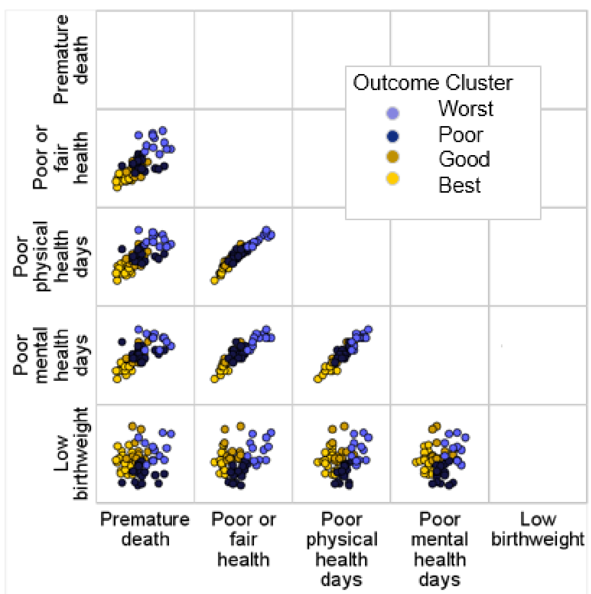
**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

N = 60

Slide 8 - Tables presentation

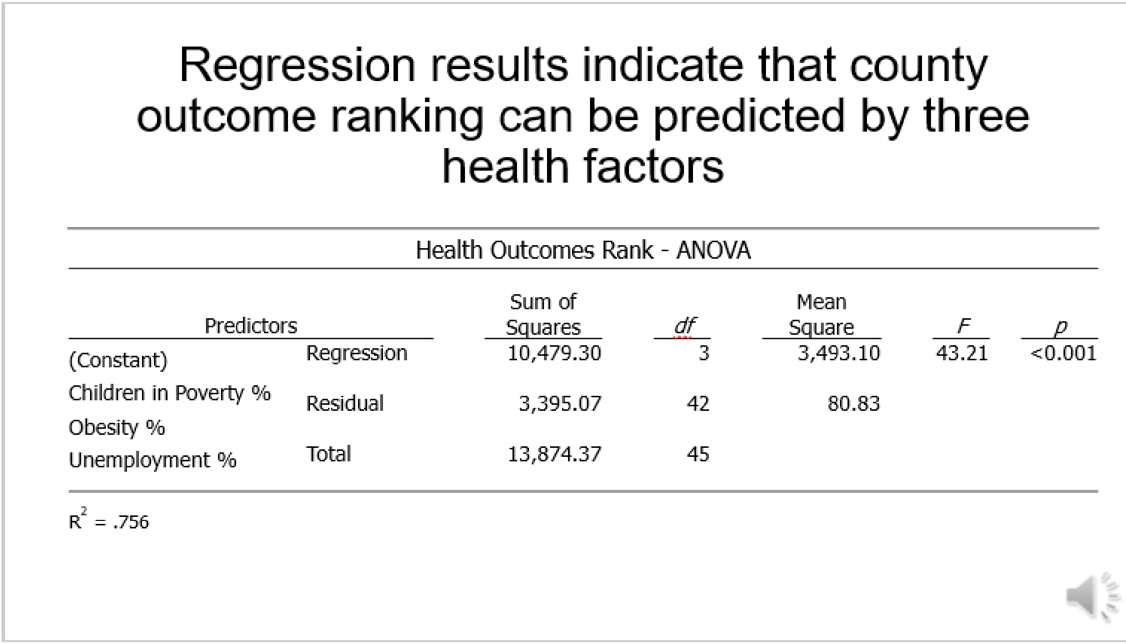
Health outcomes show positive correlations* across clusters



N = 60

*Correlations between health outcomes are all statistically significant. For low birth weight, only the correlation between birth weight and poor or fair health is significant at $p < .05$.

Slide 8 - Graphs presentation



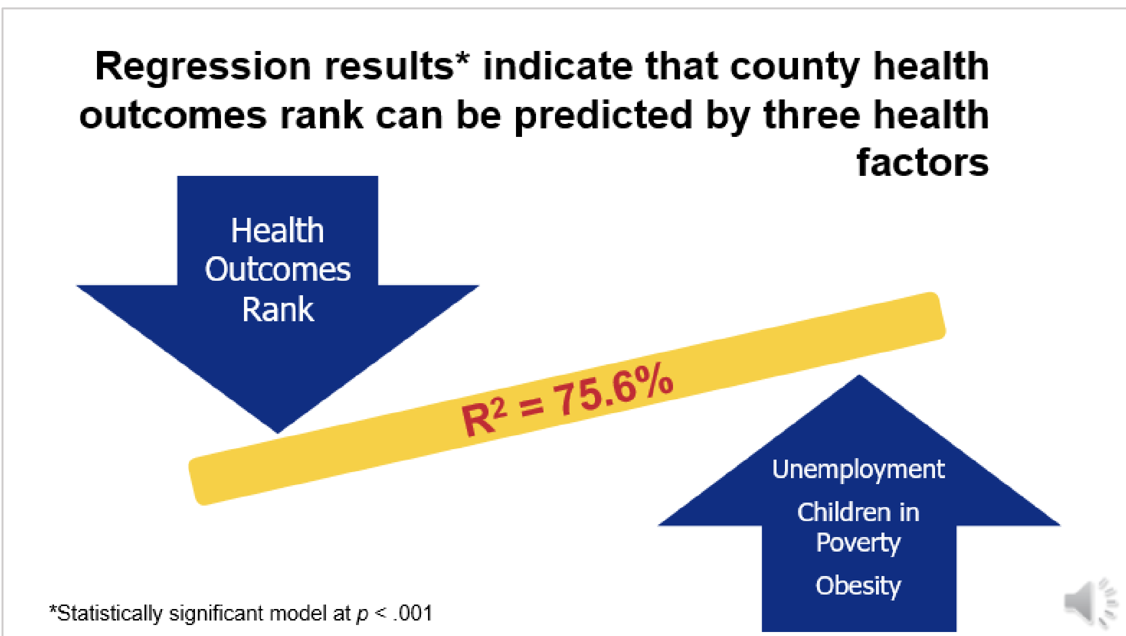
Regression results indicate that county outcome ranking can be predicted by three health factors

Health Outcomes Rank - ANOVA

Predictors		Sum of Squares	*df*	Mean Square	*F*	*p*
(Constant)	Regression	10,479.30	3	3,493.10	43.21	<0.001
Children in Poverty %	Residual	3,395.07	42	80.83		
Obesity %						
Unemployment %	Total	13,874.37	45			

$R^2 = .756$

Slide 9 - Tables presentation



Slide 9 - Graphs presentation

As adult obesity rate, percent of children in poverty, and unemployment rate rise, county health outcome rank falls

