Measuring the Influence of Alternative Treemap Visualizations on Human Judgment Tasks

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Akilah L. Hugine-Elmore

May

2013

APPROVAL SHEET

The dissertation

is submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

AUTHOR

The dissertation has been read and approved by the examining committee:

Stephanie Guerlain, Ph.D.

Advisor

James H. Lambert, Ph.D.

William T. Scherer, Ph.D.

Larry G. Richards, Ph.D.

Florence E. Turrentine, Ph.D.

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

May

2013

ABSTRACT

Visualizations of large data sets provide insight to features of the data, improve the accuracy of mental models of the information, and locate data regions of particular interest. A visualization technique that has gained popularity is the *treemap*. The treemap uses a recursive algorithm to display a hierarchical data set in the form of nested rectangles of varying size, orientation, aspect ratio, relative placement, and color to represent selected aspects of the data. This dissertation extends previous work on human perception and rectangular relative area judgments to determine how these display features affect judgments that are supported by treemap visualizations. The effort includes a novel treemap generation algorithm, which utilizes lexicographical order theory to generate treemaps that group like data elements together, to support human gestalt perception capabilities and the results of an experiment comparing this alternative approach to the current, squarified algorithm that is a current practice. A final experiment explores the use of treemaps for supporting decision making in a specific application area -- review and interpretation of surgical quality data -- and further characterizes the performance of decision makers to correctly interpret data displayed in alternative treemap formats. The combined effort thus improves on current knowledge of the capacities of individuals to make relative and proportional judgments in hierarchical visualizations, with an innovative method for visualizing hierarchically structured information in treemaps.

ACKNOWLEDGEMENTS

"Thanks be to God, who in Christ always leads us in triumph…" (II Corinthians 2: 14)

The work was supported by Grant Number T15LM009462 from the National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.

First, I would like to thank my advisor Dr. Stephanie Guerlain for her guidance and support throughout this research. She has always motivated me to strive to do my best and her advice during my time here at UVA was invaluable. She has certainly made a lasting impression on me and I will be forever grateful to have had the opportunity to work with her. Additionally, I would like to thank my advisor Dr. James Lambert for his willingness to serve as my advisor during a transitional period. I am thankful for his contributions and insights towards my dissertation research. I am extremely grateful for his unwavering support and guidance.

I sincerely thank Dr. William Scherer, Dr. Larry Richards, and Dr. Beth Turrentine for their willingness to serve on my committee and for the helpful feedback they provided on my dissertation research. Appreciation is also extended to Dr. Turrentine for her medical expertise and assistance with the surgery data that contributed to the work in Chapter 6. I would also like to thank Dr. Ellen Bass for her help with the experimental design, writeup, and data analysis related to the work in Chapter 4.

I also benefited from several research activities while at UVA that are not part of the main dissertation effort, these activities and related publications are summarized in Appendix G.

Special thanks are expressed to my colleagues that provided encouragement and gave valuable feedback towards my research. There are too many to list, but my sincere gratitude goes to Don Rude, Leigh Baumgart, and Jenny Prey.

Finally, I would like to thank my family and friends for their continued support and encouragement during this effort and also life. I would especially like to thank my parents Andrew and Abbiegail Hugine for giving me much needed advice and immense strength and support during my time here at UVA. They are truly an inspiration. Last, but not least I would to thank my husband Quincy Degen Elmore. There are no words to express how much I appreciate everything he has done for me. I would not have been able to finish this journey without his love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

Large, complex datasets have some of the following properties: A large number of records, many variables, complex data structures, and intricate patterns and dependencies in the data that require complex models and methods of analysis. Analysts are faced with the task of interpreting these large data sets for exploration to identify correlations and meaningful trends. These enormous datasets are classified as "big data", which is defined as massive amounts of data collected over time that are difficult to analyze and handle using common database management tools [1]. Big Data includes business transactions, healthcare information, surveillance videos, and activity logs [2]. The issue of big data has come to the forefront of many research organizations. In particular the National Science Foundation (NSF) and the National Institutes of Health (NIH) launched a “big data” research program that aimed to:

> “Advance the core scientific and technological means of managing, analyzing, visualizing and extracting information from large, diverse, distributed, and heterogeneous data sets in order to accelerate progress in science and engineering research. [3]”

As a result of these “big data” analytics initiatives, the demand for data visualization tools is rising sharply. Visualizing these large data sets can help users gain insight into relevant features of the data, locate regions of particular interest, construct accurate mental models of the information, and leverage parallelism in visual processing. Data visualizations range from basic charts, such as line, bar, area and pie charts, status indicators for sets of data, to more complex visualizations, such as, scatter graphs, bubble charts, and spark line charts. Data visualization is only successful to the degree that it encodes information in a manner that our eyes can discern and our brains can understand [4], which is more a science achieved by studying human perception and understanding how humans interpret data visualizations. The goal is to translate abstract information into visual representations that can be easily, efficiently, accurately, and meaningfully decoded.

To examine this notion further of human perception and accurate interpretation of data visualizations, we investigate a relatively new data visualization solution, called a treemap, that has been gaining popularity as a means to visualize large hierarchical data sets. Treemaps are space-constrained visualizations that display multidimensional data as sets of nested rectangles with area proportional to a specified dimension on the data [5] (Figure 1). The visualization combines features of multivariate coding and display layout to present hierarchies in a rich visual environment that fosters relative comparisons of structures in the hierarchy. One common layout employs a recursive "squarified" algorithm that avoids high aspect ratio rectangles and rotates the direction of each successive subdivision (horizontally and vertically) [6].

Researchers and analyst have used treemaps to visualize data sets in various domains: education and legal [7], stock market [8], air traffic flow management (ATFM) [9], and network traffic [10]. Figure 1 depicts a treemap visualization of the stock market through the Wall Street Journal website, SmartMoney [8]. Stocks are categorized by industry area (e.g. technology, health care, financial, and energy) and the size of each industry block illustrates the activity of stocks in these areas. The activity of each stock is indicated by rectangle size and the direction of change is shown through color (red=loss, green=gain).

While the treemap has utility in data analyses, several questions arise as to whether users of treemaps can purposefully and reliably interpret the information being displayed, particularly when comparing two specific nodes or comparing subsets of data. For example, when making singular comparisons where the area of two rectangles (length x width) is being used to represent a value of interest, the aspect ratio, distance, and offset angle of those rectangles can vary. Thus, it is possible that in one treemap, the same data values could be represented in multiple ways. Can people judge such varying examples as being the same area? Given a 2:3 rectangle and a 3:2 rectangle, can people accurately judge which one is bigger, and correctly estimate the difference in areas? Will such comparisons of the size of two rectangles be as accurate with, for example, a comparison between a 2:3 rectangle and a 1:1 square? Furthermore, are judgments affected when those rectangles are neither co-located nor aligned along the same horizontal, central, or vertical axis?

Figure 1. Treemap Visualization of Financial Data [8]

There are also human perceptual issues that arise when assessing proportional judgments of treemap. For instance in Figure 1, a shade of green represents an increase in stock value as opposed to a shade of red representing a decrease in stock value. With this visualization and coding, the user at a quick glance can see that many of the health care sector stocks increased for the day. However, how accurately can the user judge the percentage of health care stocks that increased for the day, compared to the percentage of energy stocks that increased for the day? The seemingly random placement of the nodes by size from the squarified treemap algorithm makes this proportional area judgment relatively difficult. The user potentially would have to cognitively group or count all of the green colored nodes together for stocks that had a gain and group the red colored nodes for stocks that had a loss for the day in order to come up with an assessment. Implications of human perception research suggest that utilizing alternative layout algorithms might help eliminate some of the information processing steps needed to make this type of judgment. An alternative algorithm could potentially group like nodes together utilizing Gestalt laws of perceptual organization [11]. The close proximity could be better suited for making relative magnitude comparisons among users.

In visualization, treemaps have value because they are able to display multiple variables in one display for a large number of data values. However, there are limitations with treemaps due to the perceptual nature needed to assess the visualization and we do not fully understand the precision with which people can make accurate assessments. Research has shown that humans generally have biases when making rectangular area judgments, which could potentially hinder the effectiveness of treemap visualizations [12] since area judging is a key aspect that is required when interpreting them. In order for a data visualization to be effective it must reinforce the human cognition of individuals who are interpreting the visualization for decision making.

The purpose of this dissertation is to characterize perceptual issues that affect the ability of individuals to interpret treemap visualizations and to develop and evaluate a new design feature of treemaps that could help with these interpretations by adding the ability to perceptually group like nodes together. The research focuses on relative area and proportional perception judgments. Relative judgments are ones that are made when there is an opportunity to compare two or more stimuli and judge the relative position of one compared with other along the same dimension, such as judging the area or magnitude between two figures. Insights from this research give a fundamental understanding of relative judgments in hierarchical data visualizations and in turn validate the effectiveness of these types of visualizations.

This dissertation is organized into six chapters, which is further conceptualized in the work flow diagram in Appendix A.

Chapter 1 provides the introduction, purpose, scope, and organization of the dissertation.

Chapter 2 provides a review of related literature. The chapter begins with a background into graphical perception, detailing common perceptual issues that occur when analyzing data graphically that affects human judgment abilities. The chapter ends with a review of data visualizations that are used to display hierarchical data.

Chapter 3 provides a review of treemap visualizations. The chapter begins with a conceptual background about treemap visualizations, including the algorithms and theories utilized to develop the visualization. Also covered are the shortcomings of existing treemap techniques, and the benefits of the treemap approach as well as its applicability.

Chapter 4 describes an experiment conducted as part of the dissertation to help us understand human perception issues related to making relative rectangular area judgments. The chapter explores the ability of decision makers to judge the percent difference in the area of two rectangles when the basic parameters that make up those two rectangles vary: relative placement, aspect ratio, and true percentage difference.

Chapter 5 introduces an alternative treemap algorithm. The algorithm is a recursive attribute grouping algorithm for treemaps that clusters like attributes together. The chapter also explores the ability of decision makers to make relative proportion judgments, i.e., how well individuals judge the proportion of rectangles within a set of nested rectangles that are of one color when the rest are in a second color. A human subjects experiment is discussed in detail that gauges the impact that the alternative attribute grouping display layout will have on interpreting proportional judgments.

Chapter 6 investigates the application of hierarchical data visualizations to surgical quality data to help clinicians make judgments about their patients. Surgery data from the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) database is mapped to both the squarified treemap layout and the alternative attribute grouping algorithm introduced in this dissertation to evaluate which layout produces better judgment accuracy and faster judgment times and further examines how well proportional judgments can be made with treemaps.

Chapter 7 presents the conclusions and summarizes the perceptual issues of humans making relative area and proportion judgments utilizing data visualizations.

Furthermore, the chapter summarizes the benefits and utility of the treemap concept for relative judgments of large hierarchical data sets. The chapter describes several future directions that have been suggested by this dissertation research.

CHAPTER 2. CONCEPTUAL BACKGROUND

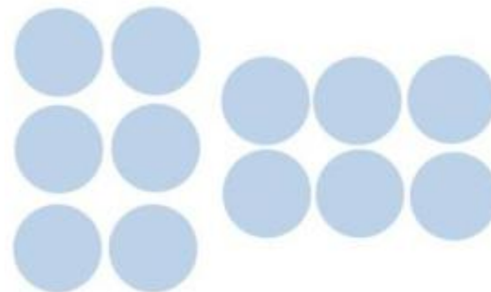
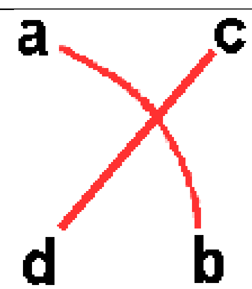
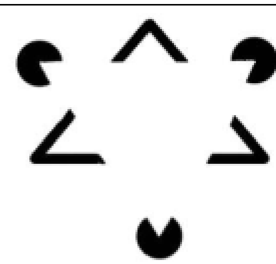
2.1 Human Perception of Data Visualizations

Data visualization is effective because it shifts the balance between perception and cognition to take full advantage of the pattern recognition abilities of the brain. Visual perception is extremely fast and efficient compared to cognition, which is much slower and less efficient for comprehension. Data visualization shifts the balance toward greater use of visual perception.

One of the earliest contributions to the science of perception was the Gestalt principles of perception [13]. The intent of the principles is to uncover how humans perceive pattern, form, and organization of visualizations. Table 1 shows a few of the principles that can inform data visualization efforts. For the law of proximity, objects that are close together are perceived as a group. For the law of similarity, objects that share similar attributes (e.g., color or shape) are perceived as a group. For the law of good continuation, objects that are aligned together or appear to be a continuation of one another are perceived as a group. Lastly, for the law of closure, open structures are perceived as closed or complete.

Wickens and Andre expanded on the human perception concept in terms of proximity, developing the proximity compatibly principle (PCP) to emphasize that there is a relationship between different types of displays and the way the information from the displays has to be used [14]. The principle asserts that tasks in which “close mental proximity” is required will be best served by more proximate (close) displays. PCP may then be understood as a set of principles that incorporates a variety of psychological mechanisms, such as attention, object perception, and working memory, to link the visual processing of display characteristics to the cognitive processing of decision task characteristics.

Table 1. Gestalt Principles of Visual Perception

Law of Proximity	
Law of Similarity	
Law of Good Continuation	
Law of Closure	

2.1.1 Graphical Perception

Researchers have investigated how visual variables such as position, length, area, shape, and color, impact the effectiveness of data visualizations [15]. Each variable plays a role in "graphical perception", which is the ability for the viewer to decode and interpret information displayed in a visualization. Graphical perception is a term coined by Cleveland and McGill who ran a set of experiments to determine the elementary perceptual processes subjects utilize when decoding different types of graphs (scatter plot, bar charts, pie charts, maps, stacked charts, and bubbles). Their work was an extension of Stevens Power Law [16], which served as a way to estimate magnitudes.

Stevens Power Law $\varphi = kS^n$

$\varphi = Pyschological\ measurement\ (perceived\ value)$

$S = Physical\ Measurement\ (magnitude\ of\ stimulus)$

$k = Constant$

$n = Power\ Index\ determined\ by\ the\ properties\ of\ the\ stimulus$

The exponent *n* is determined by the type of stimulus and the value of the number has an effect on how individuals perceive the stimulus. For instance if the stimulus is brightness (*n=0.5*), where $n < 1$, individuals will underestimate the perceived stimulus value. On the other hand if $n > 1$ (e.g. Heaviness (*n=1.1*)) people have a tendency to overestimate the perceived stimulus value. Stevens Power Law indicates a description of how people perceive values through a number of encodings, including length, area, and volume. It was found that on average people underestimate area with a range of n from 0.6-0.9 and are better at judging length with a range of n from 0.9-1.1.

Cleveland and McGill extended this work by investigating how individual perceptual tasks relate to each other. In other words which elementary perceptual tasks are easier or harder to judge? Cleveland and McGill conducted an experiment using 7 graphical encodings, and for each encoding subjects had to judge relative magnitudes. For each graph, subjects were asked to compare a standard and test stimuli, and to judge what percentage the test stimuli was of the standard. The results of the experiment were assessed using log absolute error measure of accuracy:

$$error = \log_2(|judged\ percent - true\ percent| + \tfrac{1}{8})$$

From the results it was found that encoding values by angle, slope, or area produces significantly larger errors than position along aligned or non-aligned scales or length. Cleveland and McGill produced an ordering of perception task from easiest to hardest (Figure 2). From the ordering it is inferred that when developing a data visualization, it is preferable to use an encoding scheme which leads to a decoding task as high up the Cleveland and McGill ordering as possible.

Accuracy

1. Position on a common scale
2. Position along identical, non-aligned scales
3. Length
4. Angle
5. Area
6. Volume
7. Color (hue, saturation, brightness)

Figure 2. Cleveland and McGill Perception Ordering

In order to decode a treemap a number of these perceptual tasks are needed from Cleveland and McGill list. The tasks most relevant to treemaps are: position of non-aligned scales, length, area, shading, and color saturation. Two of the most important elements for treemap interpretation (area and color) are ranked low for accuracy on the "elementary perceptual task" list.

2.1.2 Perception of Simple Figures

There have been numerous research studies that have investigated a decision makers ability to judge spatial properties of simple figures [17-19]. Regan & Hamstra [20] conducted an experiment to measure the accuracy with which subjects judge that a square rectangle is perfectly symmetrical. In other words, that the aspect ratio (a/b) is equal (a and b are vertical and horizontal dimensions, respectively). The authors used a two alternative forced-choice paradigm and subjects were required to discriminate between a reference stimulus and a test stimulus. According to Regan & Hamstra, in the visual pathway there are aspect ratio-sensitive neurons, functionally organized into two pools. One of the pools is sensitive to the vertical extent of a stimulus, while the other is to the horizontal one. Regan & Hamstra found that observers are highly sensitive to small deviations from perfect symmetry.

Morgan [21] extended the research of Regan & Hamstra by making the hypothesis that noisy estimates of width and height are in fact the basis of judgments about aspect ratio as well as area. Morgan required subjects to compare the areas of two shapes with randomly-differing widths and heights. One shape, the standard, always had the same area, but its width was a random variable. The test had a different width and area from the standard. The observer had to decide whether the test shape had a larger or smaller area than the standard. Morgan reported that accuracy of area judgments tends to be higher on trials where height and width of the comparison and test stimuli differ in the same direction, rather than in the opposite direction. He found that area discrimination is consistently worse than aspect ratio or height discrimination. For ellipses, accuracy for aspect ratio was higher than predicted by the combination of noisy width and height signals; for rectangles it was worse. He concluded that observers have no access to high-precision codes for 2-D area and that they base their decisions on a variety of heuristics derived from 1-D codes. One is that observers use a variety of heuristics for combining width and height estimates into an estimate of area (i.e. If the width is greater but the height smaller, a decision could be reached by deciding which difference from the standard is greater).

Nachmias [22] disproved Morgan's hypothesis by conducting a study measuring discrimination for height, aspect ratio, and area of ovals and rectangles. Subjects were given a test and stimulus and asked to judge one of the three properties. Random jittering of the orthogonal property (width, aspect ratio, and area) in the figure shown to the participant was used to control the criterion of the observer (Figure 3). He found that judgments about aspect ratio are far more reliable than those about total area, which correspond with the conclusion reached by Regan & Hamstra. Nachmias stated that we can clearly reject the hypothesis that people judge aspect ratio by using linear combinations of noisy estimates of height and width, since the Weber fraction for aspect ratio are lower than predicted by the hypothesis.

Figure 3. Standard Stimuli Used in Experiments: Black Square in the Center of a Circular Gray Window and a Black Oval in the Center of a Square Gray Window [22]

2.2 Hierarchical Data Visualizations

One of the most challenging types of data to convert into a chart or visualization is also one of the most common: multi-level or hierarchical data. Hierarchical data is usually associated with categorical data. A categorical variable describes a particular quality or characteristic. Categorical data consists of variables whose values comprise a set of discrete categories. Such data require different statistical and graphical methods from those commonly used for quantitative data. Categorical data dimensions appear in many real-world data sets, but few visualization methods exist that properly deal with them [23]. Datasets with many categories are often organized hierarchically: split one piece of information between several questions or ask the same question several times for cross-checking [24]. For instance, bank accounts are classified in several ways that will often involve hierarchical categorizations (Figure 4). Using these hierarchies for visualization is helpful for the analyst, because they provide a natural way of collecting and abstracting data.

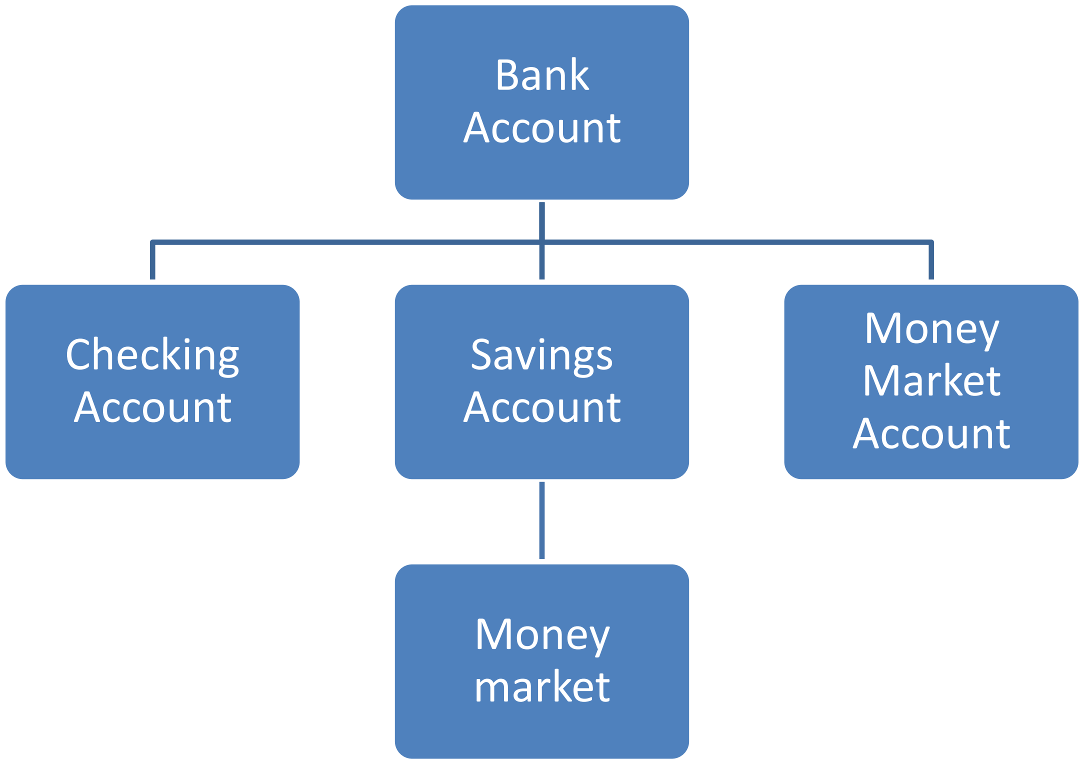


Figure 4. Classification Hierarchy of Bank Accounts

Interaction is also required, because the user will want to have the capability of switching between a more detailed investigation compared to a general search of the hierarchies. Data analysts would like to be able to explore data thoroughly, look for patterns and relationships, confirm or disprove the expected, and discover new phenomena. An important element of a successful graphical display is flexibility, both in tailoring the analysis to the structure of the data and in responding to patterns that successive steps of analysis uncover. For instance, a health care administrator may be interested in a general query such as how many patients received colorectal surgery compared to a physician who might be interested in assessing the patients they performed the surgery on, their complications, adverse events, or length of stay.

In these cases, every category of data is composed of sub-categories. Also, a change in one data point has a major effect on the surrounding data. With the inherent nature of hierarchical data there are certain limitations that arise when displaying this information. Singer & Feinstein presented the issues in displaying hierarchical data [23]. The issues are broken up into two categories focusing on visualizing the data and displaying the data. Visualizing the data refers to finding a way to actually display the data, identifying the best representation to infer the meaning of the data (i.e. displaying rank for nominal data). Displaying the data refers to the visual aspects of the data once displayed (i.e. arbitrary spacing on a bar chart), the output.