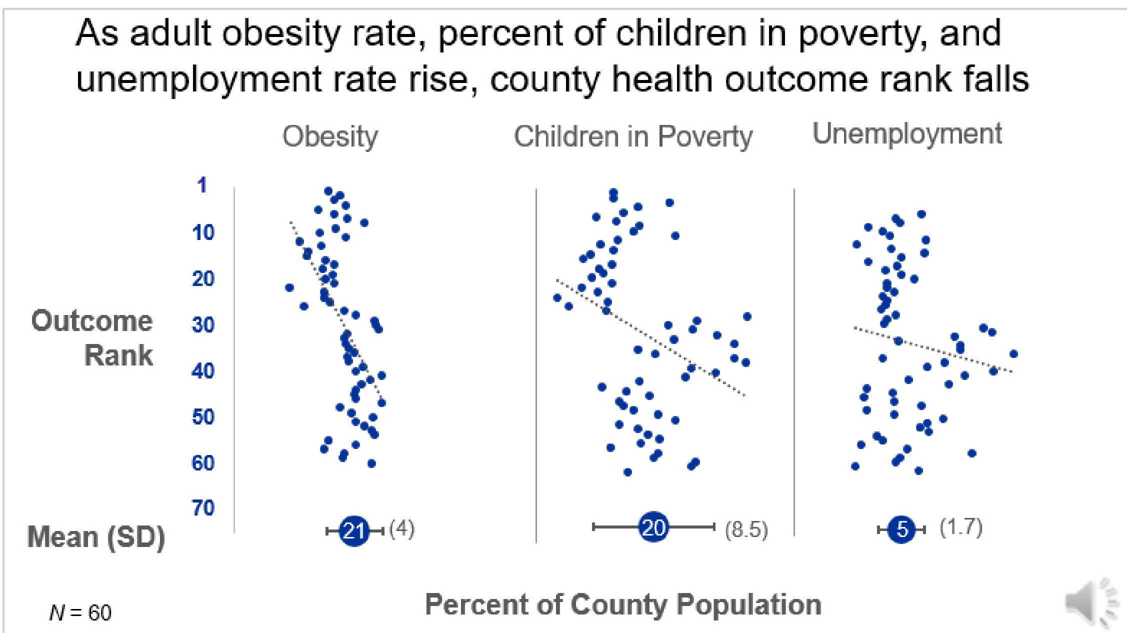
Correlations

	Health Outcomes Rank	Obesity %	Children in Poverty %	Unemployment %
Health Outcomes Rank	-			
Obesity %	-.542*	-		
Children in Poverty %	-.781*	.44*	-	
Unemployment %	-.604*	0.17	.62*	-
M	30.50	20.72	20.35	5.24
SD	17.46	4.00	8.53	1.67

*. Correlation is significant at the 0.01 level (2-tailed).

N = 60

Slide 10 - Tables presentation



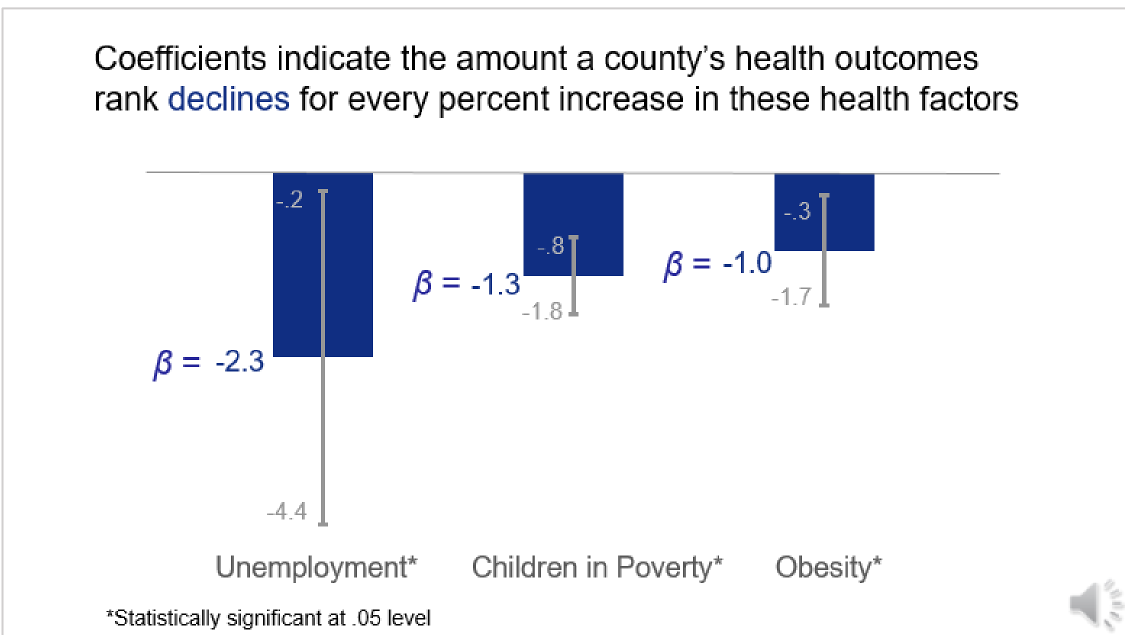
Slide 10 - Graphs presentation

Coefficients indicate the amount a county's health outcome ranking declines for every percent increase in these health factors

Model	Unstandardized Coefficients[a]				95.0% Confidence Interval for B	
	B	Std. Error	*t*	*p*	Lower Bound	Upper Bound
Unemployment	-2.3	1.0	-2.236	0.031	-0.2	-4.4
Children in Poverty	-1.3	0.2	-5.385	0.000	-0.8	-1.8
Obesity	-1.0	0.3	-2.848	0.007	-0.3	-1.7

a. Dependent Variable: Health Outcomes Rank

Slide 11 - Tables presentation

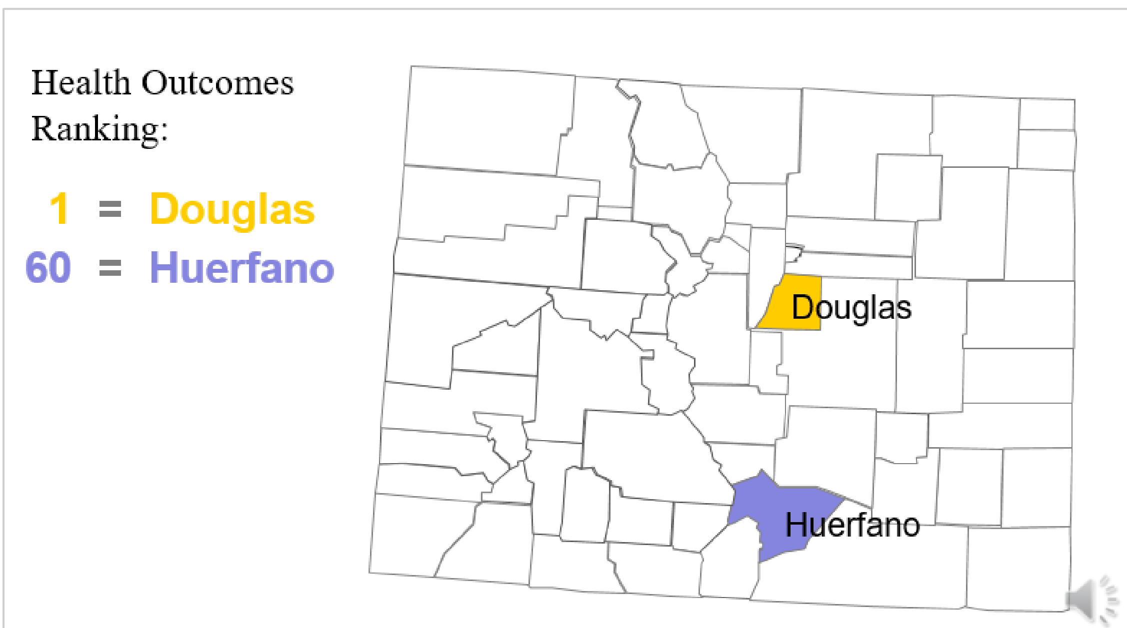


Slide 11 - Graphs presentation

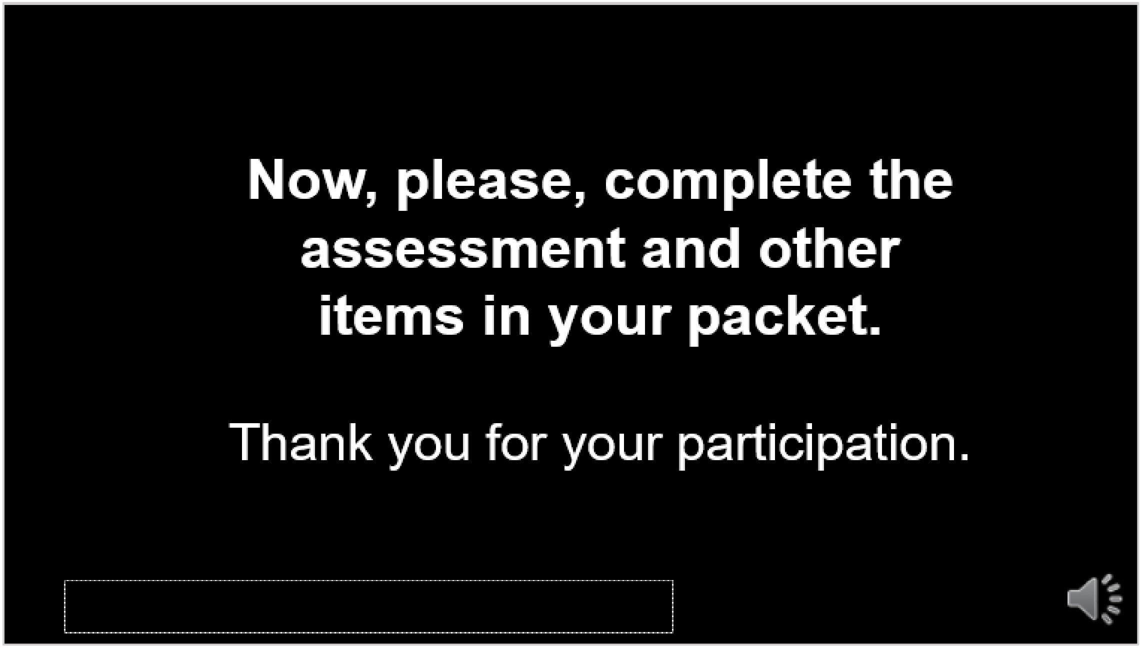
Health Outcomes Ranking

County	Health Outcome Rank
Douglas	1
Huerfano	60

Slide 12 - Tables presentation



Slide 12 - Graphs presentation



Slide 13 – This slide is the same in both presentations. It is not included in any analysis.

Appendix C –Script for Presentations by Slide

NOTE: Easy, moderate, and difficult refer to the estimated difficulty of the slide content.

Slide Pair 1

Easy - This presentation is about Colorado Health Outcomes and Health Factors that might impact population health outcomes.

Slide Pair 2

Moderate - The data used in this analysis are from the 2016 County Health Rankings for the state of Colorado. The data set has five outcome variables that are used to rank the counties from best to worst.

The five outcome variables are: premature death, poor or fair health, poor or fair physical health days, poor or fair mental health days, and low birth weight. Data on premature death and low birth weight babies are collected by the National Vital Statistics System and provided by the Center for Disease Control’s National Center for Health Statistics. Poor or fair health, poor or fair physical health days, poor or fair mental health days measure quality of life. These metrics are compiled from the CDC’s core Behavioral Risk Factor Surveillance Survey, a survey that has been in use since 1993.

The data set also includes a number of health factors or inputs. Health factors consist of a total of 30 different metrics in 4 categories. The health behaviors category includes such things as smoking rates, obesity rates, and other lifestyle factors. The clinical care category relates to the number of care providers, insurance rates, and usage rates of some services. Social and economic factors include levels of education, unemployment, poverty, and several other variables. While factors relating to the physical environment category include such things as air, water, and housing quality, and long commutes. Health factor data comes from a variety of sources.

Slide Pair 3

Easy - The 2016 county health rankings provides median values for each factor and outcome for the entire United States so that comparisons can be made with Colorado's median values. Happily, Colorado experienced 2000 fewer years of potential life lost than the whole country. Years of potential life lost calculates the total number of years lost for people who died before reaching age 75 per 100,000 population.

Slide Pair 4

Moderate - Poor or fair health is a self-reported health outcome. Coloradans report 3% fewer poor or fair health days than the average American. Unfortunately, Colorado counties experience, on average, 1% more low birth weight babies. Low birth weight is a birth weight of less than 5.5 pounds.

Slide Pair 5

Moderate - Physically unhealthy and mentally unhealthy days are also self-reported. Coloradans report somewhat more physically unhealthy days than mentally unhealthy days. The range of reported physically unhealthy days is also greater than for mentally unhealthy days.

Slide Pair 6

Easy - In an effort to explore relationships between Colorado counties, cluster analysis was conducted using the outcome variables to see if there was any link between counties with low rankings compared to counties with high rankings. The clusters are based on z-scores which are a way to standardize values based on averages. A positive z-score means that the county's score is above average and a negative z-score means that it is below average with exactly average having a z-score of zero. Using z-scores allows comparisons across all of the rankings.

On this slide we can see which counties are in which cluster. In general, the best and good clusters have low outcomes compared to poor and worst clusters. Interestingly, the counties cluster into nearly contiguous, geographic regions.

Four of Colorado's 64 counties did not have data so they aren't included in the clustering. The four counties that are not included are Hinsdale, Jackson, Mineral and San Juan.

Slide Pair 7

Difficult - On this slide we see a comparison of the four clusters by health outcomes. Remember that when using z-scores the average score is zero. So using the average as a base-line, we can see that the Worst and Poor clusters have poorer health outcomes on at least 4 of 5 metrics. These four are poor health days, poor or fair physical health days, poor or fair mental health days, and premature death. The Worst cluster has dramatically poorer health outcomes than the Poor cluster. The Best and Good clusters have better health outcomes on average on those same 4 metrics, with the Best cluster having better outcomes than the Good cluster on the four metrics. Results are switched for low birthweight babies with the Good cluster experiencing above average numbers of low birthweight babies and the Poor cluster experiencing below average numbers. For the good cluster this is contrasted with below average outcomes on the other metrics while the Worst cluster has the worst outcomes on every metric. So a difference in the two clusters with higher than average outcomes is that the Worst cluster has the highest average on all 5 outcomes while the Poor cluster is above average on most of the outcomes except low birth weight. Interestingly, this cluster which generally has poor outcomes, has the fewest low birth weight babies of all the clusters with a z-score of .80 compared to the Best cluster's z-score of .07. The Best cluster is the only cluster with below average scores on all 5 outcomes.