There is a need to have an effective visualization that makes hierarchical data intuitive and understandable. A number of researchers have developed various visual graphics to aid in displaying hierarchical data [25-30]. A few of these alternative displays are described in the sections below, broken up into three categories: traditional charts; relational data; and nested categories.

2.2.1 Traditional Charts

Traditional charts such as stacked bar charts and shapes allow for visualizing relational hierarchical data, but have limitations that affect the effectiveness of the display.

Stacked Bar

Probably the simplest and most familiar traditional chart types are stacked bar charts (Figure 5). Most common charting packages offer stacked bar charts, including Excel. At a glance this chart only allows two "levels" of depth, (parent category (Bar), and parent percentage (Inner bar)). Interactive versions of the stacked bar allow the user to click on a bar or bar component and "zoom" to see categories that make up that bar.

This chart is useful for comparing simple, broad categories with few constituents, and benefits from general familiarity, but fails to impart more than two levels of depth at a glance.

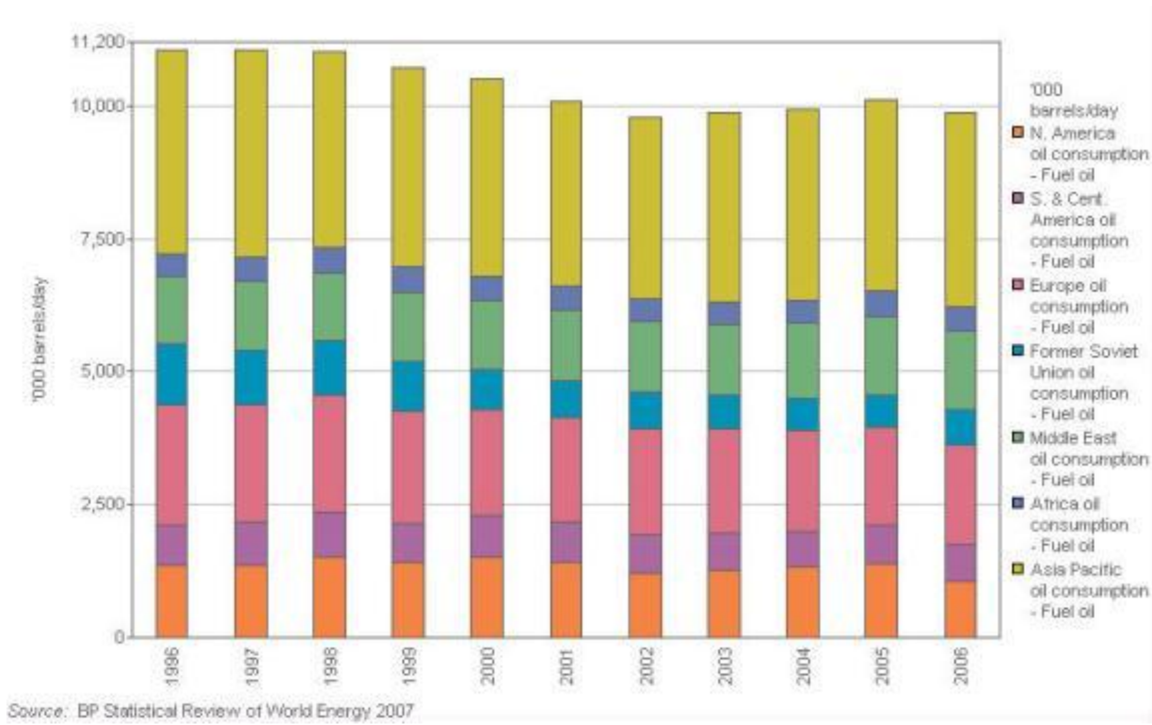


Figure 5. Stacked Bar Chart of World Fuel Consumption [23]

Stacked Areas/Shapes

Similar to stacked bar charts, an alternative chart type is stacked Area or Shape (Pyramid, Cylinder, etc) charts (Figure 6). Although these can be useful (particularly the pyramid chart, when one or two options in each "bar" take up an overwhelming percent of the total, but the lower percents must still be visible), they suffer from the pitfalls of stacked bar charts of only allowing only two levels of depth. Also, the added difficulty of perceiving the relative sizes of 3D objects is difficult for users to interpret. Often times relative proportions that are intuitive in rectangular dimensions ("*This* box is twice the area of *that* box") are lost in other shapes or in 3D ("How much more volume is in the *bottom half* of the pyramid than the *top half*?").

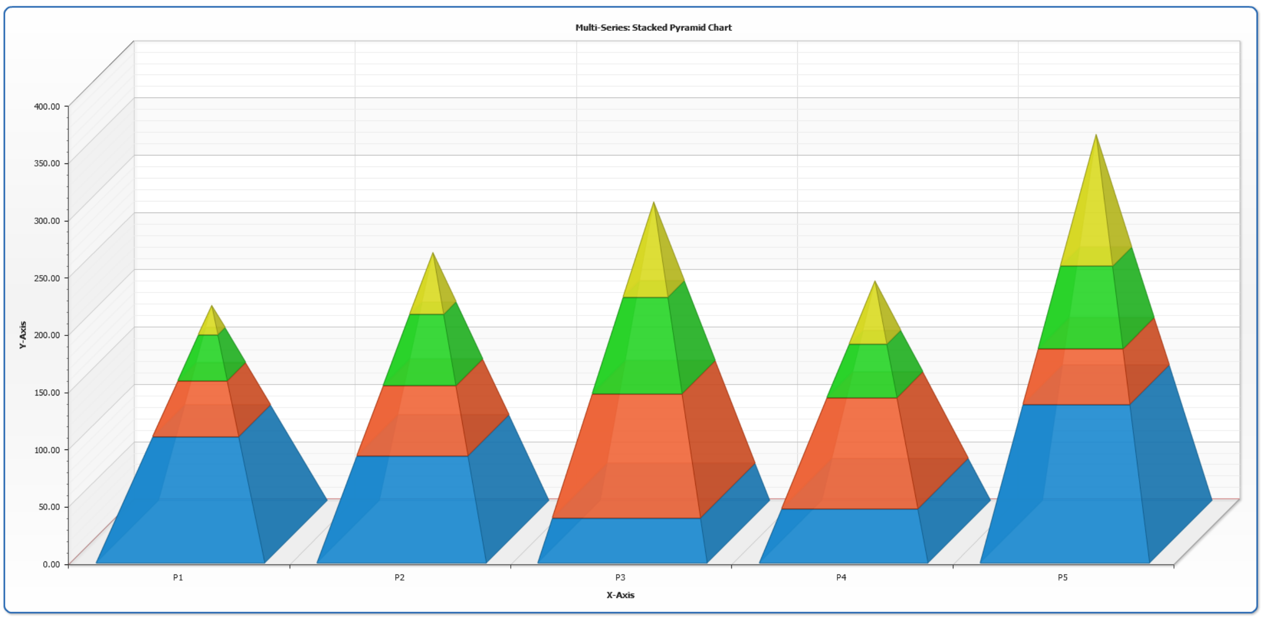


Figure 6. Multi-Series Stacked Pyramid Chart [23]

Dot Plots

Dot plot is an effective alternative to the bar graph data visualization in cases where the data contains groupings with similar values (Figure 7). The visualization is especially effective as a presentation tool to help people understand comparisons between grouped hierarchies of data. Menon and Nerella thought that dot plots are a good alternative to traditional displays (bar and pie charts), because in the dot plot the categories are listed vertically, thus attenuating the connotation of ranked categories, and the associated magnitudes are shown horizontally offering a preferred axis of display [24].

The dot plots are used to make inferences and also can be used to compare groups. The dot plot in Figure 7 not only shows the ranking of symptoms among patients and the percentages, but also shows that people have multiple symptoms since the total is greater than 100%. Adding interactivity also aids in making comparisons between grouped hierarchies in dot plots. For instance the brushing technique can be used to highlight relationships across the display. Brushing is a process in which a user can highlight, select, or delete a subset of elements being graphically displayed [31]. Using Figure 7, a user could highlight the headache row and see the number of associated symptoms in the other rows highlighted automatically.

.

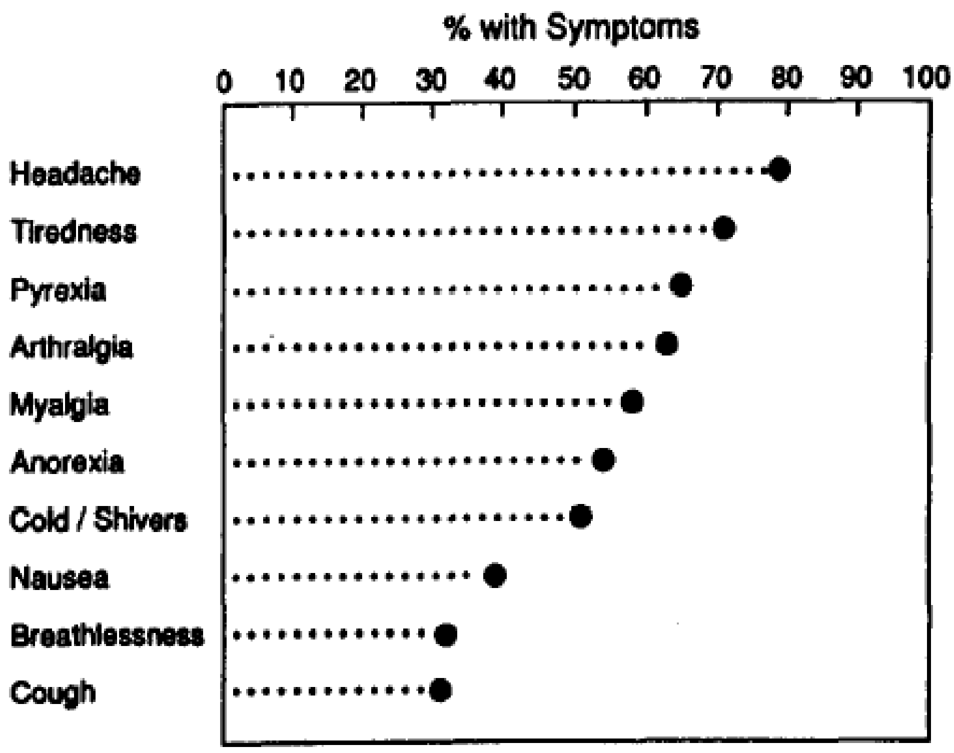


Figure 7. Dot Plot Representing Percentage of Patients with a Particular Symptom [25]

2.2.2 Relational Data

Sometimes data in hierarchies takes a more abstract form, where each element influences one or more elements, or just overlaps multiple categories. These types of data sets tend to provide an additional challenge to visualize, since laying them out optimally involves some extra computation [32].

Node-Link Diagrams

Node-Link visualizations have a statistical framework, these diagrams offer different takes on data that is highly interrelated: one element may link to hundreds (or thousands) of additional data points, or none at all (Figure 8). Node-Link charts (also called Network Diagrams) are best employed when data are related, but not necessarily in a clear hierarchy. Different dimensions of data can be shown by the size of each node, color, or even position. Many of the node-link publications deal with layout techniques complying with aesthetic rules such as minimizing the number of edge-crossings, minimizing the ratio between the longest edge and the shortest edge, and revealing symmetries [33]. Some visualization toolkits allow the links in the visualization to be different lengths, thicknesses, or colors to show an additional dimension of how two points are related.

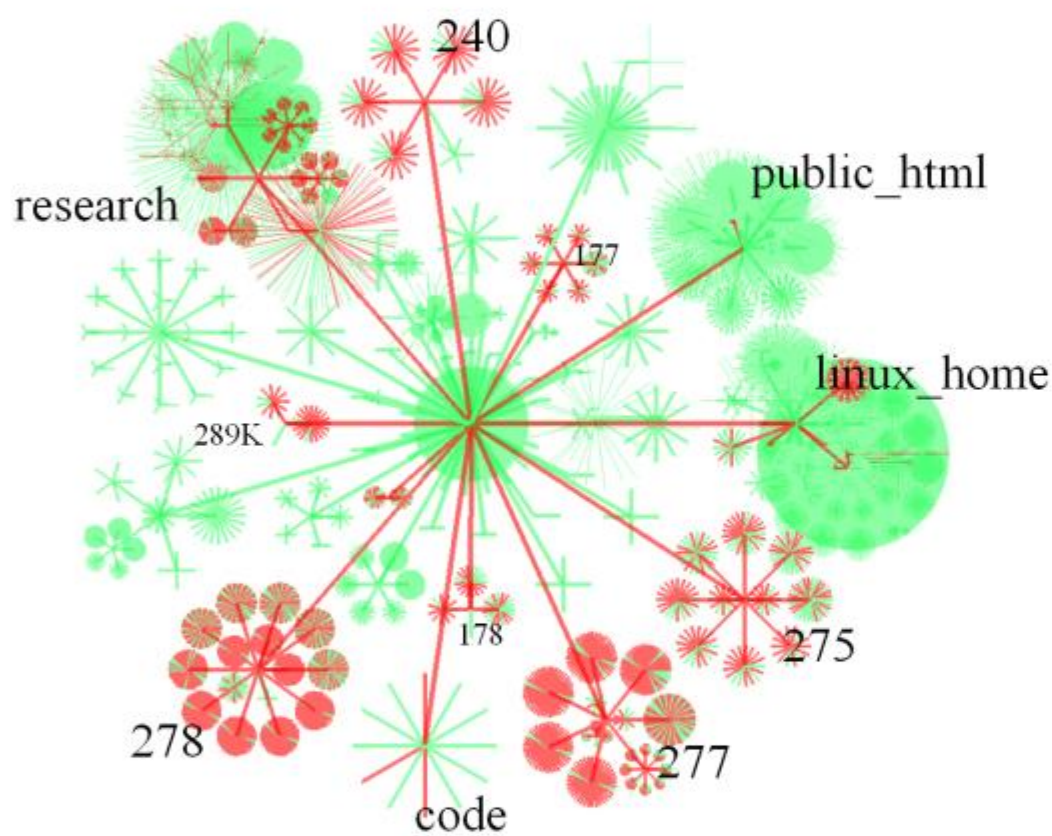


Figure 8. Node-Link Diagram of all Files in a System Directory [33]

Parallel Sets

Kosara et al. developed the concept of using parallel sets to visualize hierarchical data [26]. They developed the parallel set as a way of interacting and dealing with complex data sets. The method is based on the axis layout of parallel coordinates with the boxes representing the categories and parallelograms between the axes showing relationship of categories (Figure 9).

The basic building block of Parallel Sets is a box that represents the size of a category on one axis relative to all the data samples. Parallelograms connect categories to show how many data points are in any of the combinations between two or more categories. The color component is used to differentiate the categories and to make the connections between them easier to visualize.

Limitations of parallel sets are: 1) when there are many categories of different sizes, it can be hard to see and compare them and 2) there is a high learning curve.

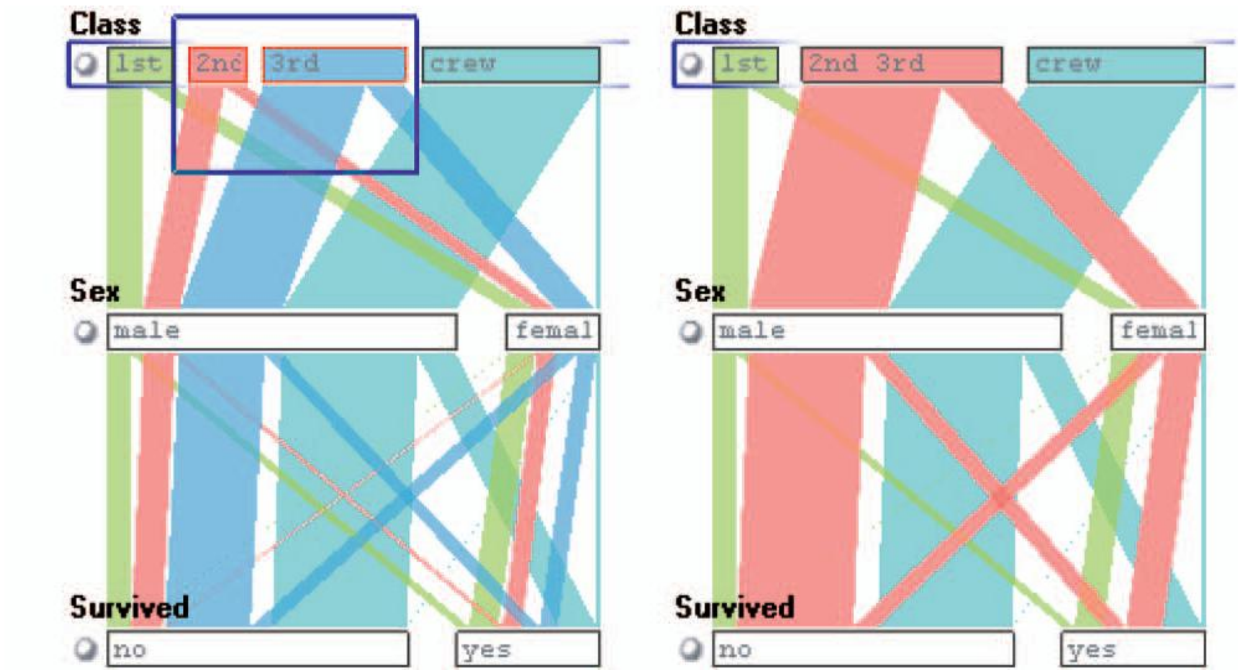


Figure 9. Parallel Set Grouping Categories of Class, Sex, and Survival Rate [26]

2.2.3 Nested Data

In general, something that is nested is fully contained within something else of the same kind. In data structures, data variables that are separately identifiable but also part of a larger data variable component are said to be nested within the larger. A nested hierarchy is a hierarchical ordering of nested sets [34]. There are a number of visualizations that are utilized to display nested data.

Tree/Flow Diagram

The simplest visualization to display hierarchical data is a tree diagram (Figure 10). A tree diagram is a representation of a tree structure, a way of representing the hierarchical nature of a data structure in a graphical form. The tree diagram starts with one item that branches into two or more, each of which branch into two or more, and so on. The visualization is used to break down broad categories into finer and finer levels of detail. Developing the tree diagram aids users in thinking about the data step by step from generalities to specifics.

Similar to tree diagrams, flow diagrams depict data in a hierarchical manner. Figure 11 shows the flow of patients through a study. The flow diagram starts from the inception of the study to a more detailed breakdown. Flow diagrams for large studies can become cumbersome, difficult to translate and read, and be time consuming in construction.

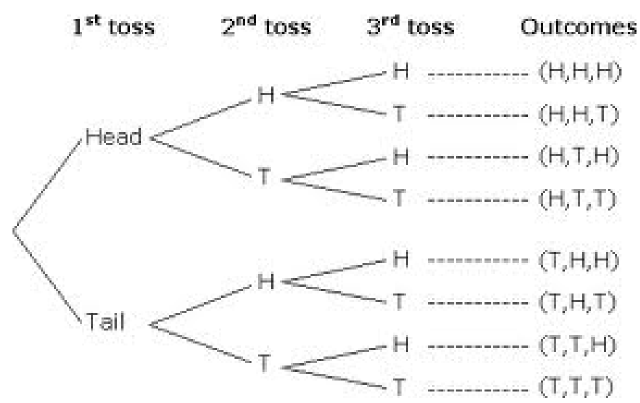


Figure 10. Tree Diagram of Coin Toss with Outcomes

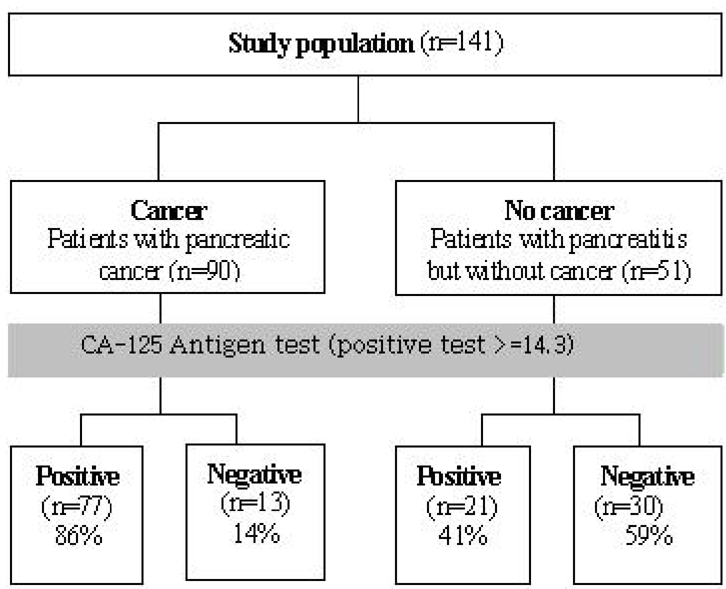


Figure 11. Flow Diagram of Pancreatic Cancer Experimental Study [29]

Multi-Level Pie

The multi-level pie chart is a data visualization format that is used for displaying hierarchical relationships. The visualization is useful when the data categories are more concise and defined. As the concentric rings go “out”, each item is sized with respect to its contribution to the inner parent category, allowing for deep hierarchies to be understood at a glance.

Figure 12 is an example of a multi-level pie chart that breaks down passenger information from the Titanic, and includes a record for each person on board, their class (First, Second, Third, Crew), gender (Male, Female), age (Adult, Child), and whether they survived (Yes, No) [35]. The idea can be extended to as many levels as desired. However the downfall of this layout is readability. As the number of levels increase, many of the ”slices” will be thin and difficult to clearly read and label.

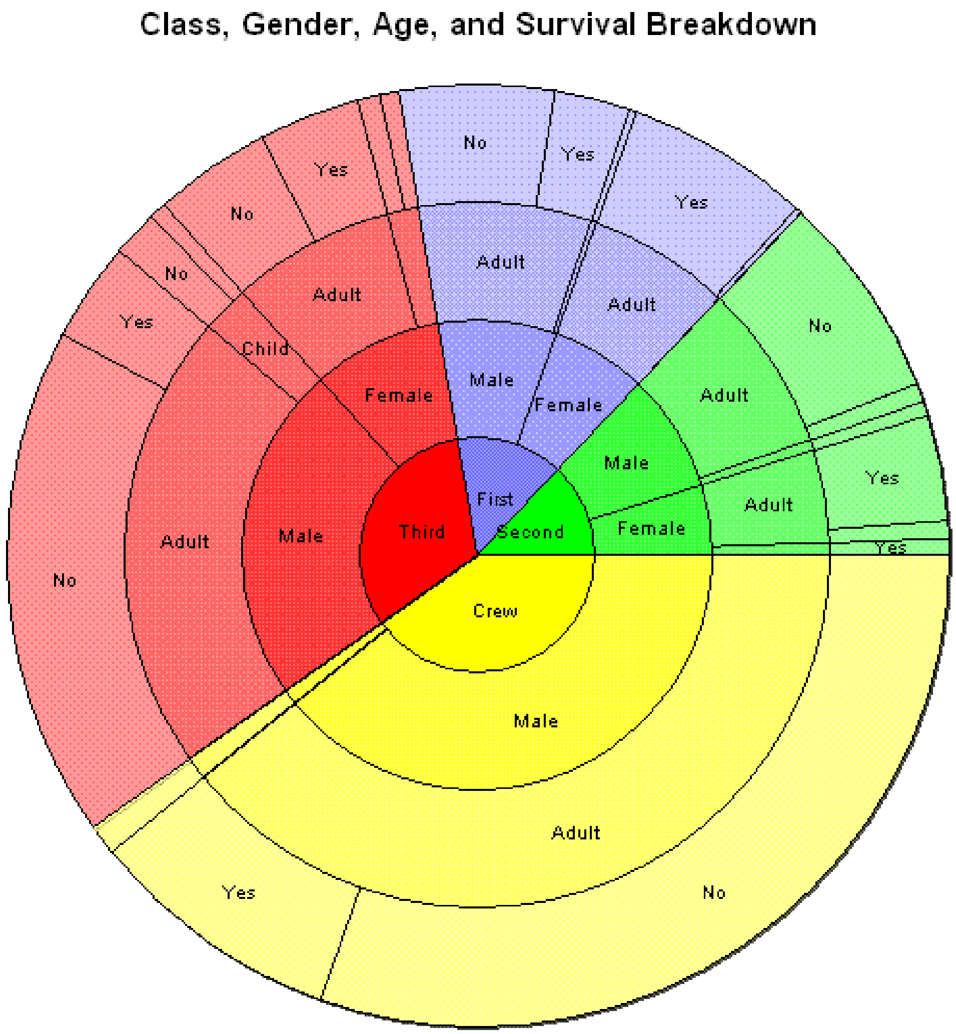


Figure 12. Multi-Level Pie Chart of Titanic Passenger Information [35]

Bubble Charts

A bubble chart displays a set of numeric values as circles (Figure 13). The visualization is useful for data sets with dozens to hundreds of values, or with values that differ by several orders of magnitude. Bubble charts are generally used to show multi-dimensional data (x, y, size, color). Each entity with associated data (v_1, v_2, v_3) is plotted as a disk that expresses two of the v_i values through the *xy* location of the disk and the third through its size [36]. Bubble charts can be considered a variation of the scatter plot, in which the data points are replaced with bubbles. The circles in a bubble chart represent different data values, with the area of a circle corresponding to the value.

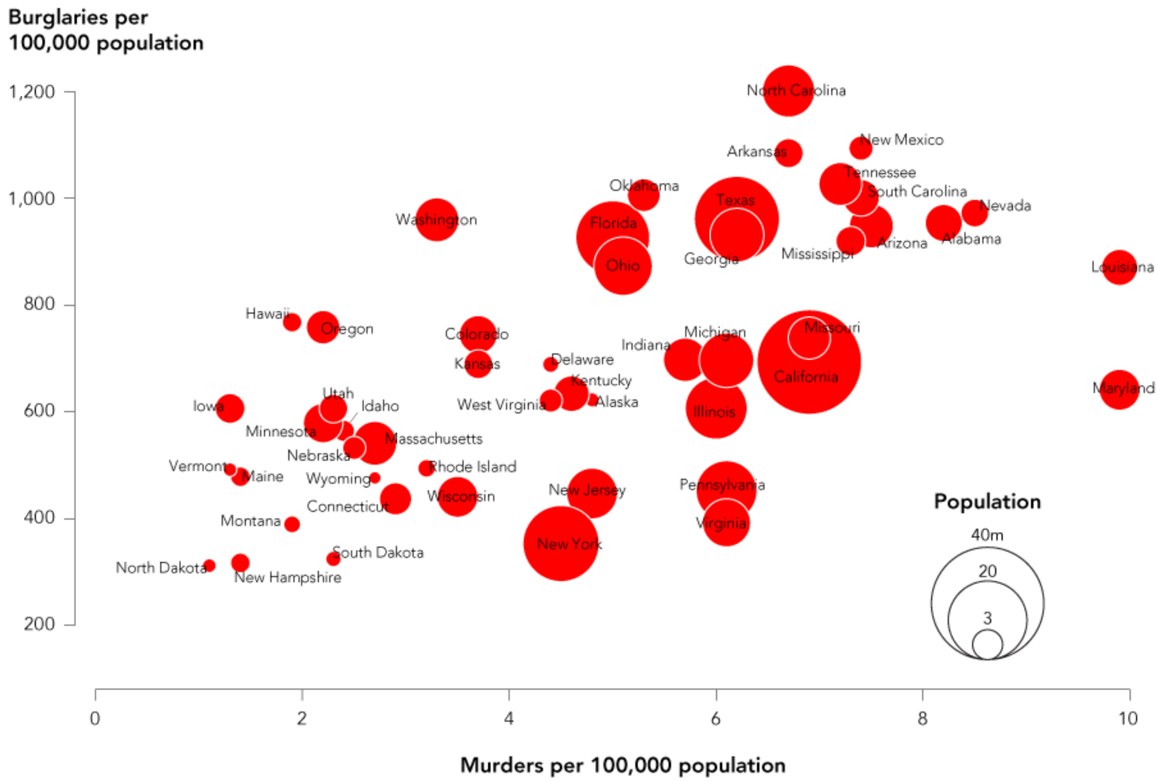


Figure 13. Bubble Chart of Burglaries and Murders for each State in the U.S. [36]

Treemaps

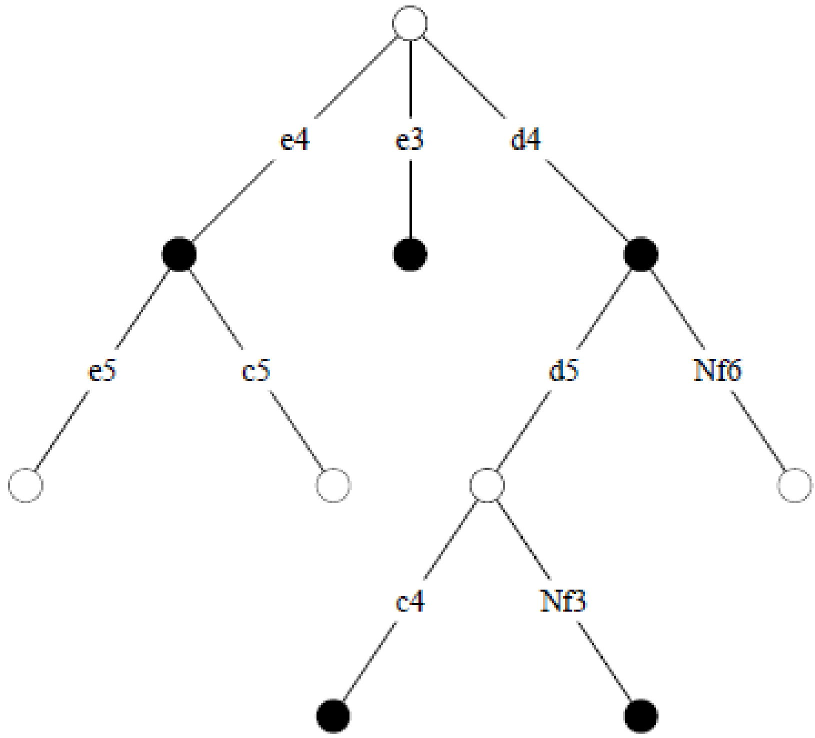
Treemap visualizations are widely used to display hierarchical data (Figure 1). In a treemap visualization, each category is sized according to what percent of the total it takes up, and child categories can be placed inside parents in a similar manner. The visualization allows aggregate categories to show through without losing the smaller constituent data. Interactive versions of the treemap allow for "drilling down" deeper into the data by clicking on a category to see the hierarchy on the entire screen.

Treemaps seem to work best when the total number of categories at each level of the hierarchy are finite (otherwise the hundreds or thousands of categories become tiny, undifferentiated squares) and each item fits neatly into a single sub-category. Since treemaps have been found to be useful for visualizing large hierarchical data sets [37-40], we chose to examine this visualization technique further, to test the validity and effectiveness of treemaps for relative area and proportional judgments. The next chapter discusses treemap visualizations in more detail.

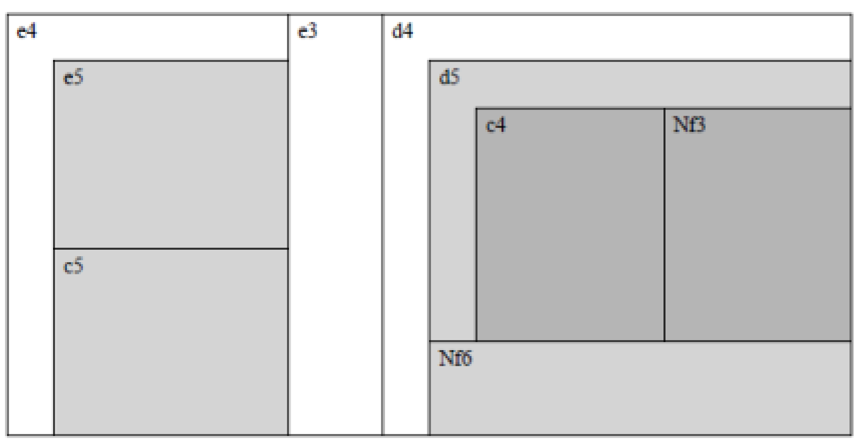
CHAPTER 3. TREEMAPS

3.1 Origins of Treemaps

Treemaps were designed by Ben Shneiderman during the 1990s in response to the problem of finding a way of producing a compact visualization of directory tree structures for a filled hard disk [37]. He created the treemap as a way for users to find large files that could be deleted, and to determine which users consumed the largest shares of disk space. Treemaps are a unique method for visualizing data in a two-dimensional display. Treemap presentations attempt to shift mental workload from the cognitive to the perceptual systems, taking advantage of the human visual processing system to increase the bandwidth of the human-computer interface [38]. A treemap data visualization uses 100% of the available display space by mapping the hierarchy onto a rectangular region.



(a) A traditional node-and-link representation of a tree



(b) The corresponding treemap

Figure 14. Principle of Treemap Visualization (a) Node-link Representation of Tree (b) Corresponding Treemap [40]

Treemaps have been described as “effectively combing aspects of a Venn diagram and a pie chart” [39]. The user interface for a treemap organizes hierarchical data as nodes, giving more space to nodes that are more relevant or important. Figure 14, demonstrates the principle of treemap visualizations from a traditional node-link tree diagram to the corresponding treemap visualization [40]. The treemap starts with a rectangular space and subdivides it recursively. The initial rectangle represents the root node of the tree. That space is divided into a number of horizontally aligned sub-rectangles, each representing a child of the root node. On each recursive call, the direction of subdivision is rotated: horizontal, then vertical, then horizontal [38]. Attributes of leaf nodes are represented using size and color coding.

3.2 Benefits of Treemaps

Treemaps were designed with four objects in mind: efficient space utilization, interactivity, aesthetics, and comprehension [41]. The visualization is used extensively in a wide variety of applications from railway monitoring [42] to Newsmap [43], which is an application that visually reflects the constantly changing landscape of the Google news aggregator. Treemap visualization is becoming increasingly popular having been featured on the TV series *24*; the Smithsonian Institution’s *HistoryWired*; and the *New York Times* [44]. Any data that are hierarchical in format can be visualized with treemaps for analysts to make a quick judgment about the data set. Table 2 illustrates some examples from various domains and how the data variables can be extrapolated for use in a treemap [41].

Table 2. Mapping Various Domain Data to Treemap Visualizations

Domain	Hierarchy	Size Measures	Color Measures
Financial	Cost Centers → line items	Amount	Year-over-year change
Project Management	Projects → Components → Tickets	Time spent, budget	Ahead/behind schedule, Over/Under budget
Blog, social sites, news	Content type → stories	Popularity, comments	Recent change in popularity
Student performance	Subjects → units → lessons	Classroom Time	Grade
US economy	Sector → subsectors → business types	Sales, employment	Change over time

There is an important aspect to compare various characteristics of treemap visualizations with traditional visualization techniques. Table 3 summarizes a comparison of treemaps to several common chart types. The ability for a treemap to represent a hierarchy of data within a restrained and rectangular space makes the visualization both efficient and effective [45]. While a dataset hierarchy can be represented in the traditional form of a network tree, there is a great deal of unused space in the visualization. The treemap fills the display space in its entirety and aesthetically, while still showing the entire structure of the data. One of the most significant features of the treemap is the ability to represent large datasets.

Table 3. Comparison of Visualization Techniques [45]

	Pie Chart	Histogram	Adjacency Matrix	Sociogram (node-link)	Treemap
Information conveyed					
Individual Nodes	No	No	Yes	Yes	Yes
Individual Ties	No	No	Yes	Yes	No
Categorical	Yes	Yes	No	No	Yes
Hierarchy	No	No	No	No	Yes
Size of Groups	Yes	Yes	No	No	Yes
Display-space utilization[1]:	75%	50%	90%	25%	100%
Free from dataset size?	Yes	Yes	No	No	Yes
Interactivity:	Static	Static	Static	Dynamic	Dynamic

3.3 Treemap Algorithms

There are a number of existing treemap algorithms [37-38, 46-47]. Traditionally, treemap algorithms exhibit a tradeoff between readability (optimizing aspect ratio) and ordered layout. The two most common algorithms are "squarified" and "slice and dice" [37-38]. Treemap algorithms vary in terms of layout stability, preservation of data order, and aspect ratio. Layout stability refers to the change in rectangle arrangements when there is a change in the data. An algorithm that depicts low stability indicates that the placement of the rectangles may be drastically repositioned even after a small change in the underlying data. This is the opposite result with an algorithm with high layout stability where the rectangle position would relatively stay the same even after changes to the underlying data. Preservation of data order is how well the layout algorithm maintains the order of the underlying data in the visualization. The treemap aspect ratio refers to the average ratio between the height and width of the rectangles. In general rectangles with lower aspect ratios (closer to squares) are favored in treemaps.

3.3.1 Squarified

The "squarified" (Figure 15) treemap algorithm divides the rectangular space into regions having an aspect ratio close to one [37]:

$$\forall r \in R, \frac{\max(r.width, r.height)}{\min(r.width, r.height)} \approx 1$$

Frames are placed around the nodes as a way to enhance the structure of the treemap. This helps to provide cues for identifying sibling relationships. The squarified algorithm starts with an initial rectangle of a certain height and width, the rectangle is split by two conditions: horizontal subdivision if the original rectangle is wider than high, vertical subdivision if the original rectangle is higher than wide. The left half is filled by adding rectangles until optimum aspect ratio is reached. The process is applied recursively to fill in the remaining portions of the rectangle. "Squarified" treemaps use approximately square rectangles, which offer better readability and size estimation than naive "slice-and-dice" subdivision.

To illustrate an example, Figure 15 shows the hierarchy of a surgery department starting with the type of surgery highlighted at the top in blue as colorectal surgery. The treemap then displays the next level of the hierarchy using the surgeon ID number of each surgeon who performed a colorectal surgery on a patient. Then under each surgeon ID is each patient for that particular surgeon. Within the treemap, hospital length of stay (LOS) and outcomes are represented by size and color, respectively. The size of the rectangles represents the LOS days for the patients. For example, we can see for Surgeon 25 in the colorectal group, patient 848 has the largest LOS days. The shading (color) of each leaf node is used to represent the outcome (1- Death (Red); 2- Alive (Green) of each patient. The treemap allows a quick visual representation of the percentage of patients with certain outcomes for each surgeon. A shortcoming of squarified treemaps is that the rectangles displayed are sorted by size rather than numerically. Many data sets contain ordering information that is helpful for seeing patterns or for locating objects in the map, which is difficult when utilizing the squarified algorithm where the original order is not preserved.

Figure 15. Squarified Algorithm of Surgery Hierarchy

3.3.2 Slice and Dice

The slice and dice algorithm (Figure 16) uses parallel lines to divide a rectangle representing an item into smaller rectangles representing its children. At each level of hierarchy the orientation of the lines (vertical or horizontal) is switched. The strengths of the algorithm are found in the change and readability measures [38]. Being an ordered treemap algorithm, it is particularly predictable in its placement of the rectangles, and hence it is easy to find a certain item among the rectangles. Though simple to implement and preserve the natural order of the data set, the “slice and dice” layout creates layouts that contain many rectangles with a high aspect ratio, this leads to long thin rectangles that can be hard to see, compare in size, and label.

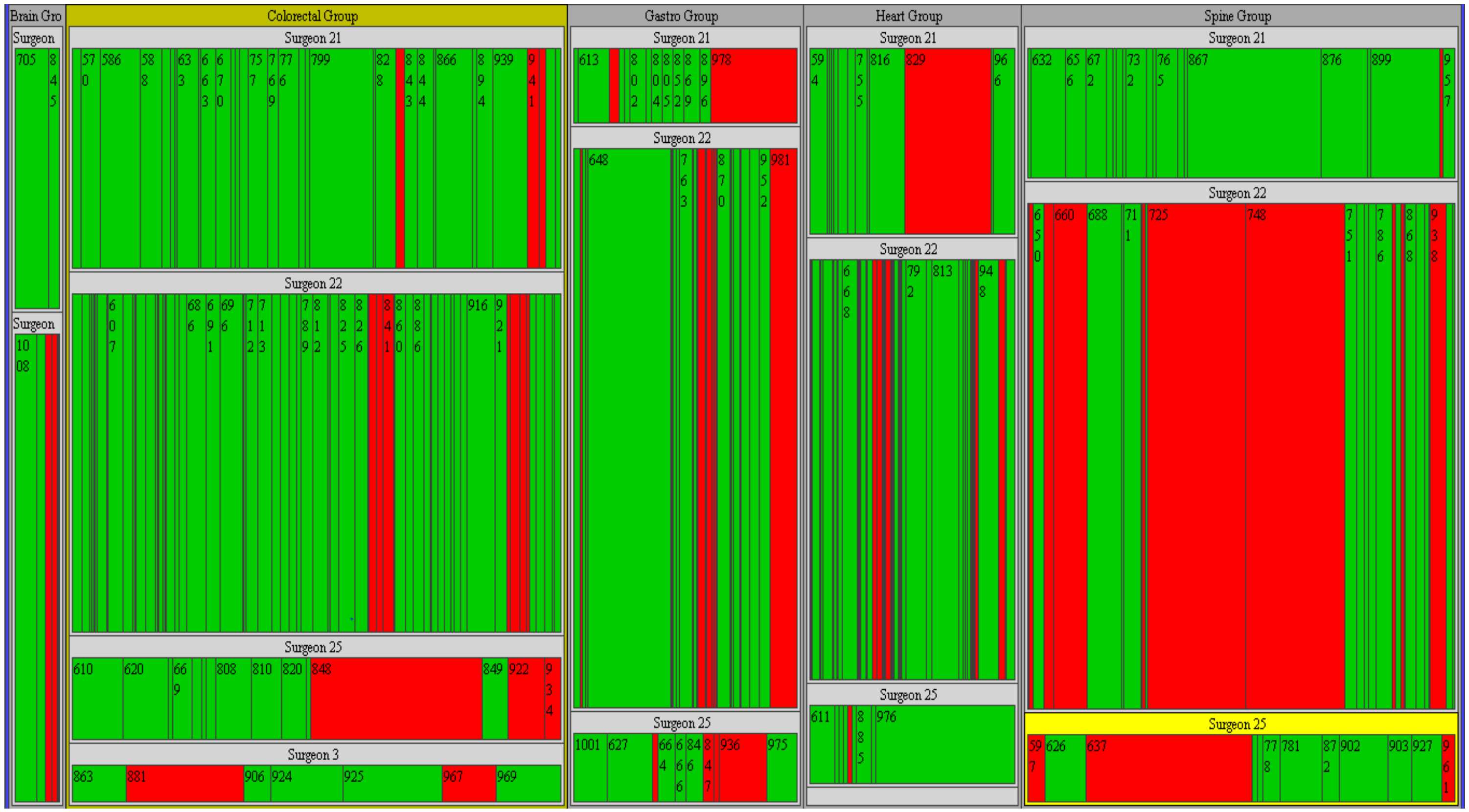


Figure 16. Slice and Dice Treemap Algorithm of Surgery Hierarchy

3.3.3 Strip

The strip treemap algorithm is a modification of the squarified treemap algorithm [46]. The algorithm works by processing input rectangles in order, and laying them out in horizontal (or vertical) strips of varying thicknesses (Figure 17). It is efficient in that it only looks at rectangles within the strip currently being processed and produces a layout with significantly better readability than the basic slice and dice treemap algorithm, and has comparable aspect ratios, and stability.

In the strip algorithm the inputs are a rectangle *R* to be subdivided and a list of items that are ordered by an index with given areas. The current strip is maintained and for each rectangle, there is a check to see if adding the rectangle to the current strip will increase or decrease the average aspect ratio of all the rectangles in the strip. If the average aspect ratio decreases (or stays the same), the new rectangle is added. If it increases, a new strip is started within the rectangle.