Slide Pair 8

Difficult - This slide shows correlations between the five health outcomes. Positive correlations indicate that as one variable goes up so does the other. Strong correlations indicate that clusters with higher scores for one variable will generally have higher scores for the other. Strong correlations exist between all the outcome variables with the exception of low birthweight. However, the correlation between low birthweight and poor or fair health is statistically significant.

Slide Pair 9

Moderate - We were also interested to see if a county's health outcome rank could be predicted by any health factors. To this end, a multiple regression was performed using health outcome rank as the dependent variable and all 30 of the health factors as independent variables. Using the stepwise method to select significant variables, a model with only three health factors as independent variables was created that predicts a county's outcome rank with almost 76% accuracy. The three health factors are the percent of population over 16 years old that are unemployed, percent of children in poverty, and percent with adult obesity.

Slide Pair 10

Difficult - Correlations between each of the three health factors, unemployment, children in poverty, and adult obesity, and health outcome rank are statistically significant and negative. A negative correlation indicates that as a variable goes up its correlate goes down so, for example, as a county's obesity rate goes up its health outcome rank goes down. For all Colorado counties, the mean percent of adults with obesity, that is a body mass index greater than 30, is almost 21% with a SD of 4. The average percent of children in poverty is about 20 with a SD of 8.5%., and the mean unemployment rate is 5 and a quarter percent with a SD of 1.7.

Slide Pair 11

Difficult - Regression results yield a coefficient for each statistically significant variable. The coefficient tells us the change in a county's health outcome rank with a one percent change in the variable if everything else is held constant. Using the unstandardized coefficients this means that a one percent increase in unemployment leads to decline in health outcome rank of 2.3. A one percent increase in the number of children in poverty leads to a lower health outcome rank by a factor of 1.3. A county's health outcome rank is inversely related by a factor of one to every one percent change in adult obesity rate.

Confidence intervals indicate the range within which 95% of the population distribution is contained.

Slide Pair 12

Easy - By now you might be wondering which Colorado counties had the best and worst health outcome in 2016. The top ranked county is Douglas County, home to the cities of Castle Rock and Parker. At the bottom of the ranking is Huerfano County. Walsenburg is the largest city there.

Slide Pair 13

Thank you for participating in this research study. Now, please, complete the assessment and other items in your packet.

Appendix D – Slide difficulty and related assessment questions

Question Counts

Assessment 1 Question Number	Assessment 2 Question Number	Slide Number	Slide Complexity
1	4	1	Easy
2	2	2	Moderate
3	5	3	Easy
4	17	4	Moderate
5	8	4	Moderate
6	14	4	Moderate
7	12	5	Moderate
8	16	5	Moderate
9	6	6	Easy
10	9	6	Easy
11	21	7	Difficult
12	13	7	Difficult
13	15	8	Difficult
14	11	9	Moderate
15	1	9	Moderate
16	18	10	Moderate
17	10	10	Difficult
18	19	10	Difficult
19	3	11	Difficult
20	7	11	Difficult
21	20	12	Easy

Totals

Difficulty Level	Number of questions
Difficult	7
Moderate	9
Easy	5

Appendix E – Principles for Creating Effective Graphs

Element	Detail	Author	Page
Aspect ratio	Orient line segments to 45 degrees.	Robbins, 2005	231 269
	Choose an aspect ratio that shows variation in the data	Cleveland, 1994	70
Axis	Avoid deceptive double Y axis	Robbins, 2005	265
	Horizontal axis should increase left to right, Vertical axis should increase bottom to top	Robbins, 2005	283
Captions	Choose the graphic that represents the information most explicitly.	Carpenter & Shaw, 1998	97
Clarity	Avoid line patterns or texture that are visually active or cause optical illusions.	Tufte, 2001 Ware, 2013	92 62
	Remove visual distortions.	Ware, 2013	66
	Make sure the combination of closure, common region, and layout are perceived as figures, not ground.	Ware, 2013	190
CLT	Don't require reader to make calculations that a computer can make more easily	Robbins, 2005	216
	Plot the variable of interest i.e. Improvement rather than before and after	Robbins, 2005	217
CLT/Gestalt	Use proximity, connectedness, and common region to associate written labels with graphical elements.	Ware, 2013	322
	Take into account human sensory capabilities so important data elements and data patterns can be quickly perceived.	Ware, 2013,	14
Color	Use color selectively to highlight	Tufte 2001	92
	Color hue is effective for distinguishing groups. Varying density or saturation can distinguish groups.	Robbins, 2005, 2005	
	Avoid using gray scale as a method for representing more than a few (two to four) numerical values.	Ware, 2013	75

Element	Detail	Author	Page
	If large areas are defined using similar colors, consider using thin border lines in a contrasting color to help define the shapes.	Ware, 2013	113
	If using color saturation to encode numerical quantity, use greater saturation to represent greater numerical quantities. Avoid using a saturation sequence to encode more than three values.	Ware, 2013	117
	Consider using red, green, yellow, and blue to color code small symbols.	Ware, 2013	123
	If colored symbols are hard to see against parts of the background, add a contrasting border, e.g. black around a yellow symbol or white around a dark blue symbol.	Ware, 2013	124
	To create a set of symbol colors that can be distinguished by most colorblind individuals, ensure variation in the yellow–blue direction.	Ware, 2013	124
	Do not use more than ten colors for coding symbols if reliable identification is required, especially if the symbols are to be used against a variety of backgrounds.	Ware, 2013	124
	Use low-saturation colors to color code large areas. Generally, light colors will be best because there is more room in color space in the high-lightness region than in the low-lightness region.	Ware, 2013	125
	When color coding large background areas overlaid with small colored symbols, use all low-saturation, high-value (pastel) colors for the background, together with high-saturation symbols on the foreground.	Ware, 2013	125
	Maintain contrast with the background when highlighting text by changing the color of the font. E.g. High-saturation dark colors on a white/light	Ware, 2013	126

Element	Detail	Author	Page
	background, low-saturation light colors on a dark background.		
	To minimize the effort of visual searches, make displays as compact as possible, while maintaining clarity.	Ware, 2013	141
	Use more saturated colors when color coding small symbols, thin lines, or other small areas. Use less saturated colors for coding large areas.	Ware, 2013	108
Connection	Link data representations using lines to show relationships.	Ware, 2013	183
	Use connecting lines, enclosure, grouping, and attachment to represent relationships between entities. The shape, color, and thickness of lines and enclosures can represent the types of relationships.	Ware, 2013	226
Data rectangle	Make data rectangle slightly smaller than the scale-line rectangle	Robbins, 2005	Visual Clarity
Data/Ink ratio	No extraneous info masking data	Robbins, 2005	154
	Eliminate ink that does not express information.	Tufte, 2001 Cleveland, 1994	93 29
	Use visually prominent graphical elements to show the data	Robbins, 2005	163
	Emphasize the data	Robbins, 2005	159
	Make the data stand out. Avoid superfluity	Tufte, 2001 Cleveland W. S., 1994	13 25
	Maximize data density and the size of the data matrix, within reason	Tufte, 2001	168
Dimensions	Do not show changes in 1-D using 2- or 3-D ie. Don't show change in length with change in area	Robbins, 2005 Tufte, 2001	203 71, 77
	Display same number of dimensions as the number of dimensions in the data.		
	Don't depict higher dimensional data with arbitrary dimensions in complex figures such at stars, ships, or glyphs.	Robbins, 2005	195
	Dimensions must be proportional to data; we visualize area, use area		