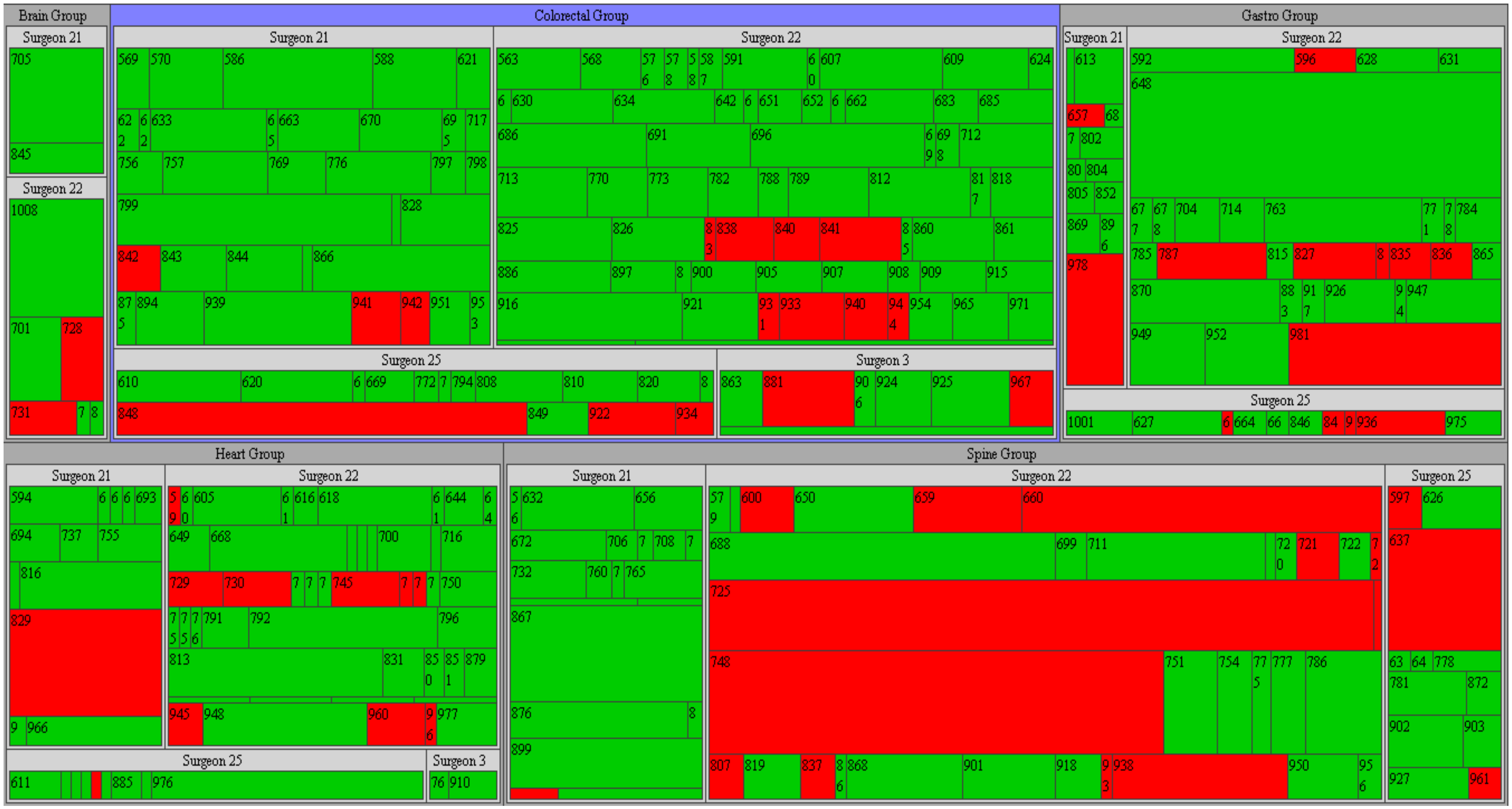


Figure 17. Strip Treemap Algorithm of Surgery Hierarchy

3.3.4 Cushion

Cushion treemaps are an extension of traditional treemap methods for visualizing hierarchical information. The standard treemap algorithms can have elongated rectangles as found in the slice and dice and strip layouts. As a result it can be difficult to compare nodes. Cushion Treemaps developed by van Wijk and van de Wetering try to alleviate that problem by applying a texture to the rectangles that make them appear like shiny cushions (Figure 18) [47]. The image shows a computer file system, with the color indicating the file type and the area the file size.

Due to the perceived discontinuity in texture between nodes, lines are no longer necessary to separate nodes, so more of the space can be used for the actual node display, and much smaller nodes can be shown than in a flat treemap. In the cushion treemap algorithm, shading is used to provide insight into the hierarchical structure. During the subdivision phase ridges are added per rectangle, which are rendered with a simple shading model. The result is a surface that consists of recursive cushions. For example, from the file directory in Figure 18, the larger the ridges or the more depth of the ridges shows higher level file directories.

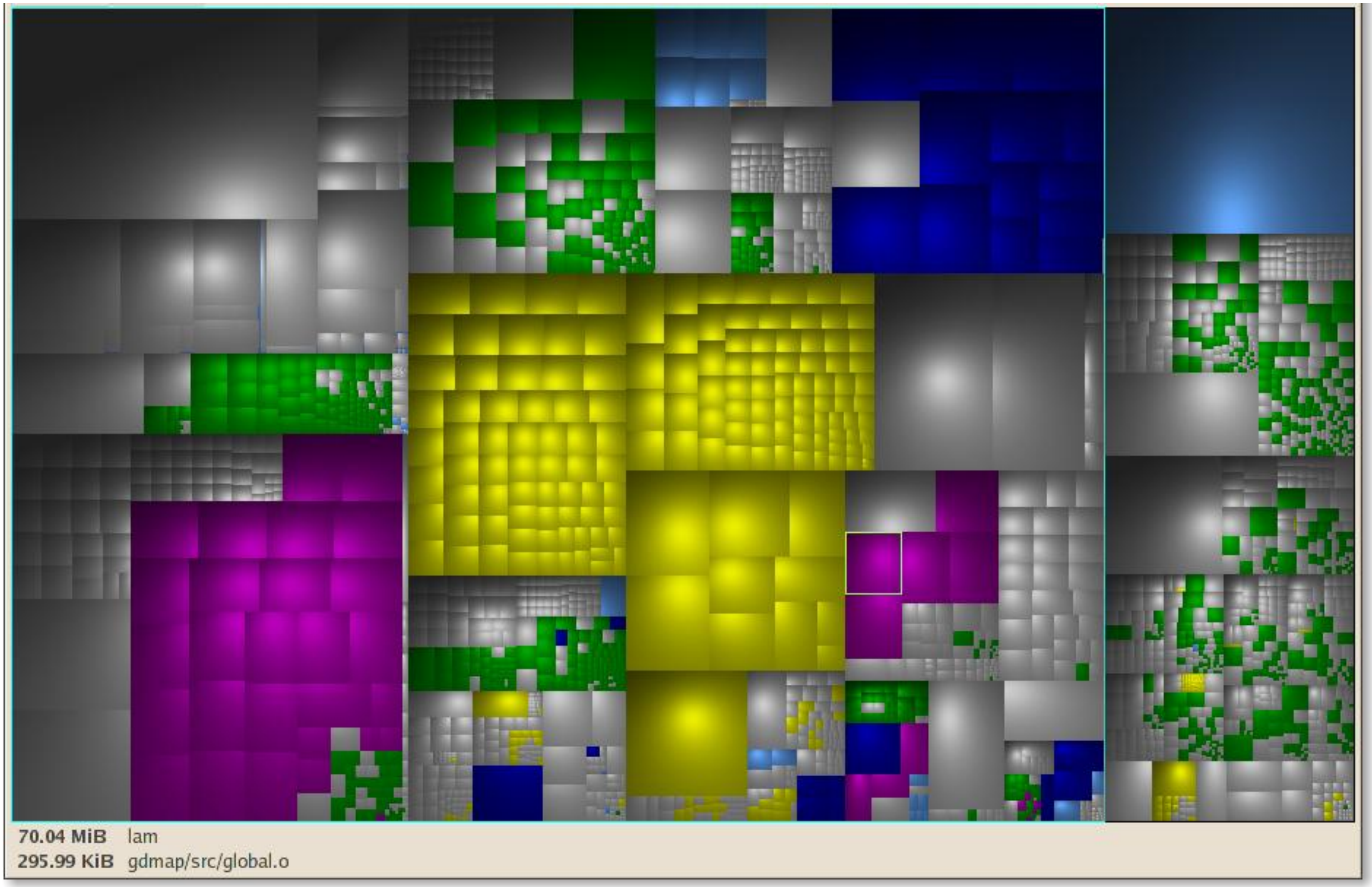


Figure 18. Cushion Treemap Algorithm of a File Structure [47]

3.4 Treemap Area Perception

To date only two researchers have examined area perception judgments within treemaps [48-49]. Heer and Bostock studied rectangular area judgment following the methodology of the Cleveland and McGill study. In the study, subjects were asked to identify which of two rectangles (marked A or B) was smaller and then estimate the percentage the smaller was of the larger making a "quick visual judgment". The experimental design of the study included a 2 (Display) x 9 (Aspect Ratio) factorial design. In the first display condition, subjects were shown two rectangles with horizontally aligned centers (Figure 19a). The second display depicted a treemap with 24 values, in which two rectangles were marked with A and B (Figure 19b). The aspect ratios were determined by the cross-product of the set $\left\{\frac{2}{3}, 1, \frac{3}{2}\right\}$.

Results from the study were analyzed using Cleveland and McGills log absolute error measure (Equation 2). There was a significant effect found in terms of aspect ratio on judgment accuracy. The comparison of rectangles with aspect ratio 1 exhibited the worst performance across both display conditions. There was no significant difference between the rectangle and treemap display conditions. The Heer and Bostock study has provided insight into rectangular judgment pertaining to treemap visualization, particularly that aspect ratio affects judgment when the ratio is close to one. However, there are several perceptual tasks involved in interpreting a treemap that may affect judgment. In essence there is a question of how luminance, area, orientation, distance, and time of judgment will affect performance.

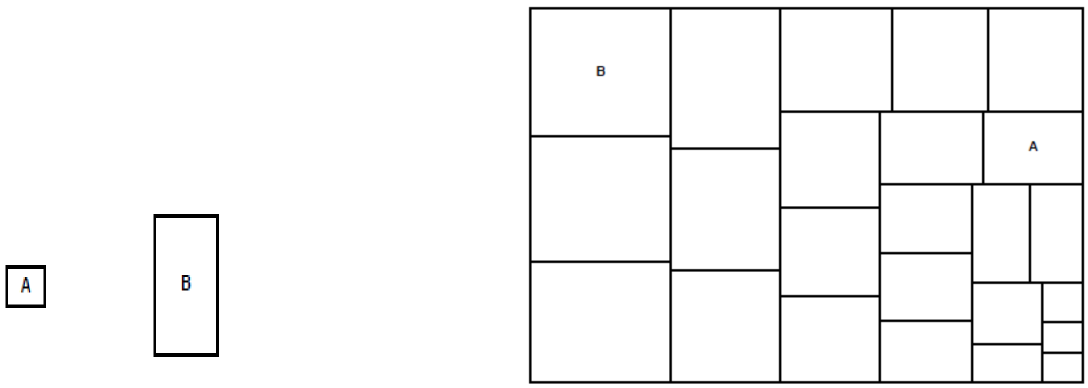


Figure 19. (a) Center Aligned Rectangles (Left); (b) 24 Value Treemap (Right) [48]

Kong, Heer, and Agrawala conducted a series of experiments to examine the effect of aspect ratio, color, and orientation on rectangular judgment [49]. In a pilot study, Kong et al. investigated the effect of luminance on the accuracy of proportion judgments. In the study, subjects were shown a 24 value treemap similar to that of Figure 19B. In each case two rectangles were marked A and B, in which subjects had to indentify which rectangle was smaller and what percent the smaller was of the larger. In each trial, the luminance of each cell was varied on a CIE L*a*b* color space according to a uniform distribution on a gray scale [51]. Results from the study concluded that there was not a significant effect on judgment accuracy due to luminance.

In a following study, subjects compared rectangular areas of varying size, aspect ratio, and orientation. Subjects were shown two center-aligned rectangles (similar to Figure 19A) and asked to identify the smaller rectangle and make a quick visual judgment on the percentage smaller. The experimental design consisted of a 4 (True Percentage) x 6 (Aspect Ratio) x 2 (Orientation). The percentages were varied between 32%, 48%, 58%, or 72%. The aspect ratios were determined by the cross-product of the set $\left\{\frac{2}{9}, \frac{2}{3}, 1, \frac{3}{2}, \frac{9}{2}\right\}$. The relative orientation was indicated by the rectangles having identical $\left\{\frac{9}{2}x\ \frac{9}{2}\right\}$ or different $\left\{\frac{9}{2}x\ \frac{2}{9}\right\}$ orientations. Results from the study found that there was not a main effect of orientation on judgment accuracy. There was a significant interaction effect between orientation and aspect ratio. Error increased when orientations differed and aspect ratios were extreme at $\left\{\frac{9}{2}\right\}$. They also found, similar to the Herr & Bostock study, that comparing squares leads to more error.

Both studies [48-49] explored rectangular judgments pertaining to perceptual tasks needed to interpret a treemap visualization. Insights rendered from the studies are:

- An aspect ratio of 1 exhibits the worst performance of judgment accuracy
- Luminance of rectangles does not affect judgment accuracy
- Conditions where orientation is different and aspect ratio close to 1 renders the worst performance
- Comparing diverse aspect ratios improves accuracy but can become difficult at extreme aspect ratios

3.5 Conclusions

This chapter has reviewed the origins of treemaps, described several alternative treemap generation algorithms, and introduced human perception research related to interpreting data elements commonly found in treemaps. The next chapter further explores characteristics of treemaps that affect human judgment and describes a human subjects experiment that was conducted to investigate these characteristics.

CHAPTER 4. STUDY 1: UNDERSTANDING HUMAN PERCEPTION OF RECTANGULAR AREA JUDGMENTS

This chapter explores the impact of treemap characteristics on the ability of a decision maker to interpret information displayed this way, by systematically varying treemap display variables and measuring the ability for a viewer to judge the percent difference in area depending on how and where the rectangles being judged are located in the treemap, extending the work of Heer and colleagues.

4.1 Introduction

With treemaps, analysts may have to compare non-contiguous rectangular cells of varying aspect ratio, area difference, and orientation displaced with respect to horizontal and vertical distance. Previous research has shown that, in general, when making area perception judgments of geometric stimuli, people tend to underestimate area differences [53-55]. In addition, relative area judgment accuracy decreases when area differences are closer to the middle of the scale (with worst performance peaking at a true difference of ~60%) [53,55]. Furthermore, human judgment performance is improved when items are closer and aligned along the same vertical scale, as translation and rotation create additional cognitive tasks. [56-59] Because rectangles in treemaps can be placed at any location on the two-dimensional treemap display, the offset angle between the longitudinal axis of one rectangle and a line connecting the two may also affect human judgment performance (Figure 20) but no previous research has systematically studied the potential effect of offset angle.

However even with aligned and similarly oriented rectangles, previous research has shown that aspect ratio can impact judgment performance. Both [48] and [49] examined relative rectangular area judgments for two center-aligned rectangles with varying aspect ratios. More square-like aspect ratios reduce performance, hypothesized due to use of one dimensional (1D) length comparisons to help estimate area [49]. [48] found that trials with the extreme aspect ratio pairs (9:2x9:2) also exhibited higher judgment errors and hypothesized that diverse but not extreme aspect ratio pairs would be preferred.

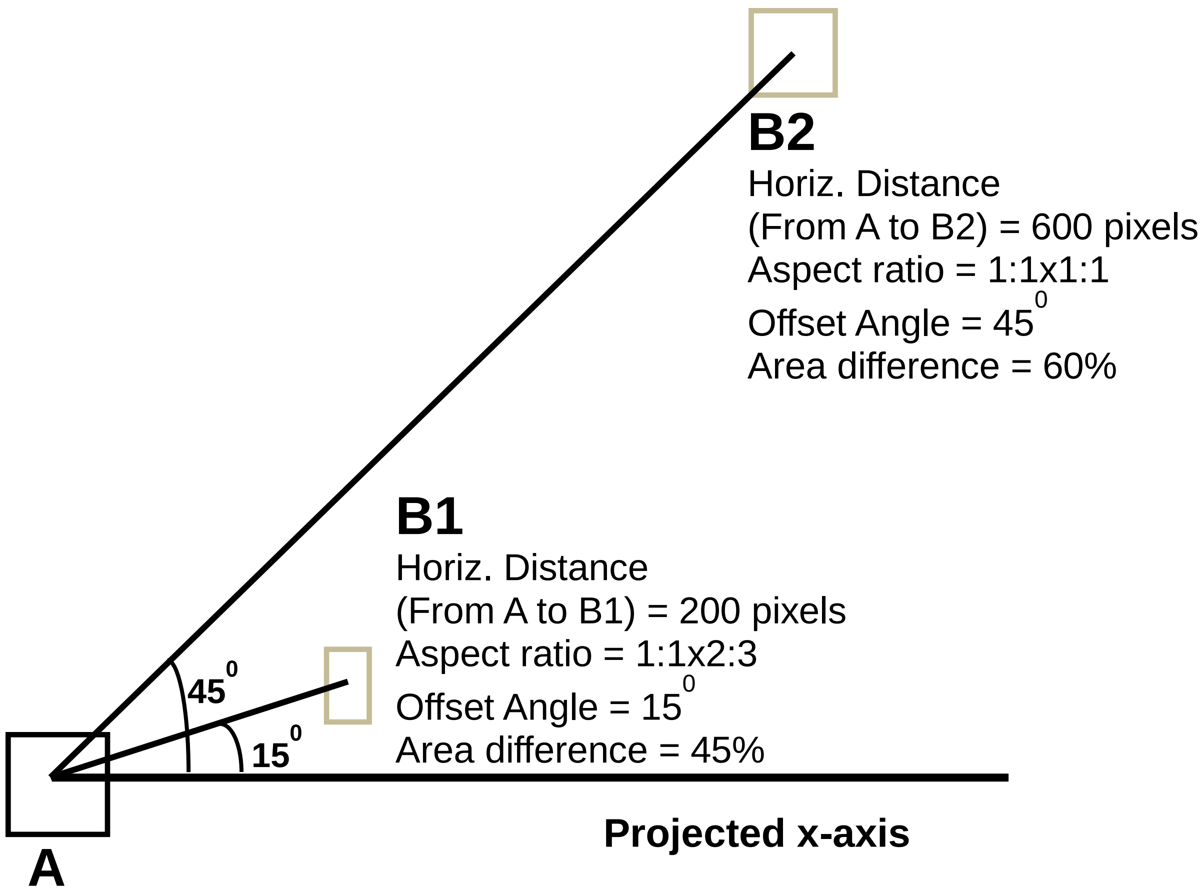


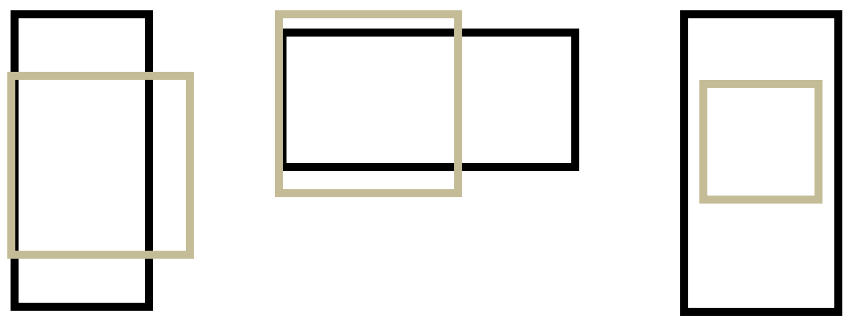
Figure 20. Geometry of Relative Rectangular Area Judgments

Complicating the analysis of human performance implications for relative rectangular area judgments is that the combination of area difference and aspect ratio creates an emergent geometric property: Whether rotated or not, either one rectangle can be contained inside the other (observable fit) or not (observable non-fit) (Figure 21). Individuals may be more accurate and make judgments more quickly when one rectangle fits within the other as mental translation can support such judgments [60]. Cave and Kosslyn [61] varied the sizes of geometric stimuli, and found that more time was required when a stimulus appeared at a different size. Anderson [62] found that when making judgment about area, humans use cognitive algebra using height and width to make a judgment. In the case of non-fit, whereby the length or width of the smaller rectangle is greater than the length or width of the larger rectangle, these cognitive algebra rules are harder for the observer to use.

$$a\mathrm{i} = h\mathrm{i} + w\mathrm{i} \text{ (height + width rule)}$$

$$a\mathrm{i} = h\mathrm{i} \text{ x } w\mathrm{i} \text{ (height × width rule)}$$

Previous research on rectangular area judgments has not explicitly addressed this concept of fit.



(a) (b) (c)

Figure 21. (a) Smaller Rectangle Does Not “Fit” within Larger Rectangle (Aspect Ratio Pair: 2:3x1:1 and Area Difference of 75%); (b) Smaller Rectangle Does not “Fit” within Larger Rectangle (Aspect Ratio Pair: 3:2x1:1 and Area Difference of 75%) (c) Smaller Rectangle “Fits” within the other Rectangle (Aspect Ratio Pair: 2:3x1:1 and Area Difference of 45%)

4.1.1 Objectives and Hypotheses

This study investigates the effect of geometric features of pairs of rectangles (in terms of true area difference, aspect ratios, horizontal distance apart and vertical offset angle) on the decision makers ability to judge the true difference in area between two rectangular stimuli. To validate the prior research and to extend it to systematically consider new features (offset angles and observable fit), we evaluate the following hypotheses:

Hypothesis 1: Decision makers will be less accurate (a) and will take longer (b) when making rectangular area judgments comparing square aspect ratio pairs as compared to non-square aspect ratio pairs.

Hypothesis 2: Decision makers will be less accurate (a) and will take longer (b) when making rectangular area judgments as offset angle increases.

Hypothesis 3: Decision makers will be less accurate (a) and will take longer (b) when making rectangular area judgments as horizontal distance increases.

Hypothesis 4: Decision makers will be less accurate (a) and will take longer (b) when making rectangular area judgments when the smaller rectangle does not fit within the larger.

Hypothesis 5: Decision makers will tend to underestimate rectangular area differences.

4.2 Methods

4.2.1 Participants

Thirty undergraduate and graduate science and engineering students participated in the study (n=15 males, n=15 females). The average age for participants was 23.3 years (*SD* = 3.14). They were each paid $10 for their participation. A protocol for this study was submitted and approved by the IRB for Social and Behavioral Sciences at the University of Virginia (#2011-0013-00).

4.2.2 Independent Variables

4.2.2.1 Within-Subject Variables

Four within-subjects independent variables were used to design the geometry of the area judgment trials: aspect ratio pairs, horizontal distance, offset angle, and true area difference (Table 4).

Aspect Ratio Pairs

Nine aspect ratios pairs were developed using all combinations of 2:3, 1:1, and 3:2 (average aspect ratio =1.04; standard deviation = 0.28). The aspect ratios used fall within 2 standard deviations of the Bruls et al. layout [6].

Horizontal Distance

Measured by projecting the centers of the rectangles to the x-axis, five horizontal distances (200, 300, 400, 500, and 600 pixels) led to trials where the centers were 2-6 inches apart (Figure 20). The horizontal position of the left rectangular stimuli was held constant; the horizontal position of the right rectangular stimuli varied in distance.

Table 4. Independent Variable Used to Create Trials

Independent Variable	Levels
Aspect ratio pairs	$\begin{Bmatrix} \frac{2}{3} x \frac{2}{3}\ ; \frac{1}{1} x \frac{1}{1} ; \frac{3}{2} x \frac{3}{2} \\ \frac{2}{3} x \frac{1}{1}\ ; \frac{2}{3} x \frac{3}{2} ; \frac{3}{2} x \frac{1}{1} \end{Bmatrix}$
Horizontal Distance	200, 300, 400, 500, 600 pixels
Offset angle	15, 30, 45 degrees
Area difference	45, 60, 75 percent

Offset Angle

The offset angle represents the minimum angle between a line connecting the centers of the two rectangular stimuli and a line projected parallel to the x-axis of the viewing window (Figure 20). The values for the offset angles were 15, 30, and 45 degrees. The vertical position of the left rectangular stimuli was held constant; the vertical position of the right rectangular stimuli varied counterclockwise.

True Area Difference

The true area difference denotes the actual (physical) percentage difference the smaller rectangle is of the larger. The values used were 45%, 60%, and 75%.

Observable Fit

By combining the area differences and aspect ratio pairs, an emergent attribute with two levels, *Fit* and *Non-Fit*, was determined based on whether one rectangle (rotated or not) could fit inside the other. Non-fit trials were those with true area difference of 75% combined with aspect ratio pairs of 2:3x1:1 and 3:2x1:1. All others were fit trials.

4.2.2.2 Between-Subject Variables

There were two between-participant independent variables: gender and block order.

Gender

Gender was modeled to account for potential differences associated with the spatial ability task of judging area between rectangular stimuli [63].