Element	Detail	Author	Page
	For viewers to detect differences in length between two line segments use a fixed percent increase in the length.		
Enclosure	Put related information inside a closed contour or use color or texture.	Ware, 2013	187
General	Mobilize every graphical element, perhaps several times over, to show the data	Tufte, 2001	139
	Make important data stand out.	Ware, 2013	14
	Greater values should be more distinct.	Ware, 2013	14
	Make all visual distinctions as subtle as possible, but still clear and effective.	Tufte, 2001	92
	Superposed data sets must be readily visually assembled	Robbins, 2005 Cleveland, 1994	167 51
	Strive for clarity	Robbins, 2005	219
	Be consistent in order color and other elements in groups of graphs	Robbins, 2005	221
	Show data variation, not design variation	Tufte	61
General	Put major conclusions into graphical form. Make captions comprehensive and informative	Cleveland, 1994	54
	A large amount of quantitative info can be packed in a small region Present many numbers in small space	Robbins, 2005 Tufte	224 13
	Different graphs emphasize different aspects of the data; 2 or more times Reveal data at several levels of detail	Robbins, 2005 Tukey Tufte	224 157 13
	Don’t quote data out of context	Tufte	77
	Describe everything that is graphed Draw attention to the important features of the data Describe the conclusions that are drawn from the data on the graph.	Cleveland, 1994	55
	Visual clarity must be preserved under reduction and reproduction	Robbins, 2005	

Element	Detail	Author	Page
Glyphs	When developing glyphs, use small, closed shapes to represent data, and use the color, shape, and size of those glyphs to represent attributes of the data.	Ware, 2013	224
	Map variables to integral glyph properties if it is important for people to respond holistically to a combination of two variables in a set of glyphs,	Ware, 2013	165
	Glyph attributes of size, lightness (on a dark background), darkness (on a light background), vividness (higher saturation) of color, or vertical position in the display can all represent quantity	Ware, 2013	168
	Use glyph length or height, or vertical position, to represent quantity. If the range of values is large, consider using glyph area as an alternative. Never use the volume of a three-dimensional glyph to represent quantity.	Ware, 2013	169
Gridlines	Deemphasize grid lines Distinguish grid lines from data	Robbins, 2005	185
	Use light gray gridlines (when used at all) so as not to distract from the data lines;	Tufte, 2001	92
Legend/labels	Key outside scale-line rectangle	Robbins, 2005 Cleveland, 1994	189 47
	Label lines directly, avoid legends and keys	Robbins, 2005	213
	Label data sets directly unless too much clutter. Clear, detailed, and thorough labeling defeats graphical distortion and ambiguity. Data labels should not interfere with the quantitative data or clutter the graph.	Robbins, 2005 Ware, 2013 Tufte Cleveland, 1994	213 321 56, 77 43

Element	Detail	Author	Page
Markers	Use filled circles instead of different letters or shapes.	Cleveland & McGill, 1984	
	Identifying and tracking a referent may be aided by the use of symbols already associated with the referent or value such as a glyph	Carpenter & Shaw	97
	Overlapping plotting symbols must be visually distinguishable Show the data, data points should not overlap Express *x* and *y* to avoid concealment by crowding	Robbins, 2005 Cleveland, 1994 Robbins, 2005 Tukey	165 50 154 157
Notes	Notes in caption or in the text	Robbins, 2005 Cleveland, 1994	189 47
Proximity	Place symbols and glyphs representing related information close together.	Ware, 2013	182
Reference line	Use a reference line to show an important value that must be seen across the entire graph, but do not let the line interfere with the data	Cleveland, 1994	42
Scales	Draw data to scale Representation of numbers should be directly proportional to the quantities represented	Robbins, 2005 Tufte	197, 155 56, 77
	Scales should be proportional to the numerical quantities represented	Tufte, 2001	92
	Zero baseline for bar charts, others shouldn't mislead	Robbins, 2005	240
	Use same scale on different panels comparing data	Robbins, 2005	289
	Do not insist that zero be included when showing magnitude	Cleveland, 1994	92
	Scale breaks only if can't be avoided. Use a full scale break. Taking log might cure the need Don't connect numerical values on two sides of a break.	Robbins, 2005 Cleveland, 1984	257, 265 2
	Logarithmic scales when need to understand percent change or multiplicative factors	Robbins, 2005 Tukey	243 157 95

Element	Detail	Author	Page
		Cleveland, 1994	
	Showing data on logarithmic scale can cure skewness toward large values	Cleveland, 1994	103
	2 Y scales for one axis if helps i.e. Fahrenheit and Celsius	Robbins, 2005	263
	All axes require scales Use common baseline to compare data sets	Robbins, 2005	277 207
	Choose scales so that the data rectangle fills up as much of the scale-line rectangle as possible	Robbins, 2005 Cleveland	285 81
Similarity	Standardize graphical symbol systems within and across applications.	Ware, 2013	17
	Use symmetry to make pattern comparisons easier. Symmetrical relations should be arranged on horizontal or vertical axes unless some framing pattern is used.	Ware, 2013	185
	Define multiple overlapping regions with a combination of line contour, color, texture.	Ware, 2013	188
	Standardize the use of visual patterns within and across applications.	Ware, 2013	220
Split Attention	Comparisons between text and graphic or between two graphics must occur within the same eye span on a page;	Tufte, 2001	92
Symbols	Use color, form, and motion to display aspects of data so that they are visually distinct.	Ware, 2013	145
	Make symbols as distinct from each other as possible, in terms of both their spatial frequency components and their orientations components.	Ware, 2013	151
	Use strong preattentive cues (color, orientation, size, contrast, motion) before weak ones (line curvature) to facilitate search.	Ware, 2013	156

Element	Detail	Author	Page
	For maximum pop out, a symbol should have a distinct feature such as being the only colored item in black and white display.	Ware, 2013	157
	Make symbols differ in both shape and color (redundant coding) to maximize distinctiveness.	Ware, 2013	159
	If symbols are to be pre-attentively distinct, avoid coding that uses conjunctions of basic graphical properties.	Ware, 2013	160
	Use symbols instead of words or icons to represent a large number of data points.	Ware, 2013	321
	To highlight two distinct attributes of a set of entities, code using different properties.	Ware, 2013	161
Textures	Make nominal coding textures distinctively different in terms textural spacing and orientation components. Make texture elements vary in the randomness of their spacing.	Ware, 2013	205
	Use simple texture parameters, such as pattern size or density, only when fewer than five ordinal steps must be reliably distinguished.	Ware, 2013	206
	Design textures so that quantitative values can be reliably judged, by using a sequence of textures that are both visually ordered (for example, by element size or density) and distinct in some low-level property.	Ware, 2013	209
	When using overlapping textures to separate overlapping regions in a display, avoid patterns that seem to move when they are combined.	Ware, 2013	212
	When using textures to overlap regions of color, use open patterns so the other data are visible through the gaps.	Ware, 2013	. 212