Block order

To account for potential order effects, trials were grouped into five equal-sized blocks. Each participant experienced one of five possible block orders designed using a Latin square randomization approach (Table 5).

4.2.3 Dependent Variables

Dependent variable measures for the study were log absolute error and completion time.

Log absolute error

Log absolute error measured the accuracy of the judged difference between the two rectangular stimuli [53]:

$$\log_2(|\text{judged percent - true percent}| + 1)$$

Completion time

The completion time was measured from the time a question appeared on the display to the time that the participant pressed the Enter key.

4.2.4 Procedure

Each session consisted of a short briefing, spatial ability assessment, training trials, and experimental trials, lasting a total of less than 1 hour.

Spatial Ability Assessment

Paper-based spatial ability assessments derived from the Newton and Bristoll Spatial Ability-Practice Test [64] assessed the spatial ability of each participant (Appendix B). The first two parts included rotation tasks where participants were presented with a figure and were asked to find the two-dimensional rotation. The third part was a 25 question shape-matching task where participants were asked to match the 25 shapes found in one panel with the 25 rotated shapes in the other. New participants would be recruited to replace anyone with an assessment score below 80%.

Relative Rectangular Area Judgment

Participants completed the area judgment portion of the study using a custom web application. Participants entered demographic information and then completed two area judgment training trials. For these trials, participants were provided with feedback on the accuracy of their judgments. The participants then completed 270 area judgment trials (see Figure 22) without feedback in blocks of 54 with breaks between blocks. In each trial, participants viewed a 1024 x 768 pixel image containing two rectangles and then identified which of the rectangles (A or B) was smaller and estimated the percentage the smaller was of the larger. Participants were encouraged to work quickly and to make “quick visual judgments.”

Table 5. Latin Square Design for Group Trials

Block Order	Order of blocks				
1	A	B	C	D	E
2	B	C	D	E	A
3	C	D	E	A	B
4	D	E	A	B	C
5	E	A	B	C	D

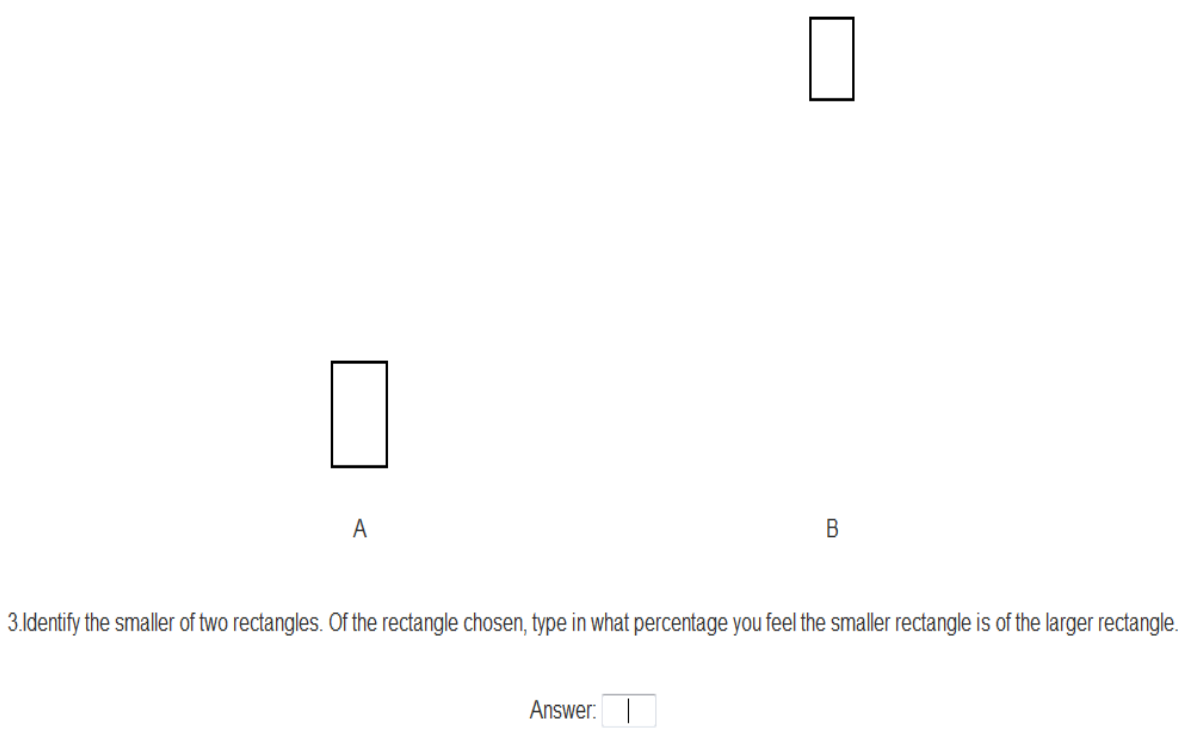


Figure 22. Example Trial: Aspect Ratio Pair 2:3x2:3; Area Difference 60%; Offset Angle 45 Degrees; Distance 400 Pixels

4.2.5 Experimental Design & Analysis

Spatial Ability Assessment

The percentage of correct responses out of the 27 total was calculated to assess the spatial ability assessment score.

Statistics Used

This study employed a repeated measures design with aspect ratio pairs, horizontal distance, offset angle, and area difference as within-subject factors and gender and block order as between-subject factors. Each of the 30 participants completed 270 trials. 54 trials were assigned to each of the five blocks. Six participants (3 female; 3 male) were randomly assigned to each block order.

Data were analyzed using repeated measures MANOVA with Wilks lambda. Effects found to be significant in the MANOVA were analyzed using repeated measures ANOVA for each dependent measure. For the significant effects, Bonferroni-adjusted post-hoc analysis was used to determine which levels were significantly different from the others.

To address the observable fit, repeated measures MANOVAs with Wilks lambda were conducted with the independent variables: fit, distance, and offset angle.

Binary logistic regression was used to evaluate the tendency for participants to overestimate or underestimate the judged area difference compared to the actual difference. Results are reported using $\alpha = 0.05$ for significance and 0.10 for trends.

4.3 Results

The MANOVA results revealed that there were no significant main effects for gender, $\Lambda = 0.88$, $p = 0.525$ or block order, $\Lambda = 0.65$, $p = 0.730$. Therefore the between-subject variables were not included in subsequent repeated measures ANOVA.

4.3.1 Spatial Ability Assessment

All participants met the inclusion criterion of scoring higher than 80% on the spatial ability assessment. There was no significant difference between males ($M = 95.5$, $SD = 5.28$) and females ($M = 95.7$, $SD = 5.33$), $t(28) = –0.1246$, $p = 0.901$.

4.3.2 Log Absolute Error

All participants correctly assessed which rectangle was smaller than the other for all trials. However, based on the log absolute error, judging the relative area between two rectangular stimuli with varying parameters proved to be a difficult human perception task ($M = 3.30$, $SD = 1.42$). This is approximately a difference of 10 in absolute error when judging relative area difference.

Table 6 shows the ANOVA results for absolute log error as a function of aspect ratio pair, horizontal distance, offset angle and area difference and Figure 23 shows the main and two-way interaction effects. The main effect of aspect ratio pairs was significant for relative area judgment. Aspect ratio pairs 1:1x1:1 ($M = 3.37$, $SD = 1.42$) significantly decreased accuracy more than pairs 2:3x2:3 ($M = 3.22$, $SD = 1.42$) and

2:3x3:2 (*M* = 3.22, *SD* = 1.42). Results also indicated a main effect for offset angle, such that the 45 degree offset angle (*M* = 3.34, *SD* = 1.43) decreased accuracy more than the 15 degree offset angle (*M* = 3.18, *SD* = 1.43) and the 30 degree offset angle (*M* = 3.21, *SD* = 1.43). These main effects were qualified by a significant interaction between aspect ratio pairs and offset angle. Figure 23 illustrates that while the accuracy is better at 30 degrees than at 45 degrees for all tested aspect ratios, there was no such consistent pattern for 15 degrees, where the accuracy varies depending on aspect ratio.

Table 6. Repeated Measures ANOVA for Log Absolute Error and Time as a Function of Aspect Ratio (AR) Pair, Horizontal Distance (Distance), Offset Angle (Angle) and Area Difference

Effects	Log Error	Time(sec)
AR Pair	$F(5,145)=2.25$; $\boldsymbol{p=.050}$	$F(5,145)=3.06$; $\boldsymbol{p=.017}$
Distance	$F(4,116)=1.60$; $p=.108$	$F(4,116)=1.09$; $p=.210$
Angle	$F(2,58)=12.01$; $\boldsymbol{p=.002}$	$F(2,58)=2.98$; $\boldsymbol{p=.018}$
Area Difference	$F(2,58)=1.04$; $p=.262$	$F(2,58)=2.05$; $\boldsymbol{p=.090}$
AR Pair x Distance	$F(20,580)=.903$; $p=.322$	$F(20,580)=1.17$; $p=.232$
AR Pair x Angle	$F(10,290)=2.15$; $\boldsymbol{p=.009}$	$F(10,290)=2.82$; $\boldsymbol{p=.005}$
AR Pair x Area Difference	$F(10,290)=2.86$; $\boldsymbol{p=.005}$	$F(10,290)=2.21$; $\boldsymbol{p=.010}$
Distance x Angle	$F(8,232)=2.13$; $\boldsymbol{p=.015}$	$F(8,232)=1.27$; $p=.204$
Distance x Area Difference	$F(8,232)=1.31$; $p=.238$	$F(8,232)=1.07$; $p=.310$
Angle x Area Difference	$F(4,116)=2.75$; $\boldsymbol{p=.020}$	$F(4,116)=1.15$; $p=.239$

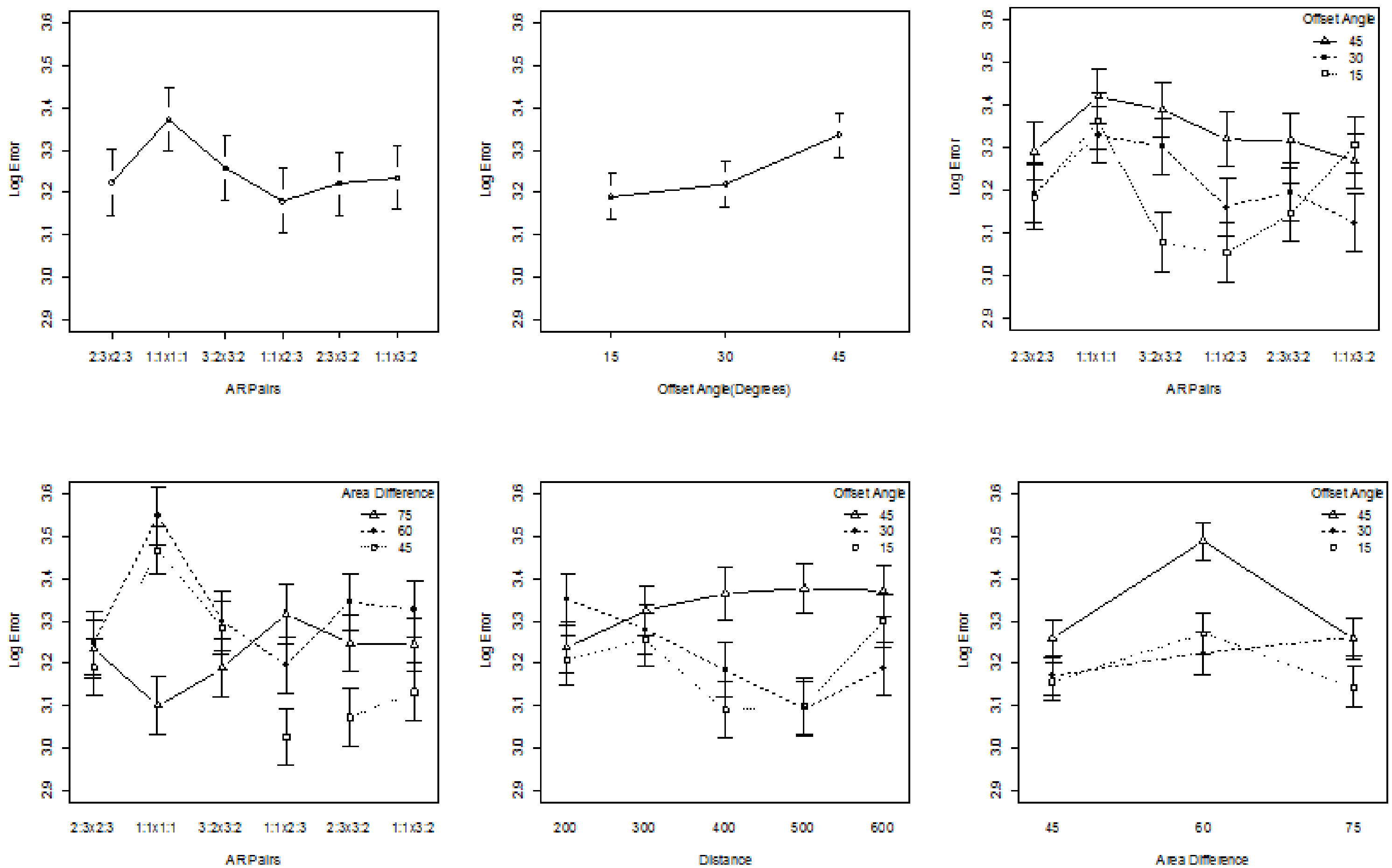


Figure 23. Main effects and Two-way Interaction Effects for Log Error as a Function of Aspect Ratio (AR) Pair, Horizontal Distance (Distance), Offset Angle (Angle) and Area Difference

Aspect Ratio Pairs × Area Difference, Offset Angle × Distance, and Offset Angle × Area Difference also had significant interaction effects for relative area judgment. Post-hoc analyses revealed that, for the Aspect Ratio Pairs × Area Difference interaction, accuracy decreases for squares (1:1x1:1) when the true difference is smaller (45% or 60%) but this degradation in accuracy goes away when the true area difference is larger (75%). For the Offset Angle × Distance interaction, area judgment error trends higher as the distance increases for offset angle 45, but trends lower as the distance increases for offset angles 15 and 30 (except at the extreme distance of 600 pixels). The Offset Angle × Area Difference interaction revealed that the two extreme areas (45% and 75%) are easier to judge than the middle area difference (60%) for offset angles 15 and 45 degrees, but at 30 degrees, we see a linear trend of accuracy decreasing as area difference increases. In general, looking at the related interaction plots in Figure 24, we can see that relative area judgment accuracy tends to decrease significantly when offset angle is 45 degrees as compared to 15 and 30 degrees.

A separate ANOVA analysis investigated the effect on log absolute error as a function of horizontal distance, offset angle and observable fit (Table 7). There was a main effect for Offset Angle and Observable Fit, with a trend towards significance for Distance. Post-hoc tests revealed that the 45 degree offset angle (M = 3.53, SD = 1.41) decreased accuracy more than the 15 degree offset angle (M = 3.20, SD = 1.41) and the 30 degree offset angle (M = 3.23, SD = 1.41). Non-fit (M = 3.43, SD = 1.41) trials decreased accuracy more than fit (M = 3.24, SD = 1.41) trials. The only significant interaction was for Offset Angle x Distance (see Figure 24).

Table 7. Repeated Measures ANOVA for Log Absolute Error and Time as a Function of Horizontal Distance (Distance), Offset Angle (Angle) and Observable Fit (Fit)

Effects	Log Error	Time(sec)
Distance	$F(4,80)=1.96$; $p=.097$	$F(4,80)=1.08$; $p=.364$
Angle	$F(2,40)=13.78$; $\boldsymbol{p=.001}$	$F(2,40)=5.21$; $\boldsymbol{p=.020}$
Fit	$F(1,20)=9.64$; $\boldsymbol{p=.002}$	$F(1,20)=.767$; $p=.410$
Distance x Angle	$F(8,160)=2.06$; $\boldsymbol{p=.035}$	$F(8,160)=1.88$; $p=.123$
Distance x Fit	$F(4,80)=.857$; $p=.488$	$F(4,80)=.856$; $p=.489$
Angle x Fit	$F(2,40)=.027$; $p=.972$	$F(2,40)=2.63$; $\boldsymbol{p=.071}$

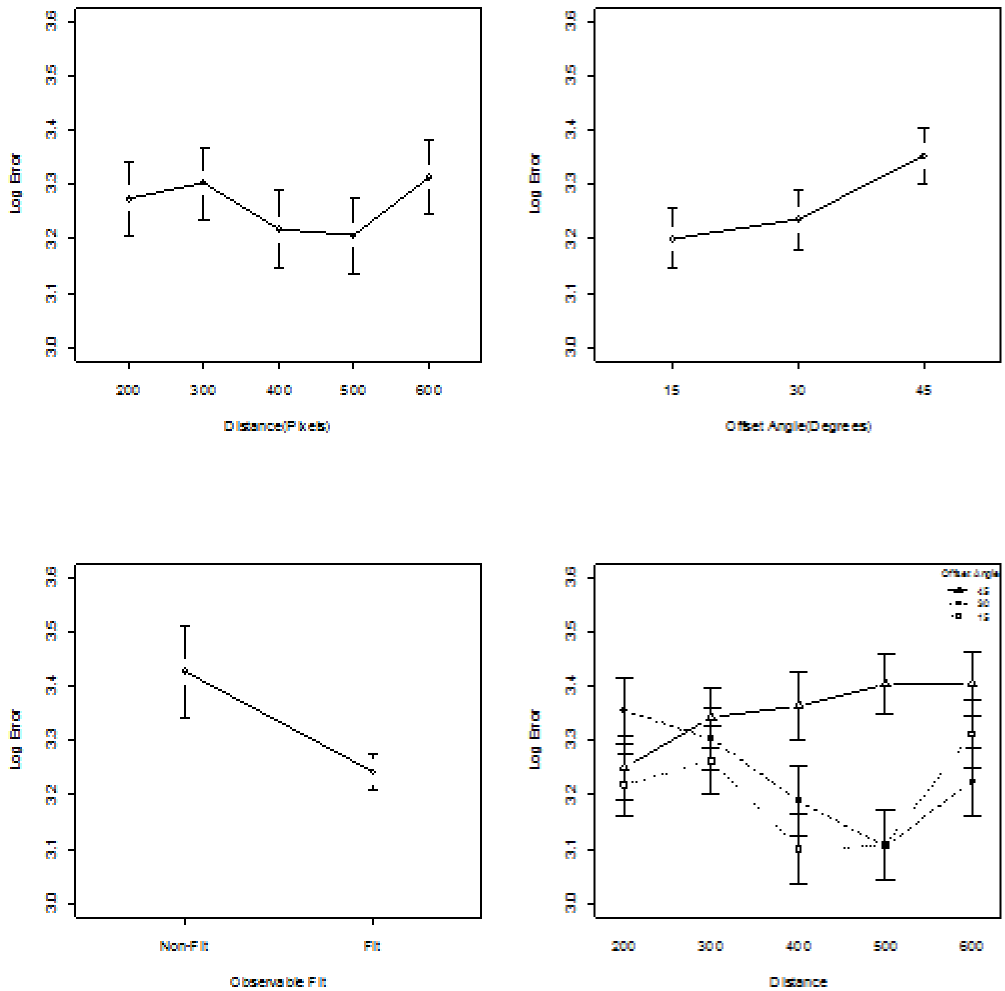


Figure 24. Main effects and Two-way interaction Effects for Log Error as a Function of Horizontal Distance (Distance), Offset Angle (Angle) and Observable Fit (Fit)

4.3.3 Judgment Time

Participants on average took 6.22sec, (*SD* = 6.39sec) to process relative area judgments of two rectangular stimuli with varying parameters. Results from the repeated measures ANOVA reveal that the main effect of aspect ratio pairs was significant for relative area judgment time. Aspect ratio pairs 2:3x2:3 (*M* = 6.70sec, *SD* = 6.39sec) significantly increased judgment time more than pairs 3:2x3:2 (*M*=5.92sec, *SD*=6.39sec) and 1:1x3:2 (*M* = 5.90sec, *SD* = 6.39sec). Offset angle was also a significant main effect for judgment time, such that the 45 degree offset angle (*M* = 6.34sec, *SD* = 6.39sec) increased judgment time more than the 15 degree offset angle (*M* = 6.13sec, *SD* = 6.39sec) and the 30 degree offset angle (*M* = 6.17sec, *SD* = 6.39sec). There was a linear trend in that as the offset angle increased from 15 degrees to 45 degrees, so did the judgment time (see Figure 25). Examining the interaction effects revealed significant two-way interactions for Aspect Ratio Pairs x Offset Angle and Aspect Ratio Pairs x Area Difference (Table 7 and Figure 25).

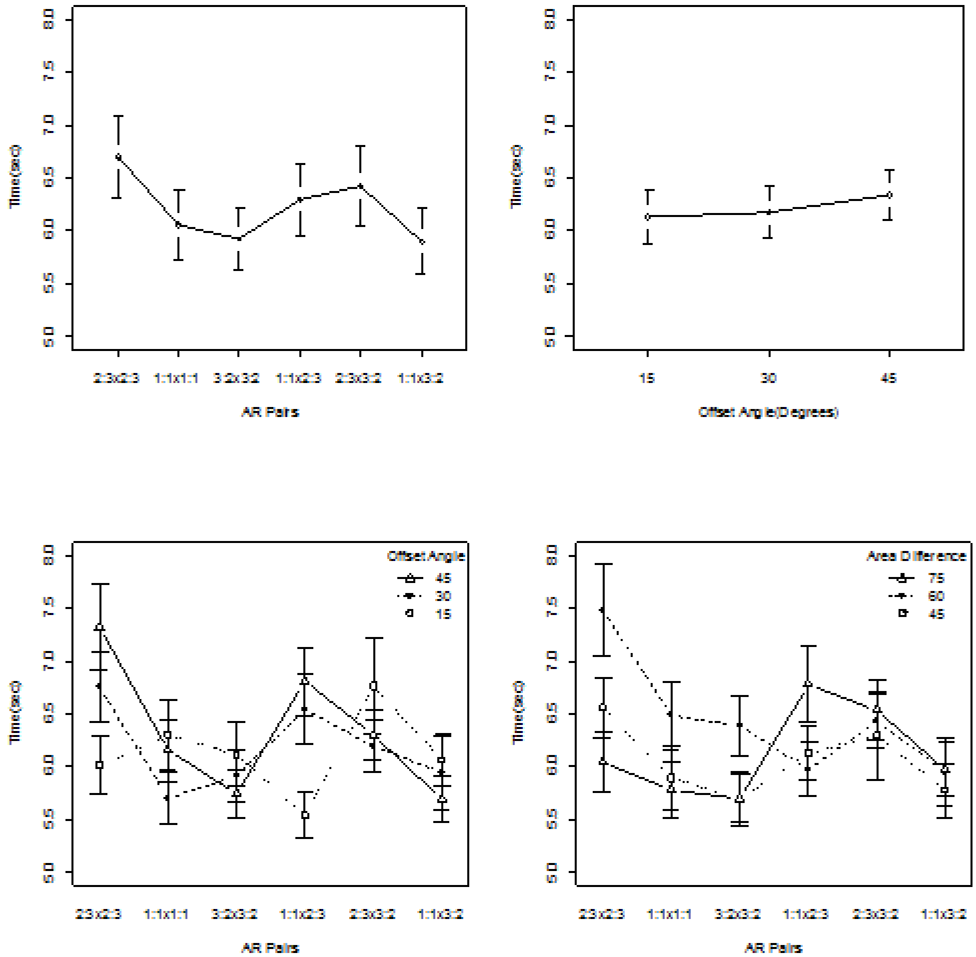


Figure 25. Main Effects and Two-way Interaction Effects for Time as a Function of Aspect Ratio (AR) Pair, Horizontal Distance (Distance), Offset Angle (Angle) and Area Difference

A separate analysis for judgment time was conducted using repeated measures ANOVA that included the independent variables of Distance, Offset Angle, and Observable Fit. Results of repeated measures ANOVA showed a significant main effect for offset angle (Table 7). From post-hoc analysis the 45 degree offset angle (*M*=6.34sec, *SD*=6.39sec) increased time more than the 15 degree offset angle (*M*=6.13sec, *SD*=6.39sec) and the 30 degree offset angle (*M*=6.17sec, *SD*=6.39sec), $p < .001$. The interaction effect Offset Angle x Observable Fit approaches significance. Figure 26 illustrates that Non-fit trials took longer to judge over the fit trials for the 30^0 and 45^0 offset angles, but this result was reversed at the smaller offset angle (15^0).