Element	Detail	Author	Page
	Ensure high contrast between texture elements in the foreground and color-coded data in the background when using open textures overlapping colored regions.	Ware, 2013	213
Tick Marks	Point outward	Robbins, 2005	179
		Cleveland,	39
	Limit quantity	1994	154
	Use sensible values	Robbins, 2005	
	Don’t use equally spaced tick marks		281
	for uneven intervals	Robbins, 2005	287
	Tick marks should include the data		80
	range	Cleveland,	
		1994	

Appendix F – APA Table Construction Guidelines

Guideline	Page
No vertical lines	140
Place items that are to be compared next to each other	107
Place labels so that they clearly connect with the elements they are labeling	107
Use fonts that are large enough to read	107
Color only if it is crucial for understanding	109
Different indices should be segregated into different parts or lines of tables	110
Arrange data so its meaning is obvious	110
Table should have a title. Title should be brief,, clear, explanatory	116
Headings identify columns. Should not be many more characters in length than the widest entry. Standard abbreviations can be used with no explanation. Non-standard explained parenthetically following entries in stub (leftmost) column.	117
Each column needs a heading, including the stub	118
Indent subordinates in stub column instead of different column	118
Use variable names (not numbers) in stub.	120
Headings identify what's under, not next to them	120
Column heads span one column, column spanners span 2 or more columns with their own heads. Decked headings have spanner above heads. If possible, not more than two levels in decked headings.	120
Table spanners span the body of the table to divide the table.	121
Parallel syntax for items within a column.	121
Use singular unless referring to groups	121
Carry comparable values to same level of decimal if possible	122
Dash replaces 1 on diagonal of correlation matrix	123
Leave blank cells blank. If data was missing use a dash and explain the dash.	123
Display confidence intervals with point estimates (means, correlations, regression slopes). Use brackets or separate columns for upper limit and lower limit.	123
Don't include columns of data that can be easily calculated from other columns.	123
General notes qualify, explain or provide information. Ends with explanation of abbreviations, and symbols. Acknowledgements that the table is reproduced from another source. Designated by the word *Note* followed by a period	125

Specific notes refer to a particular column, row or cell. They are indicated by superscript lowercase letters. Within headings and table body, order the superscripts from left to right, top to bottom. Begin note's footnote with the superscript lowercase letter.	125
Probability note indicates how asterisks and other symbols are used to indicate *p* values	126
Report exact probability to two or three decimal places. Can use "$p <$" if exact probabilities make the table "unruly"	126
Order of notes should be general note, specific note, and probability note.	127
Notes begin flush left on a new line below the table. Start each type on a new line, run each type together.	127
Limit lines to those necessary for clarity	127
White space can substitute for rules (lines)	127
Single or double space. Consider legibility	128

Source: Concise rules of APA style, 2011

Appendix G – Recruitment materials

Protocol Analysis Recruitment email

Dear [*name*],

My name is Holly Roof and I am a PhD candidate from the Research Methods and Statistics Department at the University of Denver. I am writing to invite you to participate in my mixed methods research study with the working title: Is a Picture Worth a Thousand Numbers? Effects of Data Display Format on Memory for Research Results. You're eligible to be in this study because you are an advanced graduate student in the RMS program. I obtained your contact information from Dr. Nicholas Cutforth, Professor and Department Chair Research Methods and Information Science Morgridge College of Education.

If you decide to volunteer for this study, you will participate in a protocol analysis of your interpretation of visual presentation of research results. Using the "think aloud method" you will be asked to articulate your thought processes while viewing a short slideshow. A brief training period will be provided. The entire process will take approximately 15 minutes of your time. I would like to audio record your verbalizations and then use the information to gain a greater understanding of the process by which individuals with statistical training interpret research presentations. As a token of my appreciation, you will receive a $5 gift card to a local coffee shop.

Your participation is completely voluntary. If you would like to participate or have any questions about the study, please email me at holly.roof@du.edu or contact me at 970-376-7565.

Thank you very much.

Sincerely,

Holly Roof

Experimental Component Recruitment Email

Dear [Instructor of Record Name]

As you know, I am a PhD candidate in the Research Methods and Statistics Department at the University of Denver. My dissertation is a mixed methods research study with the working title: Is a Picture Worth a Thousand Numbers? Effects of Data Display Format on Memory for Research Results. The experimental component of the study involves comparing viewers' initial understanding and recall of research results with different presentation formats. I am writing to request access to your Analytics [I or III] class as participants for my study.

Your class will watch one of two PowerPoint presentations of a quantitative research report based on 2016 Colorado Health Outcomes. Immediately following the presentation, participants will take a short assessment to gauge their understanding of the material. At this time they will be asked demographic questions as well. This should take about 20 minutes of class time. Approximately two weeks later, I would like to administer the assessment a second time to gauge your students' recall of the information presented. This will take about 10 minutes. Your students will have the option to opt out of participation.

Would it be okay for me to take 20 minutes of class time next week and 10 minutes of class time in two weeks to conduct my experiment? If you agree what days and times are best for you? If you have any questions about the study, please email me at holly.roof@du.edu or contact me at 970-376-7565.

Thank you very much,

Sincerely,

Holly Roof

Appendix H – Introductory Material

Experiment Introduction

Hello – My name is Holly Roof and I am a PhD candidate from the Research Methods and Statistics Department at DU. I'm here to talk to you about participating in my research study. This is a study about the effects of data display format on memory for research results. You are eligible to be in this study because you are in this class. Your teacher has given me permission to be here.

If you decide to participate in this study, you will watch a short video presenting the results of a research study, then you will take a short quiz about the presented material and answer a few questions about yourself. In a couple of weeks, you will take another quiz about the same material. The results of the quizzes will be used to compare effects on understanding and retention of different display formats. Any personal information will be kept confidential.

Your participation is completely voluntary and will not affect your grade in this class. If you choose not to participate you may remain in the classroom or return at _________ o'clock.

Does anyone have any questions at this time?

If you have any more questions about this process or if you need to contact me about participation, I may be reached at [write email address on the board].

I'm passing out a consent form for this research. Take as much time as you need to read the document. If you choose to participate, please sign this consent form.

Thank you so much,

Protocol Analysis Introduction

Hello – My name is Holly Roof. Nice to meet you. Thank you for agreeing to participate in my dissertation research. It's a mixed-methods study about the effects of data display format on memory for research results. For the qualitative portion, I am interested in the thought processes people use as they interpret displays of statistical information.

If you decide to participate in this study, you will watch a short PowerPoint slide show that presents the results of a research study. As you watch the slide show, you will be asked to keep talking out loud as you attempt to interpret the material on each slide. I will audio record you speaking and transcribe the data later. Also, you will need to answer a few questions about yourself. We will go through a practice exercise first to warm up and become comfortable with the procedure

Any personal information, as well as all data and transcripts will be kept confidential.

Your participation is completely voluntary and will not affect your grade in any class.

Do you have any questions?

This is a consent form for this research. Take as much time as you need to read the document. If you choose to participate, please sign this consent form.

Thank you so much,

Okay, let's get started.

Protocol Analysis Instructions

This method is called Think aloud. In a moment I'll turn on a PowerPoint slide show. The slides are timed to mimic an actual presentation. You are asked to interpret each slide in the way you would if you were watching a research presentation. While you do so, try to say everything that goes through your mind.

Appendix I – Assessment questions

1. The presentation is about health outcomes and factors in:
   - ☐ United States
   - ☐ Denver
   - ☐ Colorado
   - ☐ Front Range Communities

2. The variables used to rank Colorado counties health outcomes are:
   - ☐ Health behaviors, Clinical Care, Social & Economic Factors, Physical Environment
   - ☐ Health behaviors, Premature Death, and Physical Environment
   - ☐ Clinical care, Premature Death, Poor or Fair Health, and Low Birthweight
   - ☐ Premature death, Poor or Fair Health, Poor Physical Health Days, Poor Mental Health Days, and Low Birthweight

3. Compared to all U.S. counties, do Colorado counties experience greater or fewer average years of potential life lost due to premature death?
   - ☐ Greater years of potential life lost
   - ☐ Fewer years of potential life lost

4. Compared to all U.S. counties, do people in Colorado report more or less poor or fair health days?
   - ☐ More poor or fair health days
   - ☐ Less poor or fair health days

5. Compared to all U.S. counties, are more or less low birthweight babies born in Colorado?
   - ☐ More low-birthweight babies are born in Colorado
   - ☐ Fewer low-birthweight babies are born in Colorado

6. Low birthweight is defined as a birthweight less than 4.75 pounds?
   - ☐ True
   - ☐ False

7. The average number of physically and mentally unhealthy days reported by Colorado residents is:
   - ☐ Less than 3
   - ☐ Between 3 and 3.5
   - ☐ Between 3.5 and 4
   - ☐ Greater than 4

8. Colorado residents report a wider range of physically unhealthy days than mentally unhealthy days.
   - ☐ True
   - ☐ False

9. What areas of Colorado are characterized by poorer health outcomes? Choose all that apply.
   - ☐ Central mountains
   - ☐ Eastern plains
   - ☐ Urban areas
   - ☐ Southern counties
   - ☐ Western counties

10. What areas of Colorado are characterized by better health outcomes? Choose all that apply.
    - ☐ Central mountains
    - ☐ Eastern plains
    - ☐ Urban areas
    - ☐ Southern counties
    - ☐ Western counties

11. Which clusters have more poor health outcomes?
    - ☐ Good & Best
    - ☐ Worst & Poor
    - ☐ Good & Poor
    - ☐ Best & Worst

12. Which clusters have more low-birthweight babies?
    - ☐ Good & Best
    - ☐ Worst & Poor
    - ☐ Best & Poor
    - ☐ Good & Worst

13. Which health outcome variable shows the weakest correlations with the other health outcomes?
    - ☐ Premature death
    - ☐ Poor physical health days
    - ☐ Low birthweight babies
    - ☐ Poor mental health days
    - ☐ Poor or fair health days

14. Regression results indicate that health outcome rank can be predicted with just three health factors. What are the three health factors?

- ☐ Adult Smoking, Physical Inactivity, Teen Births
- ☐ Air pollution, Violent Crime, Adult Obesity
- ☐ Severe housing problems, Unemployment, Income inequality,
- ☐ Unemployment, Children in Poverty, Adult Obesity
- ☐ Children in Poverty, Adult Smoking, Unemployment

15. The degree of accuracy with which the statistically significant regression model predicts health outcome rank is measured with R^2. What percent of variation is explained by the model?

    R^2 = ________

16. Correlations between the independent variables (health factors) and the dependent variable (health outcome rank) indicate that as the health factors rise a county's health outcome rank rises.
    - ☐ True
    - ☐ False

17. Of the three health factors, which has the lowest average rate?
    - ☐ Adult obesity rate
    - ☐ Percent of children in poverty
    - ☐ Unemployment rate

18. Of the three health factors, which has the most variability?
    - ☐ Adult obesity rate
    - ☐ Percent of children in poverty
    - ☐ Unemployment rate

19. Which of the three health factors has the greatest impact on a county's health outcome ranking?
    - ☐ Adult obesity rate
    - ☐ Percent of children in poverty
    - ☐ Unemployment rate

20. Which health factor has the narrowest confidence interval?
    - ☐ Adult obesity rate
    - ☐ Percent of children in poverty
    - ☐ Unemployment rate

21. Which Colorado County had the best health outcome rank in 2016?

    ____

Appendix J – Demographic Questionnaire

About You

Participant Characteristics
Age ____________

Sex

☐ Male ☐ Female

Student Status

☐ Undergraduate ☐ Graduate

Program ____________________________

Number of quarters you have completed towards your degree? _____________________

How many statistics courses have you completed? ________________

In what nation did you earn your secondary (high school) education?

____

My attitude toward statistics is: (circle one)

Very favorable	Somewhat favorable	Neutral	Somewhat Unfavorable	Very unfavorable

My attitude toward empirically based research is: (circle one)

Very favorable	Somewhat favorable	Neutral	Somewhat Unfavorable	Very unfavorable

I am a Colorado resident

☐ Yes ☐ No

My level of interest in Colorado health issues is: (circle one)

Very interested	Somewhat interested	Neutral	Somewhat uninterested	Very uninterested

The number of PowerPoint presentations I have viewed in past month is approximately

- ☐ None
- ☐ 1 - 5
- ☐ 5 – 10
- ☐ Too many to count

The number of Power Point presentations that I have created is approximately

- ☐ None
- ☐ 1 - 5

- ☐ 5 – 10
- ☐ Too many to count

Appendix K – Protocol Analysis Slide-by-slide Results

Images of all slide pairs can be found in Appendix D

Slide 1 – This slide is the same in both presentations and was classified as easy. It includes the title of the presentation and a picture of a "Welcome to Colorado" highway border sign.

Only three participants said anything. All three read the title. Comments included: "I like the pictures on the first slide" and "Now I know what to expect."

Slide 2 – This slide is also the same in both presentations and classified as moderate. It presents the types of data in the analysis as health outcomes and health factor categories in lists that reveal one item after another.

All participants read the title and most read list items as they were revealed. Two questioned if the list items would be used to measure effect (P2) or as predictors (P4).

Slide 3 – This slide is classed as easy and presents a comparison of a single data point, potential years of life lost due to premature death for the United States and the state of Colorado. The graph shows the US flag and the Colorado flag atop lines (lollipop style) representing the value of the data point. All three participant who saw the tables presentation said "okay" after reading the title. Whereas one participant who saw the graphs presentation criticized the title as wordy and another said "Colorado flag clearly indicates hundred" and the third said "interesting".

Slide 4 – Moderate. This slide displays a comparison of two outcomes, poor or fair health and low birthweight, for the US and Colorado. Participants who saw the graphs presentation seemed to read the title and immediately understand the side-by-side column charts. Two commented that the slide was "interesting", P6 said "so that's pretty clear" and "the color coding is very clear what I'm looking at" and P4 said "I like how the US is color-coded to match the colors on the bar chart so that's clear for me."