4.3.4 Judgment Bias

On average, participants underestimated area differences between the two rectangles (Figure 27). Cross tabulations on each independent variable show that participants had a higher percentage of underestimates for each level of the variables. Results from the logistic regression revealed that aspect ratio pairs (p = .001), distance (p = .011), and area difference (p = .001) were all significant main effects. The tendency to underestimate is more prevalent in trials where the rectangular stimuli judgments have a closer relative placement (Distance = 200 pixels and Offset Angle =15^0).

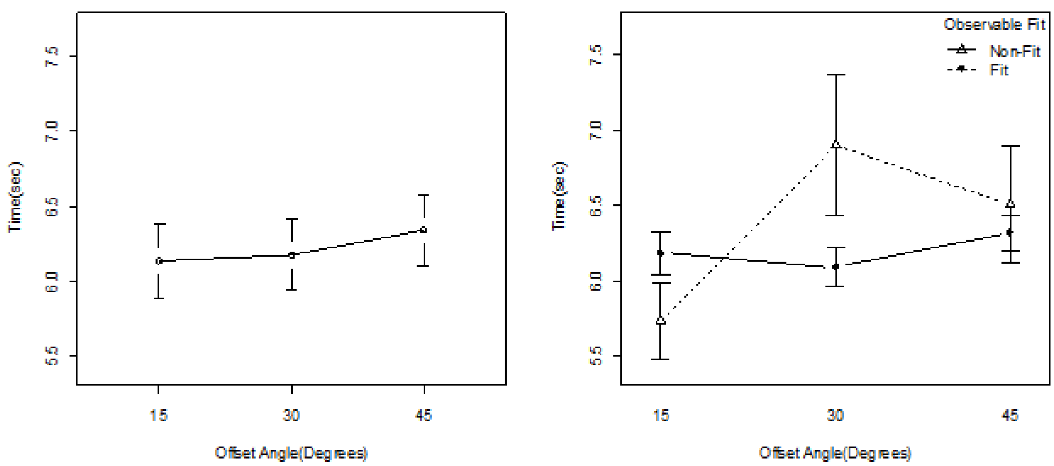
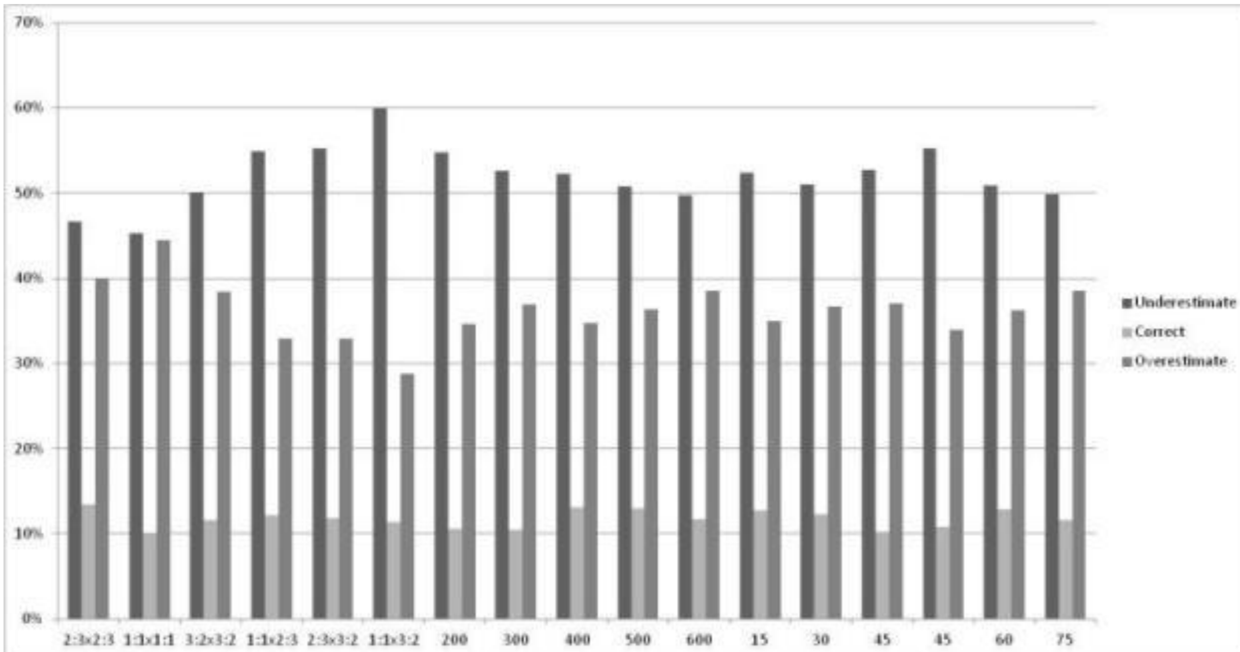


Figure 26. Main Effects and Two-Way Interaction Effects for Time as a Function of Horizontal Distance (Distance), Offset Angle (Angle) and Observable Fit (Fit)



Aspect Ratio Pairs Distance Offset Angle Area Difference

Figure 27. Cross Tabulations of Responses for Each Independent Variable

4.3.5 Results by Hypothesis

Hypothesis 1a: The square aspect ratio pair {1:1x1:1} yielded the highest error rate among participants (*M*=3.37, *SD*=1.39). Post hoc tests revealed that aspect ratio pair {1:1x1:1} was significantly different from other pairs. This result indicates that comparison of squares hinder judgment accuracy, consistent with Hypothesis 1a.

Hypothesis 1b: We hypothesized that square aspect ratio pairs {1:1x1:1} would result in longer judgment times. However, aspect ratio pair {2:3x2:3} had the longest judgment time (*M*=6.7sec, *SD*=6.39). Post hoc tests revealed that aspect ratio pair comparisons that included the 2:3 aspect ratio increased judgment time more than other rectangular area comparisons including aspect ratio 1:1 and 3:2.

Hypothesis 2a: There was a linear effect in that as the offset angle increased from 15 degrees to 45 degrees, so did degree of error, consistent with Hypothesis 2a.

Hypothesis 2b: We hypothesized that higher offset angles would increase judgment time. There was a significant linear trend for offset angle as the angle increases from 15 to 45 degrees, participants had longer judgment times. This confirms Hypothesis 2b.

Hypothesis 3a: We hypothesized that as distance increased, accuracy would decrease. Results from our study found that distance did not have a significant effect on the accuracy of area judgments by itself. However, the two-way interaction between distance and offset angle was significant in the repeated measurers ANOVA model. For offset angle 45 degrees, as distance increases so does log absolute error. This result implies that relative placement negatively effects area judgment accuracy, specifically large distances

(600 pixels) and offset angles (45 degrees).

Hypothesis 3b: We hypothesized that as distance increases participants would have longer judgment times. The main effect of distance on judgment time was found not to be significant. This result is not consistent with Hypothesis 3b.

Hypothesis 4a: Consistent with Hypothesis 4a, participants made significantly smaller judgment errors for the fit (M = 3.23, SD = 1.41) trials than for the non-fit (M = 3.43, SD = 1.41) trials.

Hypothesis 4b: There was no significant difference in judgment time for fit vs. non-fit trials, which indicates a lack of support for the hypothesis. However, there was a trend towards significance for the interaction Offset Angle x Observable Fit. Non-fit trials took longer to judge over the fit trials for larger offset angles, than for the smaller offset angles.

Hypothesis 5: Participants underestimated area 52% of the time when making rectangular area judgments, consistent with our hypothesis that participants would have a tendency to underestimate perceived area.

4.4 Discussion

The above experiments have investigated how geometric properties (aspect ratio pairs, true area difference, distance, and offset angle) influence area judgments when comparing rectangular stimuli. The study also investigated the perception of observable non-fit and observable fit trials. Results from the study are consistent with other studies assessing relative area magnitude judgments of rectangular stimuli that square aspect ratios are perceptually harder to judge [48,49] suggesting that the squarified treemap algorithm could potentially benefit from eliminating square aspect ratios. Surprisingly, square aspect ratios did not have the longest judgment time. Our study found that trials that included a rectangle with a 2:3 aspect ratio led to the longest judgment times. The study conducted by Heer and Bostock[48], on rectangular area judgments did not assess judgment time, but their results also found that trials that included a rectangle with a 2:3 aspect ratio led to higher error. It is not clear whether the vertical orientation of the aspect ratio may have attributed to longer judgment time as compared to trials that only included the 3:2 and 1:1 aspect ratios. Future studies could help to explain what heuristic is utilized when making the relative area comparison for a specific aspect ratio.

Our work further investigated how relative placement (horizontal distance and offset angle) impacts area judgment. The horizontal distances between graphical elements seem to bear no relationship to how accurately relative area is judged. Offset angle has an impact on rectangular area judgment both standalone and interacting with horizontal distance in terms of relative placement. An increase in offset angle size resulted in larger errors and longer judgment time. This coincides with related research that investigated the impact that angles have on proportion comparisons between geometric figures [65].

For both the proportion judgments and our relative area magnitude judgments, angle judgments were less accurate and took the most time to judge. This effect is compounded when examining the interaction of horizontal distance and offset angle. At the 45 degree angle error increased as the horizontal distance increased. Far distance and large acute angles lead to more difficult cognitive processing than linear aspects. Prior research [15] confirms that non-aligned geometric objects are perceptually harder to compare.

Our results suggest that participants were more accurate when the compared rectangles had observable fit versus observable non-fit. This was also directly correlated with relative placement (horizontal distance and offset angle). For the non-fit trials at the maximum horizontal distance and offset angle, participants had a higher error and longer judgment times. An empirical explanation of why participants have a difficult time judging non-fit trials can be implied from prior research [58] on cognitive stages humans use when comparing geometric objects. Johnson [59] developed cognitive processing steps that individuals make during perceptual comparison task. In the instance of making observable non-fit comparisons that are far in proximity, extra cognitive stages may potentially be added after the individuals rotate the two rectangular stimuli where there is a step of cognitively sliding the geometric stimuli to be closer in proximity and then deciding whether the rectangles fit or not. With the additional cognitive steps, we would expect individuals to take a longer time and have a harder time when making rectangular area judgments where the smaller does not fit within the larger (Figure 28). However, there was no significant difference in time found. Additional experimentation is needed to form a more accurate and complete perceptual model.

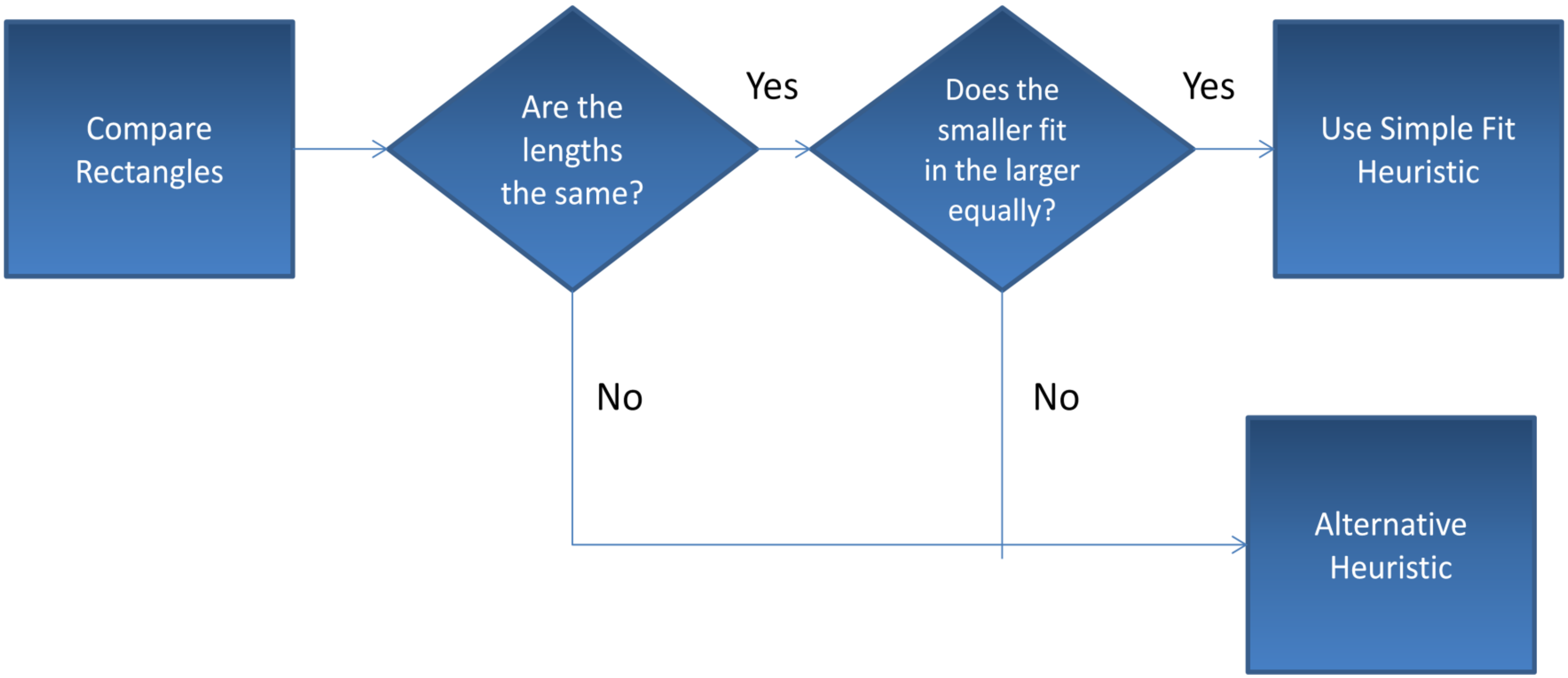


Figure 28. Rules for Area Judgment Heuristic in Observable Fit Case

Our study implies that geometric properties that systematically vary when representing key information in treemaps (the aspect ratio, true size difference, and relative distance and offset angle of the rectangles being displayed) as well as observable non-fit adversely affect judgment accuracy of rectangular area comparisons. These properties have more of a negative impact on rectangular area judgment and judgment time when there is an interaction between two of the properties. Participants on average differed in absolute error greater than 10 and the variance in estimates were large, indicating that this was a difficult perceptual task that will lead to decreased accuracy. For rectangular area judgment the geometric features of rectangle placement in squarified treemaps require consideration of relative distance and offset angle. It is sometimes necessary for users of treemaps to make such rectangular area comparisons to gain insight into relevant features of the data, construct accurate mental models of the information, and leverage parallelism in visual processing.

There remains a challenge of how to make individuals better at these rectangular area comparisons or whether we can adjust the algorithms or provide additional cues to help users, such as making such judgments automatically if there is an easy way for users to highlight the relative variables of interest. Implications from our study on rectangular area judgments suggest that utilizing alternative layout algorithms can help eliminate some of the information processing steps needed to make this type of judgment. The algorithm could potentially group like nodes together utilizing Gestalt laws of perceptual organization [11]. The close proximity could be better suited for making relative magnitude comparisons among users. For follow-on work, discussed in the next chapter, we implement a treemap system that contains a grouping algorithm to help in examining the intrinsic details of rectangular area comparisons and compare it to a traditional system. An alternative solution to aid the user in making such judgments would be to develop mechanisms for users to select subsets of the data of interest and to create specialized support displays to better enable users to make comparisons of selected features of the dataset being viewed.

There are certain limitations in the work presented. We conducted our experiment with participants examining two standalone rectangles instead of within a treemap. A previous study [48] demonstrated no significant difference between standalone rectangular area judgment tasks and rectangular area judgment tasks conducted within nodes of a treemap. We also did not vary the luminance of the rectangles, as is common in treemap layouts, but a previous study [49] demonstrated no significant difference depending on the luminance of the rectangles being compared.

Based on these results, we find that particular geometries of rectangular stimuli in area judgment result in people making systematic judgment errors. In general, relative area judgments of rectangular area comparisons result in low accuracy and do not support fast judgments. We can conclude that size difference (aspect ratio and area difference) and relative placement (distance between rectangles and offset angle) impact the decision makers ability to make relative rectangular area judgments. While utilizing treemaps to analyze large hierarchical data can help support fast characterization of the data, they may not be well-suited to making inferences on subsets of the data that require highly accurate rectangular area comparisons.

CHAPTER 5. STUDY 2: DESIGN AND VALIDATION OF AN ATTRIBUTE GROUPING TREEMAP ALGORITHM

This chapter describes a novel attribute grouping algorithm that was created as an alternative treemap layout to help support proportional area judgments. The attribute grouping algorithm was used in a human subjects experiment to validate the effectiveness of the algorithm.

5.1 Introduction

Many tasks performed by a human require the viewing of graphically displayed data. Individuals frequently study graphical visualizations to get a "feel" for their data.

Managers and economists study plots of various indices of production, employment, etc. Investors plot stock market averages and medical personnel plot outcomes and patients vital signs. An individuals real-time perception of the properties of graphically displayed data can influence his or her decision-making behavior [66]. This applies to human interpretation of the data displayed by treemap visualizations. We refer back to Figure 1, which depicts a treemap visualization of the United States S&P 500 stock market [8]. Stocks are categorized by industry area (e.g. technology, health care, financial, and energy) and the size of each industry block illustrates the activity of stocks in these areas. The activity of each stock is indicated by rectangle size (larger rectangles represent a stock that had higher trading than smaller rectangles) and the direction of price change for the stock is shown through color (red=loss, green=gain) with the brightness representing the amount of change (bright red or green means a bigger loss or gain, respectively, than dark red or green).

As with most treemaps in use today, the one shown in Figure 1 was generated by a “squarified” algorithm. Such an algorithm generates a treemap that avoids high aspect ratio rectangles (e.g., very narrow rectangles, where the length and width are quite different from each other), forcing the nodes to have aspect ratios closer to a square [6], since visually judging the area of rectangles with high aspect ratios is difficult [49].

When evaluating a treemap visualization, at a lower level, interpretation depends on the decision makers ability to perceive area, aspect ratio, luminance, and shading. On a higher level, an individual must be able to interpret how the variables fit within the nested hierarchies, in order to estimate the percentage that a variable corresponds to the whole hierarchy. For instance from Figure 1, the user at a quick glance can see that many of the health care sector stocks increased for the day (they are displayed in a shade of green, meaning their value went up, as opposed to a shade of red, meaning that their value went down). However, it remains a question how accurately can the user judge the percentage of health care stocks that increased for the day, compared to the percentage of energy stocks that increased for the day. The random placement of the nodes by size from the squarified treemap algorithm makes this proportional area judgment more difficult. The user potentially would have to cognitively group or count all of the green colored nodes together for stocks that had a gain and group the red colored nodes for stocks that had a loss for the day in order to come up with an assessment.

Perceptual principles identify ways that people quantify and estimate items. Goldstone [67] conducted a study where subjects were asked to estimate the percentage of display items that had a particular feature. Features were either randomly distributed or spatially clustered so that the features of the same type tend to be close. It was found that subjects overestimated randomly clustered figures. Numerosity research has identified ways that people quantify items. For up to six items, people can accurately and nearly instantaneously quantify the items [68]. This ability is called *subitization*. Subitizing is "instantly seeing how many", which is attributed to the recognition of patterns [69]. Individuals can also keep a running count while enumerating each item. Counting can theoretically quantify any number of items, but can be time consuming with larger sets. There is a thought that people can estimate or count the number of items in a particular region and then extrapolate. This method requires less effort than counting, especially when regions are small and easily identifiable, but still requires attentional focus and nontrivial calculations. Research also supports the notion that individuals use area occupied by items as a cue for the total quantity [70]. Our intuition is that the perceived estimation of rectangles occupying an area depends on how people perceptually group items and the strategies used to arrive at the grouping.

These implications suggest that utilizing alternative treemap layout algorithms can help eliminate some of the information processing steps needed to make proportional judgments. In this research we explore an alternative treemap algorithm that we call “attribute grouping” that groups like color nodes together, generating a treemap utilizing Gestalt laws of perceptual organization and Wicken’s proximity compatibility principle (PCP) [11, 71]. Gestalt laws account for the observation that humans naturally perceive objects as organized patterns and objects. The principle of proximity states that, all else being equal, perception tends to group stimuli that are close together as part of the same object, and stimuli that are far apart as two separate objects [11]. Wicken’s PCP indicates that in a display relevant items should be rendered close together in perceptual space (close display proximity). The proximity referred to in the principle is perceptual and spatial. Perceptual proximity refers to the perceptual similarity between different components of a display. This includes distance, color, shape, and physical dimensions. Spatial proximity refers to the distance between the items of the displays.

We might expect that due to the clustering nature of the new attribute grouping layout, judging total percentage across continuous rectangles of a certain color will result in decreased error and time. The close proximity would be better suited for making relative magnitude comparisons among users. A human subjects experiment was conducted to compare the new attribute grouping algorithm to the existing squarified treemap algorithm.

5.2 Attribute Grouping Treemap Algorithm Theory

Insights from aforementioned studies [66-70] indicates that grouping items with like attributes aids in successful completion of perceptual grouping and proportion task, compared to completing the task when attributes are dispersed randomly. Our intuition is that creating an attribute grouping algorithm for treemaps that groups same color nodes together, but still keeps the area dimensionality will improve on proportional area judgments. Previous treemap algorithms have proposed creating more useful displays by controlling the aspect ratios of the rectangles that make up a treemap and preserving the order of the underlying data mapped to how the nodes are placed on the treemap. While these algorithms do improve visibility of small items in a single layout, they do not look at improving the clustering of a subset of the data that is colored.

The attribute grouping algorithm creates rectangles in a visual order that match the input to the treemap algorithm, specifically creating a layout in which attributes that are represented by the same or similar color in the data input are adjacent in the treemap. On the front end of the algorithm the colors are first ordered by hue (h), saturation (s), and brightness (b) utilizing a concept similar to lexicographic ordering [72].