On the other hand, one participant who saw the table presentation noted that the slide was "interesting" but then re-read the slide title and signaled an attempt to process the information saying "Let's see…" This participant looked for effect size and P2 asked "are we comparing the mean" with the total average in the states? P2 and P1 ultimately related information in this slide to previously held knowledge about Colorado with P6 saying, "that makes sense because Colorado has all this hiking" and P1 concluding, "but we already know that a lot of the low birth weight is because of the altitude, it's not because of anything else."

Slide 5 – Moderate. This pair of slides compares mean reported physically unhealthy days with mean reported mentally unhealthy days in Colorado. Standard deviations are included in both presentations.

All participants except one read the slide title. Two of the participants that saw the table presentation, seemed to have unanswered questions about the material with P3 saying "So we don't know if those unhealthy days are above or below the average", and P1 saying "okay so you're looking at standard deviation units but they don't tell you…" In contrast, P4 said "I see more standard deviations in the physical" upon inspecting the graphs. And P5 was able to quickly interpret the graph saying "so I'm seeing 3.29 physical 3.18 mental range."

Slide 6 – Easy. This slide pair shows results of a cluster analysis that groups Colorado counties into four outcomes: best, good, poor, and worst. The graphs presentation slide shows a map of Colorado with counties color-coded by cluster with additional information and a key to color coding included in the title. The information that was provided in the title on the graphs slide was provided in a note on the tables slide per APA guidelines (Concise rules of APA style, 2011).

P2, who saw the tables presentations, concluded that clustering was based on previously presented outcome metrics and twice said the information presented was "interesting" while P1 complained "This cluster membership is way too many words on one slide I have no idea what I'm supposed to get from this" and commented "I don't get much information just from looking at all these words on a slide." This participant also questioned if county demographics played a role in the clustering.

In contrast, P6, who saw the graphs presentation, noted "it's color coded so it's pretty easy to understand" and "it's well labeled, all the counties are in there, it doesn't look too busy or cluttered" and "they managed to get everything in one big graphic without it looking messy. I like this one. I think it's really easy to interpret." As for demographics this participant noted, "it seems like the middle of the state has better outcomes than the, you know, east, west and south and they seem to be the more rural areas" and concluded "being in Denver or more populated areas means you probably have access to better hospitals, more clinics that sort of thing". This was echoed by P4 who said "I see a clearly color coded map of all the counties" and "the clusters of best and good are mostly in the metro area and the central Colorado area" while noting an "interesting" juxtaposition of counties in the southwest corner of the state. P5, who also saw the color-coded map, commented three times that something in the slide was "interesting".

Slide 7 – difficult. This slide presented normalized scores for the five output measures by cluster – a total of twenty different data points. The table had a heading that said "normalized average outcome"; the graph did not have a heading. While the title signaled the comparisons being made, the word 'normalized' was not included on the graph slide. The graph showed bars stretching to the left or right of a line that represented the average or a z-score of zero. The information about the quantities represented on both the table slide and the graph slide was provided in the audio for both presentations which the think aloud participants did not hear. This slide proved to be the most difficult for participants to interpret regardless of presentation.

The participants that saw the tables presentation all read the slide title then all three searched for clarification of the metric and found the word "normalized" in the table title. Once they understood the metric, they made direct comparisons of the numbers provided in the table to confirm the statements made in the slide title.

All three participants that saw the graphs presentation expressed some degree of confusion. Saying "so this one is a little bit more challenging to interpret cause you really have to read the text to understand it" (P6) and "this one is a little more confusing to me to interpret" (P4). P5 noted that she "paid attention to the color scheme (it's consistent) which actually helps."

Slide 8 – difficult. This slide pair shows correlations for the five outcome variables including statistical significance and population size (upper case *N* was used per

APA standards to denote population size rather than *n* to denote sample size) (Concise rules of APA style, 2011). The graph slide was arranged in a grid of scatter plots with dots color coded to the different clusters providing some additional information over the simple correlation table in the other presentation.

Of the tables group, one person did not read the title. Participants all confirmed correlation values were positive with one proclaiming the relationships were "a no-brainer".

The graphs group, all three read the title which said "health outcomes show positive correlations across clusters". Upon looking at the graph, P6 exclaimed "it sure does". All three participants checked for statistically significant relationships and all three said "interesting" to some element of the slide.

Slide 9 – Moderate. This slide pair presents results of regression modeling that indicates outcome rank can be predicted by three health factors. The slides include R^2 and *p* values.

Tables participants: all three read the title, then proceeded to parse the ANOVA table to determine what predictors were used (P2), what the effect size was (P3), and what sample size was used (P1). None of these participants expressed interest in the content.

Graphs participants: Two participants who saw the graphs presentation found this slide "interesting" with P4 saying "I like this visual showing how that these predictors are driving up the, well, driving down the health outcomes rank". P6 pointed out that a general audience would probably need an explanation of R^2.

Slide 10 – difficult. These slides show negative correlations between outcome rank and the three health factor predictors along with means, standard deviations, population size, and statistical significance. Again, P3 was the only participant that did not read the title aloud.

Tables: All read the title, two went immediately to interpreting the title, for example, confirming that correlations were negative, looking for means, and sample size. P1 jumped immediately to the critique that a slide with correlations between predictors and the dependent variable should appear before the regression results and were, in a sense, redundant to those results.

Graphs: All three read the title, and proceeded to verify its statements. In contrast to the tables group, these participants also drew comparisons between the different health factors such as: "we've got the standard deviation lines here as well so there is a lot of deviation with the children in poverty one" (P6), "So the mean 21, 20, and 5" and "This is much more dispersed" (P5).

Slide 11 – difficult. These slides present unstandardized coefficients and confidence intervals from the regression analysis. The table shows standard errors, *t*- and *p*-values while the graph includes a statement about statistical significance.

Tables: P2 and P3 read the title and, as before, parsed specifics provided in the table such as standardized versus unstandardized coefficients (P3), and the meaning of the *t* statistic. P1 went straight to the criticism that these tables are "are always shown" but "it's just more information, it's not telling you how this is useful or helpful or what the next steps are."

Graphs: All three read the title. P6 had trouble locating labels for the predictors and ran out of time to reach understanding. P4 and P5 focused on specific beta values and the impact those values have on outcome rank. P5 noticed relative differences between predictors.

Slide 12 – easy. These slides present the top and bottom ranked counties in Colorado in terms of health outcome. The table lists county name and rank in labeled columns. The graph shows the entire state with the top and bottom ranked counties filled with the color of their cluster.

Tables: In spite of column labels and sparse text, participants needed information that was not provided. For example, P2 said "Okay, that's the ranking I don't know how many counties we have, or I forget, in Colorado" and P3 laughed "I don't know where the heur? Heurfano is that's the worst county in the health, okay, assessment".

Graphs: Two participants concluded that they were not surprised that Douglas was ranked first because of its affluent, h nature. The third (P5) tried to figure out how to pronounce the name of the lowest ranked county, Heurfano.

Slide 13. This slide was a black slide with white lettering with text asking participants to complete their assessment packets. It was not included in this analysis.