Given two partially ordered sets A and B, then the lexicographical order on the Cartesian product of $A \times B$ is defined by [72]:

$$(a,b) \leq (a',b') \text{ if and only if } a < a' \text{ or } (a = a' \text{ and } b \leq b').$$

The function derives from the order used in a dictionary, where strings are compared in alphabetical order, from left to right.

The colors are compared using the precedence of h > s > b, to return the color order. Once the order is determined a built-in function is used to convert the resulting (H, S, B) to RGB values for actual colors. The RGB gives the amount of each red, green, and blue primary stimulus in the color. The transformation from HSB to RGB coordinates is a transformation from a Cartesian coordinate system to a cylindrical coordinate system [73-74]. To convert from HSB to RGB, the steps as well as the equations are shown in Table 8 followed by the code for this conversion concept in Figure 29. We then utilize the resulting RGB answer to sort the data by color and then size. The algorithm concept is depicted in java code in Figure 30. This code is provided in full in Appendix C.

Table 8. Equation and Steps for Converting HSB to RGB

1. Compute chroma, by multiplying saturation by the maximum chroma for a given brightness.	$C = V \times S_{HSV}$
2. Find the point on one of the bottom three faces of the RGB cube, which has the same hue and chroma as our color.	$H' = \frac{H}{60^\circ}$ $X = C(1 - \vert H' \bmod 2 - 1 \vert)$ $(R_1, G_1, B_1) = \begin{cases} (0,0,0) & \text{if } H \text{ is undefined} \\ (C,X,0) & \text{if } 0 \leq H' < 1 \\ (X,C,0) & \text{if } 1 \leq H' < 2 \\ (0,C,X) & \text{if } 2 \leq H' < 3 \\ (0,X,C) & \text{if } 3 \leq H' < 4 \\ (X,0,C) & \text{if } 4 \leq H' < 5 \\ (C,0,X) & \text{if } 5 \leq H' < 6 \end{cases}$
3. Add equal amounts of R, G, and B to reach the proper lightness or value.	$m = V - C$ $(R, G, B) = (R_1 + m, G_1 + m, B_1 + m)$

```
/**
 * Converts an HSB color value to RGB.
 * Assumes h, s, and b are contained in the set [0, 1] and
 * returns r, g, and b in the set [0, 255].
 *
 * Number  h                Hue
 * Number  s                Saturation
 * Number  l                Brightness
 * @return  Array           RGB representation
 */
function hsbToRgb(h, s, l){
    var r, g, b;

    if(s == 0){
        r = g = b = l; // achromatic
    }else{
        function hue2rgb(p, q, t){
            if(t < 0) t += 1;
            if(t > 1) t -= 1;
            if(t < 1/6) return p + (q - p) * 6 * t;
            if(t < 1/2) return q;
            if(t < 2/3) return p + (q - p) * (2/3 - t) * 6;
            return p;
        }

        var q = l < 0.5 ? l * (1 + s) : l + s - l * s;
        var p = 2 * l - q;
        r = hue2rgb(p, q, h + 1/3);
        g = hue2rgb(p, q, h);
        b = hue2rgb(p, q, h - 1/3);
    }

    return [r * 255, g * 255, b * 255];
```

Figure 29. Conversion Code of HSB to RGB

```
public int Compare (object x, object y) {
// local variables
color       Cx,Cy;
float       Hx, Hy, Sx, Sy, Bx, By;

// get Color Values
cmbColorField.dataProvider = new ArrayCollection([
{label:'Outcome',       data:'outcome'},
Cx=((outcome) x), color;
Cy=((outcome) y), color;
// get hue values
Hx = Cx.GetHue ();
Hy = Cy.GetHue ();
// get saturation values
Sx = Cx.GetSaturation ();
Sy = Cy.GetSaturation ();
// get brightness values
Bx = Cx.GetBrightness ();
By = Cy.GetBrightness ();

Cx=(Hx,Sx,Bx);
Cy=(Hy,Sy,By);


// determine order
// 1: Cx < Cy
if ((Hx < Hy) || (Hx == Hy && Sx < Sy) || (Hx == Hy && Sx == Sy && Bx < By))
return Cx < Cy;
else{
    // 2: Cx = Cy
    if (Hx == Hy && Sx == Sy && Bx == By)
    return Cx = Cy;
    else {
        // 3: Cx > Cy
        if ((Hx > Hy) || (Hx == Hy && Sx > Sy) || (Hx == Hy && Sx == Sy && Bx > By))
        return Cx > Cy;
        else return 0;
    }
        }
//Converts resulting HSB order to RGB
Cx=(Hx,Sx,Bx);
Cy=(Hy,Sy,By);
public static function convertHSBtoRGB(hue:Number, saturation:Number, brightness:Number)

//Color array which holds each elements Color Values
var colors:Object = {
'Cx'        :0xFF0000,
'Cy'        :0xFF0000,
}

if(Treemaps.SORT_ON == 'sortOn')//Group according to Color
v2 = Vector.<ITreeMapItem>(VectorUtil.sortOn(values, ['sortOny', 'sortOnx'],
                [Array.DESCENDING, Array.DESCENDING|Array.NUMERIC])); //Sort the list first by Color
else
v2 = Vector.<ITreeMapItem>(VectorUtil.sortOn(values, [Treemaps.SORT_ON], [Array.DESCENDING|Array.NUMERIC])); //Sort by size
var items:Vector.<ITreeMapItem> = map.compute( v2 );
```

Figure 30. Attribute Grouping Algorithm Components

The squarified treemap algorithm is then used to determine the proportion size of the nested rectangle. The algorithm starts with a rectangular space and subdivides it recursively. On each recursive call, the direction of subdivision is rotated: alternating horizontally and vertically [75]. For instance if the given data has areas of 6, 6, 4, 3, 2, 2, and 1 that need to be visualized in a treemap, the initial rectangle would need to be subdivided into seven rectangles in order to achieve this task. The first step of the algorithm is to split the original rectangle either horizontally or vertically. In this case horizontal subdivision is chosen. The left half is filled in first, since the biggest real estate area is given to the highest area (Figure 31). The first rectangle has an aspect ratio of 8:3, the second rectangle is placed on top of the first and the aspect ratio improves to 3:2. Next, the rectangle with area 4 is added above the original rectangles and the aspect ratio is worse at 4:1. The optimum real estate space for the left side of the rectangle is achieved when this rectangle is added, therefore the rectangle is processed on the right half of the total area. In this case, vertical subdivision is chosen since the rectangle has a greater length in height than it does width. In step 4, the rectangle with area 4 is added followed by the rectangle with area 3, which results in a decrease in aspect ratio. Adding the next area of 2, still does not improve the aspect ratio so it is assumed that the optimum real estate is realized and top right of the partition will be filled. A horizontal subdivision is used to add the area 2 rectangle, followed by the second rectangle with area 2. Once the last rectangle (area 1) is added the aspect ratio decreases, therefore the subdivision is changed to a vertical subdivision to maximize the real estate space. In the algorithm the steps are repeated until all of the rectangles are processed. The algorithm lays out rectangles in horizontal and vertical rows. Each time a rectangle is placed, the algorithm goes through a 2 choice decision paradigm. Either the rectangle is added to the current row or the current row is fixed and a new row is started in the remaining subrectangle.

After this component of the algorithm is completed, the resulting attribute grouping algorithm renders a layout where similar color nodes are grouped together keeping the aspect ratio values (Figure 32a) generated by a squarified treemap algorithm. Compared to the squarified treemap layout (Figure 32b) the layout groups the nodes together to hopefully decrease the cognitive steps needed for individuals to make proportional judgments from the treemap.

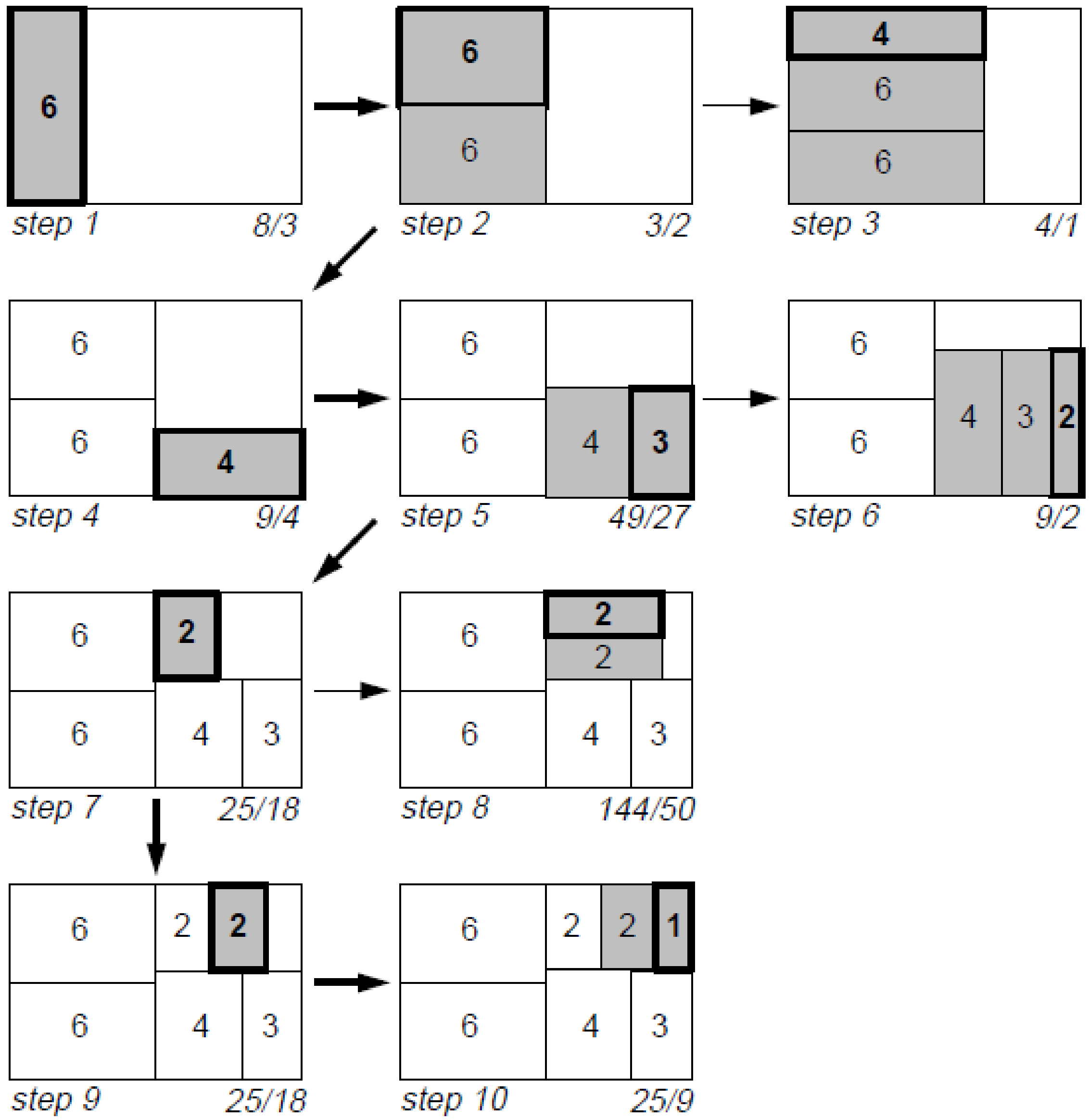


Figure 31. Example of Subdivision Treemap Algorithm

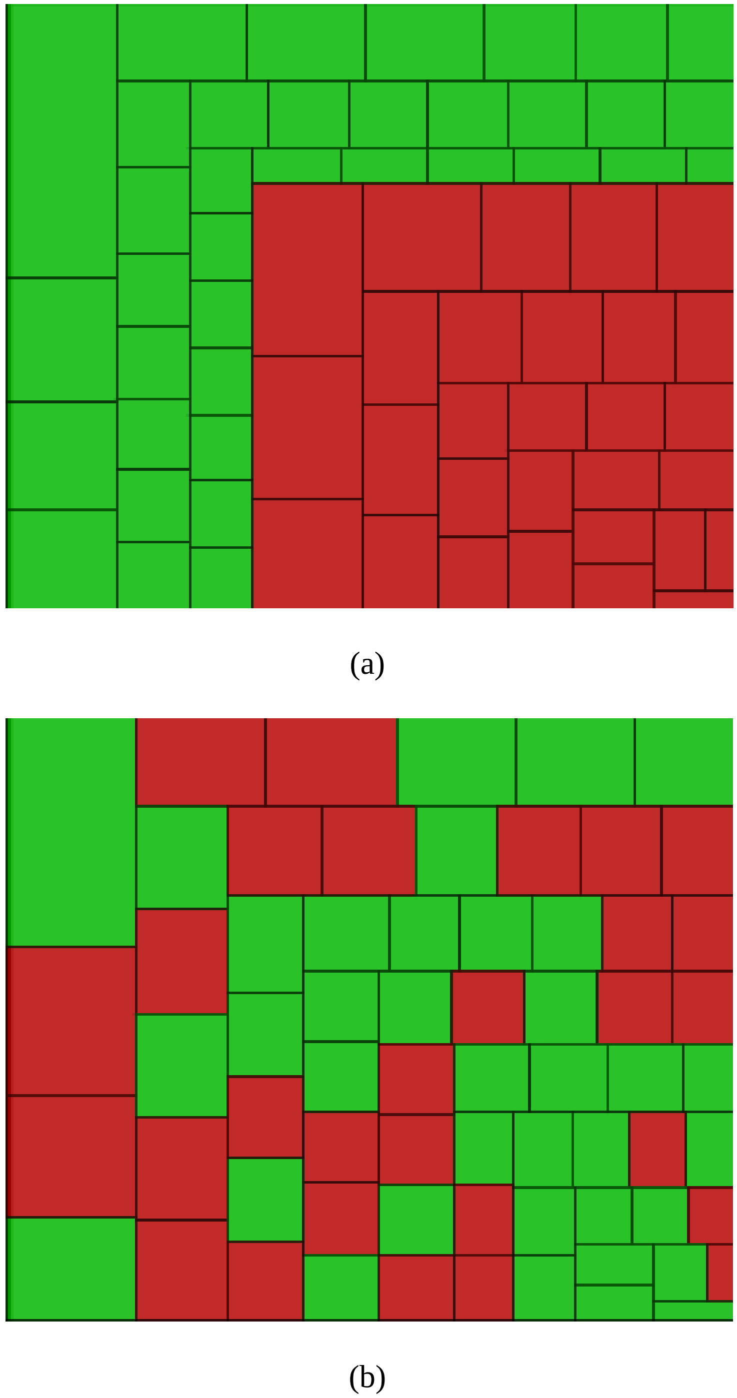


Figure 32. (a) Attribute Grouping Layout (b) Squarified Layout

5.3 Methods

To test the effectiveness of the attribute grouping layout we conducted a within-subjects experiment to compare the new layout to the common squarified treemap layout. Each participant performed a proportional judgment task using both layouts. Specifically, they were shown either a squarified or grouped treemap similar to those shown in Figures 33 and 34 respectively, and asked, “What percentage of the highlighted area is red?” Participants were encouraged to work quickly and to make “quick visual judgments.” This task was chosen as a representative sampling of the types of perceptual problems a user might run up against when analyzing the data. To reduce learning effects we utilized different order groupings for subjects (where subjects completed the first set of trials using the either the squarified or grouped treemap before moving on to the other half of the trials).

5.3.1 Participants

Sixty undergraduate and graduate science and engineering students from the University of Virginia volunteered for the study (n=28 males, n=32 females). The average age for participants was 22.2 (SD = 5.10). They were each given a $5 gift card for their participation. Subjects were randomly assigned to one of two conditions: Attribute Grouping Layout first or Squarified layout first, with half of the subjects performing each of these conditions. None of the participants had previous experience with the concept of treemaps. A protocol for this study was submitted and approved by the IRB for Social and Behavioral Sciences at the University of Virginia (#2011-0013-00).

5.3.2 Apparatus

A computer-based color blind assessment derived from Ishihara Color Test assessed the red-green color deficiencies of each participant (Appendix D) [76]. Participants performed the experiment on a workstation with a 22 inch flat panel monitor set to a 1024 x 768 pixel resolution, using a web-based custom apparatus utilizing a password protected server. Each of the responses for the participants were recorded using a MySQL database.

5.3.3 Procedure

Each session lasted 30 minutes or less. After a briefing and informed consent process, each used the computer apparatus to enter demographic information and complete a color blind test. The color blind assessment was used to ensure that the participants did not have any visual color deficiencies. Participants could not proceed with the study if they answered 2 or more questions incorrectly. Then each participant completed two training trials, the first using a squarified treemap (Figure 33) and the second using a perceptually grouped treemap (Figure 34). For these trials, participants were given feedback regarding the accuracy of their judgments. Participants then completed 50 experimental trials without feedback. Participants assigned to Order 1 saw the 25 squarified layout trials first followed by the 25 perceptually grouped trials and vice versa for participants assigned to Order 2.

Absolute error and completion times were recorded for each trial. The absolute error measure of accuracy is represented by:

$$|\text{judged percent - true percent}|$$

The completion time was measured from the time a question appeared on the screen to the time the subject pressed Enter on the keyboard.

5.3.4 Statistical Analysis

We utilized a within-subjects repeated measures design with a total of 50 trials per subject (n=60) for a total of 3,000 responses. Descriptive statistics were used to report measures of central tendency and variability for error and time on each independent variable. Main and interaction effects of the independent variables (order, layout, gender, and classification (undergraduate or graduate)) on the dependent measures (time and absolute error) were assessed using repeated measure multivariate analysis ANOVA. Tukey HSD post-hoc analysis was used to identify significant differences between the levels of the effects. Results for the main and interaction effects are reported using α = .05 for significance.

Figure 33. Squarified Treemap Layout for Experiment Trials

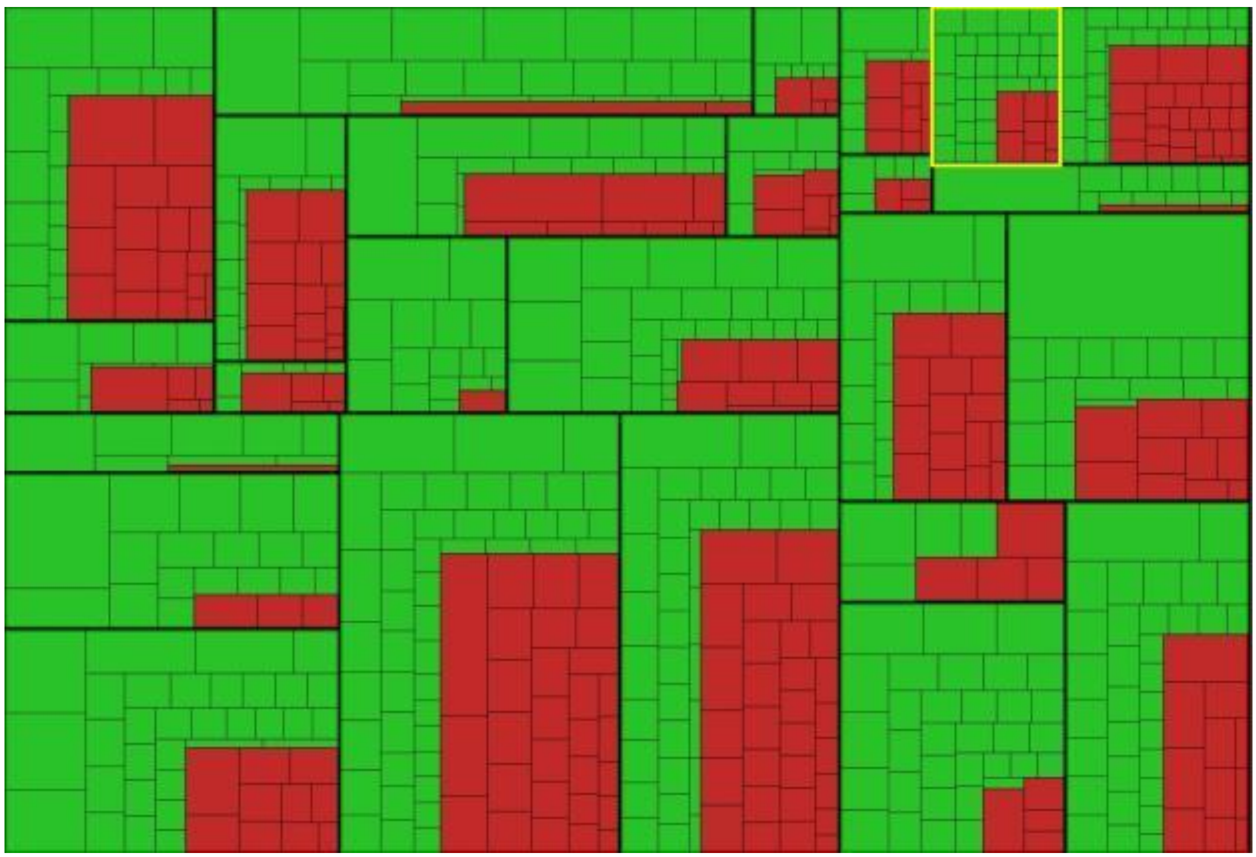


Figure 34. Attribute Grouping Layout for Experiment Trials

5.4 Results

The average completion time was *M* = 7.99 sec, (*SD* = 6.28 sec). Overall the mean absolute error for judgment was 8.1 (*SD* = 6.22). The significant main effects are shown in Table 9. None of the 2-way or 3-way interactions were significant.

The main effect of order was found to be significant for both absolute error and time. Order 2 (*M* = 7.61, *SD* = 7.67) yielded better judgment results than Order 1 (*M* = 8.53, *SD* = 8.71). Participants also recorded faster completion times for Order 2 (*M* = 7.61 sec, *SD* = 8.71 sec) than that of Order 1 (*M* = 8.37 sec, *SD* = 9.86 sec). The main effect of layout was also significant for absolute error. Participants made more accurate judgments for the attribute grouping layout (*M* = 7.57, *SD* = 7.33) compared to the squarified layout (*M* = 8.57, *SD* = 8.99). The impact of the main effects order and layout on absolute error are further examined in Figure 35, which plots the estimates as a function of answer for both variables.

Table 9. ANOVA Table of Absolute Error and Time ***Note*****: Significant at the p<0.05 Level in Bold.**

Effects	*Df*	Error Sig. (*p* =)	Time (sec) Sig. (*p* =)
Order	1	**0.0020509**	**0.0107948**
Layout	1	**0.0008405**	0.8880012
Sex	1	0.9099979	**2.176e-05**
Classification	1	0.4081782	**9.061e-07**

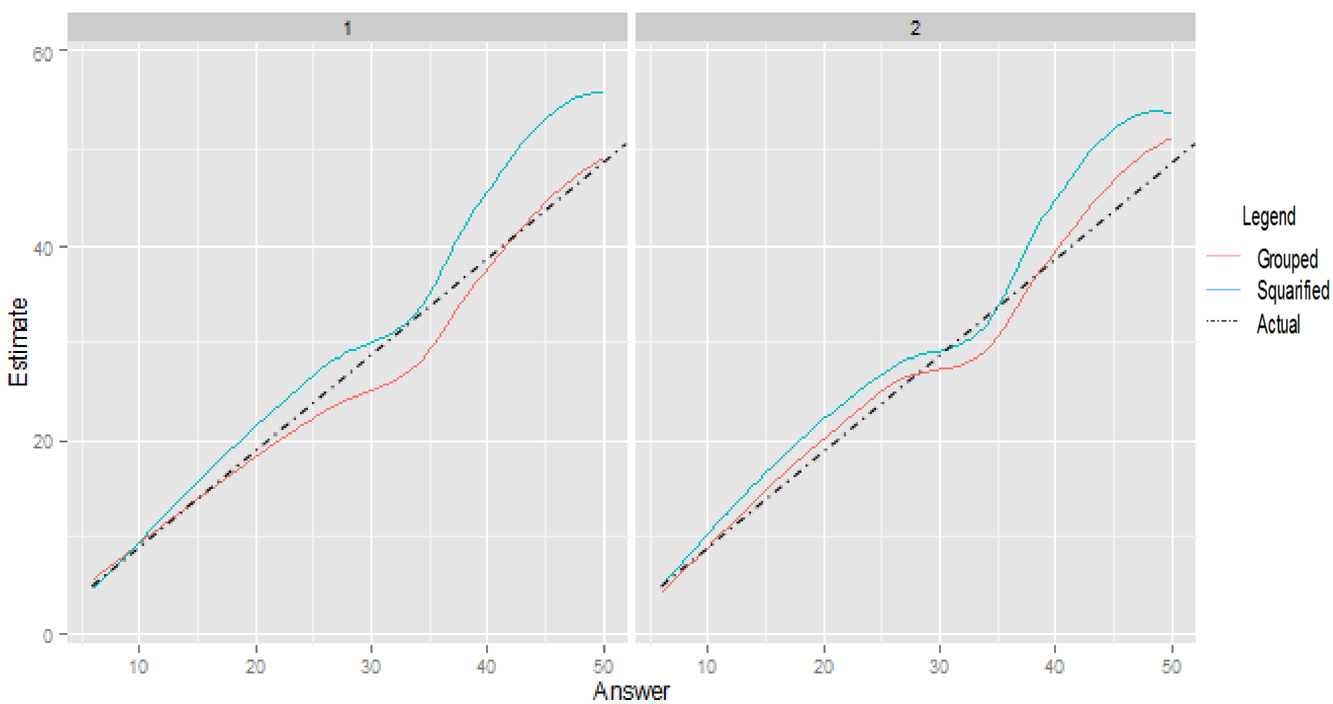


Figure 35. Estimate as a Function of Answer for Layout and Order

Figure 35 shows graphically that for both orders, participants are better able to estimate the correct answer when viewing grouped treemaps. ("Grouped" estimates are more aligned with the correct answer, represented by the diagonal dashed line, labeled as "Actual" in the figure). With the squarified layout there is a tendency for participants to overestimate the percentages. This overestimation increases when the actual answer increases in value (Answer > 40). For the squarified layout when the answer is greater than 40, the average absolute error for participants is 13.19 (SD = 8.22). For Order 1, there is more deviation from the correct answer line, compared to Order 2. Post hoc tests showed that Order 1 produced statistically significantly higher absolute error results than Order 2, p=0.002337.

There was not a significant effect of layout on the dependent variable time. However, there was a significant effect of gender and classification on time. Females (M = 7.34 sec, SD = 6.56 sec) were significantly faster than males (M = 8.74 sec, SD = 9.79 sec) and undergraduates (M = 7.12 sec, SD = 5.63 sec) were significantly faster than graduates (M = 8.98 sec, SD = 10.40 sec).

5.5 Discussion

This chapter has introduced an attribute grouping treemap layout algorithm that clusters nodes by color in order to aid users when making proportional judgments when examining treemaps. To validate the effectiveness of this algorithm we conducted a user study to compare the attribute grouping layout to that of the squarified layout. We found that the attribute grouping layout led to significantly better proportional judgment results than the squarified layout. This aligns with perceptual research that found that elements that are in closer proximity instead of randomly distributed clusters are easier to judge the total percentage of that particular element [11, 71].

It was also found that individuals tend to overestimate percentages for the squarified layout compared to the attribute grouping layout. This tendency increased for percentages that were greater than 40%. The results of overestimating for the squarified layout coincides with research by Goldstone [67] who conducted a study where subjects were asked to estimate the percentage of display items that had a particular feature. Features were either randomly distributed or spatially clustered so that the features of the same type tend to be close. Results from the study found that the subjects systematically overestimated the prevalence of features in randomly clustered displays.

Relative to the order in which participants saw the experiment trials, surprisingly we found that order had an impact on the absolute error and completion time. Participants who were given Order 2, where they saw the attribute grouping layout first followed by the squarified layout had better proportional judgments and faster times. The pattern of results is best explained by a training mechanism for proportional judgments.

Since the attribute grouping algorithm renders better proportional judgments, when seeing these trials first the like color nodes are already grouped together so the user does not have to cognitively group the colors together. It is reasonable that this helps train individuals, once they see the squarified layout they are accustomed to seeing the layout grouped together by color and are able to make a better judgment estimate compared to when individuals examine the squarified layout first they have to mentally group all the like colors together before making a judgment, which takes up more information processing steps and can lead to incorrect proportional judgments that carry over when examining the attribute grouping layout.

In general, proportional percentage judgments while examining treemaps are a difficult task and can lead to low accuracy. We can conclude that having alternative treemap layouts can impact ability of decision makers to make relative proportional judgments. While utilizing treemaps to analyze large hierarchical data can help support fast characterization of the data, they may not be well-suited to making inferences on proportional judgments utilizing traditional layouts. The use of an attribute grouping layout that clusters like nodes together enable more reliable and accurate judgments of the proportion of rectangles within a set of nested rectangles.

Chapter 6. Study 3: Using Treemaps to Visualize Healthcare Quality Data

This chapter describes applying treemap visualizations to real world data to test the effectiveness of utilizing this data visualization to analyze a large hierarchical data set. One area where treemaps can be useful in displaying information is in the medical arena. This concept of representing medical data using treemaps has not been widely explored. Baehrecke et al., utilized treemaps to visualize microarray and gene ontology data [77]. With the recent trend of employing electronic medical records (EMRs) for patient information, treemaps are potentially useful for health care professionals to assess trends amongst large populations of patients. Traditional visualizations used to display medical data such as bar charts, histograms, pie charts, and box plots are limited when displaying large quantities of data with several variables. If there is an outlier in the data or large variance, proportions can be thrown off in these displays. These visualizations are often inadequate when there is a need to display multiple variables and observe the interactions between the variables, and there is often not enough space to display relevant information.

In order to gauge whether treemap displays would be more beneficial, we investigated whether the visualization can help surgeons and other health personnel make assessments about their patients. A human subjects experiment was conducted applying both the traditional ("squarified") treemap layout and the novel attribute grouping layout discussed in Chapter 5 to surgical quality data to validate the usefulness of using treemap to interpret these type of data.

6.1 Introduction

Quality in healthcare has come to the forefront in recent years as an essential concept in providing adequate and efficient healthcare for patients. The Institute of Medicine (IOM) has defined quality of healthcare as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" [78]. An important aspect of measuring quality in healthcare is directly attributed to healthcare data. For healthcare organizations, data is central to both effective healthcare and to financial survival. Data about the effectiveness of treatment, accuracy of diagnoses, and practices of healthcare providers is crucial to maintain and improve healthcare delivery. High quality data effectively satisfies its intended use in decision making and planning.

In maintaining the standard for effective data quality in healthcare, the American College of Surgeons created the National Surgical Quality Improvement Program (ACS NSQIP) database to help with quality data reviews. The program employs a prospective, validated database to quantify 30-day risk-adjusted surgical outcomes to measure and improve the quality of surgical care, which allows valid comparison of outcomes among hospitals [79]. ACS NSQIP collects data on 136 variables, including preoperative risk factors, intraoperative variables, and outcomes for patients undergoing major surgical procedures in both the inpatient and outpatient setting. Hospitals utilize the NSQIP database for continuing education, quality improvement, and research. The semi-annual summaries, provided by the ACS, assist in targeting problematic surgical outcomes.

Even though the NSQIP database is very effective, it is challenging to review vast amounts of quality data. Surgeons can be presented with individual, sectional, divisional and departmental level data so it may be difficult to quickly determine where the priority clinical issues exist. The current output for the database is in the form of traditional summary graphs that often have multiple pages of information. Clinicians have limited time to spend reviewing quality data, therefore they could use some means to visualize the entire data set to quickly get an overview of what is going on, with access to details as needed, without having to query the system repeatedly, similar to having a dashboard system that can quickly relay key information. Data visualization can meet the need of optimizing the ability of a clinician to summarize, synthesize and prioritize the large amounts of data available to them.

6.2 NSQIP Data and Treemaps

The goal of this research is to map data from the NSQIP database to treemaps to test the validity of using this visualization to analyze data quickly and efficiently. There are several healthcare quality questions that surgeons and health personnel utilize the NSQIP data to answer, that would translate to potentially being visualized with treemaps.

Table 10 gives an example of these use case questions for surgery data. The table also illustrates the type of data explored in each use case [80]. The questions range from simple monitoring and search questions to more intensive analysis questions involving prediction and finding new phenomena.

Table 10. NSQIP Database Use Cases; (LOS) = Length of Stay in the Hospital (Measured in Days)

Use Case	Data Type(s)
Search and compare patterns of outcomes and LOS by treatment method for certain patient populations.	Nominal; Ordinal
Search and compare patterns of LOS and adverse events to confirm or disprove expected outcome	Nominal
Search and compare patterns of outcomes, frequency, and discharge status	Nominal; Ordinal
Discover new phenomena, e.g., predict outcome of patients given their history and planned intervention.	Nominal; Ordinal; Ratio
Analyze the structure of the data, e.g., effect of missing values on efficacy of analyses	
Look for trends in the data such as the clustering, or distribution of data or case load comparison across surgeons	Nominal; Ordinal; Interval
Search for patients with abnormal labs to give a pre-operative intervention.	Interval
Compare patterns of performance and outcomes by surgeon (e.g., length of case and outcomes of surgeries)	Nominal; Ordinal
Search by surgeon or surgery type and compare LOS and outcomes.	Nominal; Ordinal

The last use case is a way for surgeons to compare their performance to their counterparts by looking at outcomes for each of the surgeons' patients based on their length of stay (LOS) days by surgeon group. To illustrate an example of how treemaps can visualize this particular question, Figure 36 shows a hierarchy of a surgery department. Within the surgery department, various types of surgeries are performed by different physicians on each patient. Each patient will have a multitude of attributes and information recorded during their surgery process, i.e. length of stay (LOS) (measured in days in the hospital) and outcomes.

The treemap visualization in Figure 37 shows the hierarchy starting with the type of the surgery highlighted at the top as Colorectal surgery, followed by the surgeon ID for each surgeon who performed colorectal surgeries, followed by the patients who received colorectal surgeries by that surgeon. Each rectangle thus represents a patient, with the size of the rectangles representing the LOS days for the patient and the color representing the outcome for the patient (green = survival, red = death). Thus, when LOS is encoded in this way, the treemap gives the surgeon with the largest total LOS for his set of patients the most visual real estate.

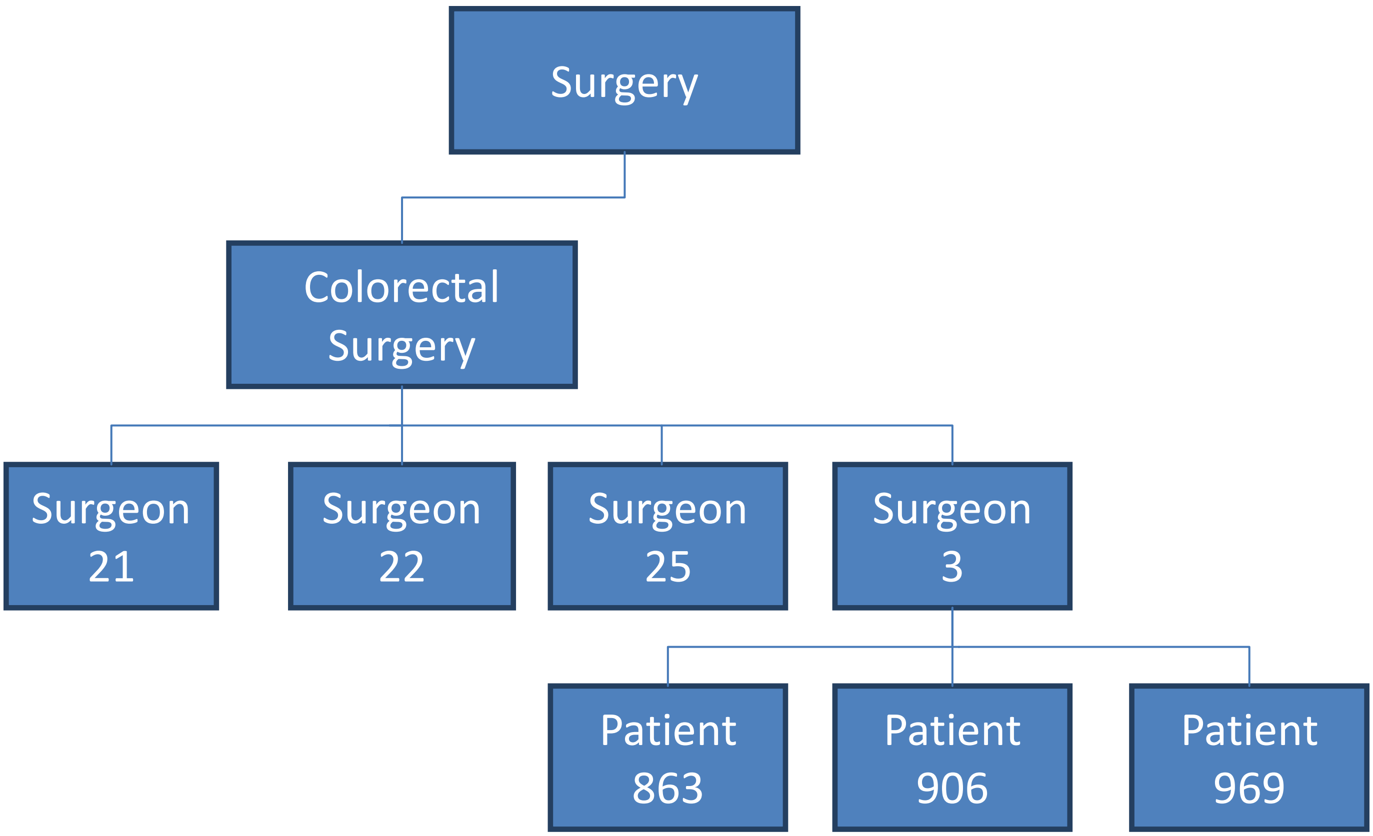


Figure 36. Hierarchy of Surgery Department, including Surgery Type, Surgeon, and Patient

Figure 37. Treemap Example of Surgery Hierarchy, including Surgery Type, Surgeon, and Patient

From Figure 37, we can see how many deaths occurred compared to survivors for each surgeon who performed colorectal surgeries. The question is how well an individual can estimate this percentage and how does this relate to the whole hierarchy. For instance how well can an individual estimate the percentage of red rectangles (deaths) that occurred in the colorectal group and for Surgeon 22 how much larger is their death percentage compared to other surgeons in the colorectal group? Being able to accurately judge these percentages or estimates is essential to help individuals analyze and interpret the data rendered. In this example, it is worth noting that the percentage of deaths indicates the number of squares of one color vs. another color, *independent of the size of the squares,* since the area (size) of the square represents length of stay. Thus, it is possible that one surgeon has a patient with a long length of stay who died, but who has a much smaller percentage of deaths (e.g., 1 out of 100 patients) than another surgeon who had several patients die (e.g., 5 out of 20 patients) but for those 5 patients, their lengths of stay were short, thus the total percent area painted red is smaller.

Previous research studies [81-83] evaluating the effectiveness of utilizing treemaps to visualize large data sets have consisted of subjects completing a series of tasks. In general the studies found that treemaps were easier to use and more effective to analyze and compare information when used in hierarchy exploration compared to other visualization tools. Examples of the tasks are shown in (Table 11).

Most of the tasks examined how well people could find extreme values in the hierarchy for a single data element. The tasks did not address how well people can understand and estimate the percentage the variable represents of the whole hierarchy, e.g. what percentage of colorectal surgeries result in death, to compare or find the extreme values for such percentages within the hierarchy, e.g., across multiple surgeons or surgery types.

The purpose of this study was to examine how viable treemap visualizations are in helping people analyze and interpret data when comparing groups of data elements to each other, comparing typical search and analysis tasks of the NSQIP database using alternative types of treemap visualizations.

Table 11. Prior Treemap Evaluation Tasks

Reference	Question
User Experiments with Tree Visualization Systems [81]	• Find the name of the parent directory of the directory "BMW” • Which directory has the greatest number of immediate subdirectories? • Locate the file labeled 1990.htm. • What is the name of the largest file in the eBay items hierarchy?
Extending Tree-Maps to Three Dimensions: A Comparative Study [82]	• Locate the largest file • Locate the largest file of a certain type • Locate the directory furthest down in the hierarchy structure • Name the most common file type
Usability Evaluation Method Applying AHP and Treemap Techniques [83]	• What is the most significant category? • Select all viewpoints in which system A total score is higher than that of B. • What is the most important check item?

6.3 Methods

6.3.1 Participants

One hundred twenty (120) undergraduate and graduate students of the University of Virginia volunteered for the study (n=56 males, n=64 females). The average age for participants was 23.6 (SD = 5.02). A protocol for this study was submitted and approved by the IRB for Social and Behavioral Sciences at the University of Virginia (# 2012-0422-00).

6.3.2 Apparatus

A computer-based color blind assessment derived from Ishihara Color Test assessed the red-green color deficiencies of each participant (Appendix D) [76]. Each participant performed the area judgments on a computer set to a 1024 x 768 pixel resolution. Participants completed the training and the experimental trials using a web-based custom apparatus utilizing a password protected server. Each of the responses for the participants were recorded using a MySQL database. The schema utilized to setup up the HTML script to the SQL database can be found in Appendix E.

6.3.3 Procedure

Each session lasted 20 minutes or less. After a briefing, each participant completed a test for color blindness (see Appendix D). The color blind assessment was used to ensure that the participants did not have any visual color deficiencies. Participants could not proceed with the study if they answered two or more questions incorrectly. Then each participant completed web-based judgment training and experimental trials.

Participants entered demographic information and then completed two proportional judgment training trials. For these trials, participants were given feedback regarding the accuracy of their judgments. The participants then completed 20 judgment task questions without feedback. Participants were encouraged to work quickly and to make “quick visual judgments.” In each trial, participants viewed a 1024 x 768 pixel image of a treemap layout that represented a subset of data from the NSQIP surgical database. The mapping of the variable to the NSQIP database can be found in Appendix F.

The surgeon hierarchy in the treemap starts with the surgeon groups (Bariatric, Colorectal, Hepatobiliary, and Vascular) → surgeon → patient. The color correlates to the outcome for each patient, where Red=Death and Green=Alive. The image contained either a squarified or grouped treemap layout. The grouped layout places rectangles of the same or a similar color adjacent to each other within each nested hierarchy in the treemap to help with making proportional judgments compared to the squarified layout (where nodes are ordered along some other dimension, such as patient ID).

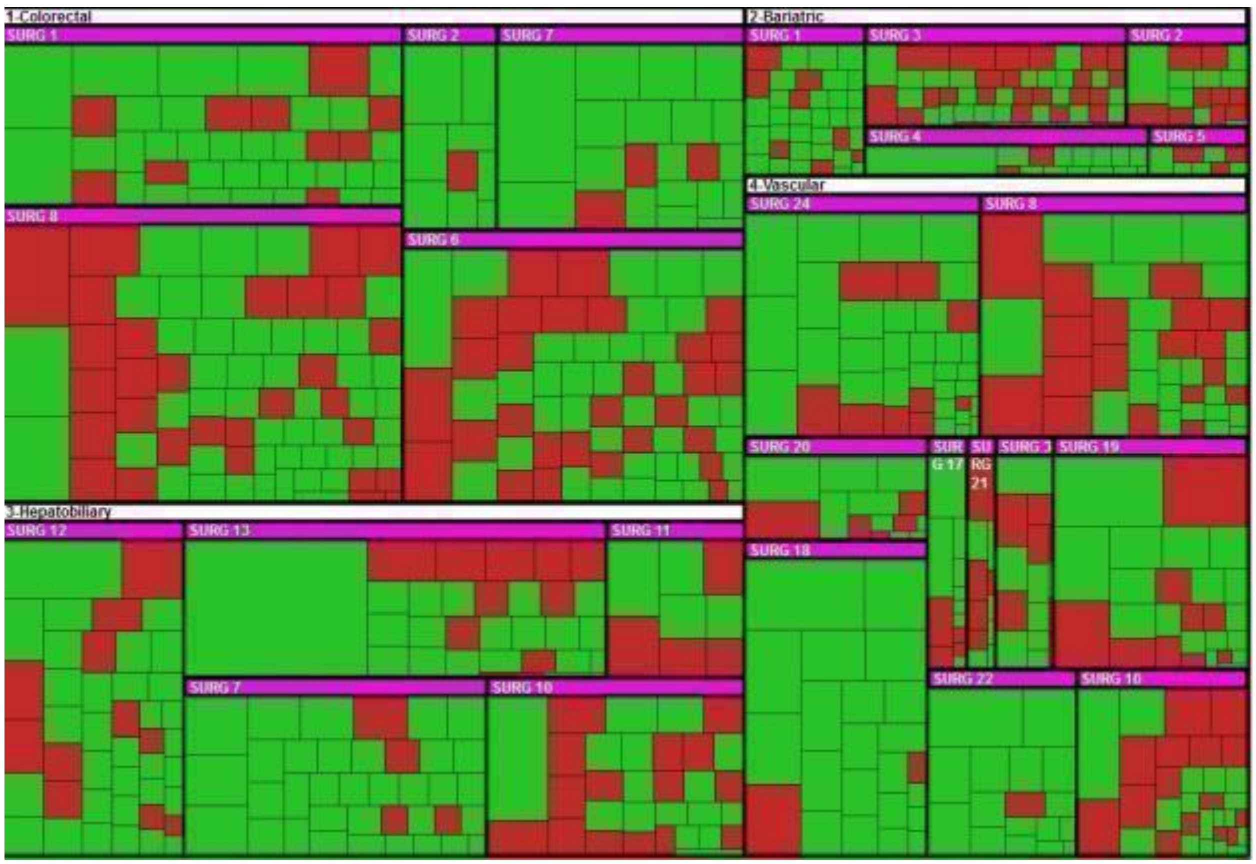


Figure 38. Squarified Treemap Layout Surgeon Hierarchy with Varied Node Size

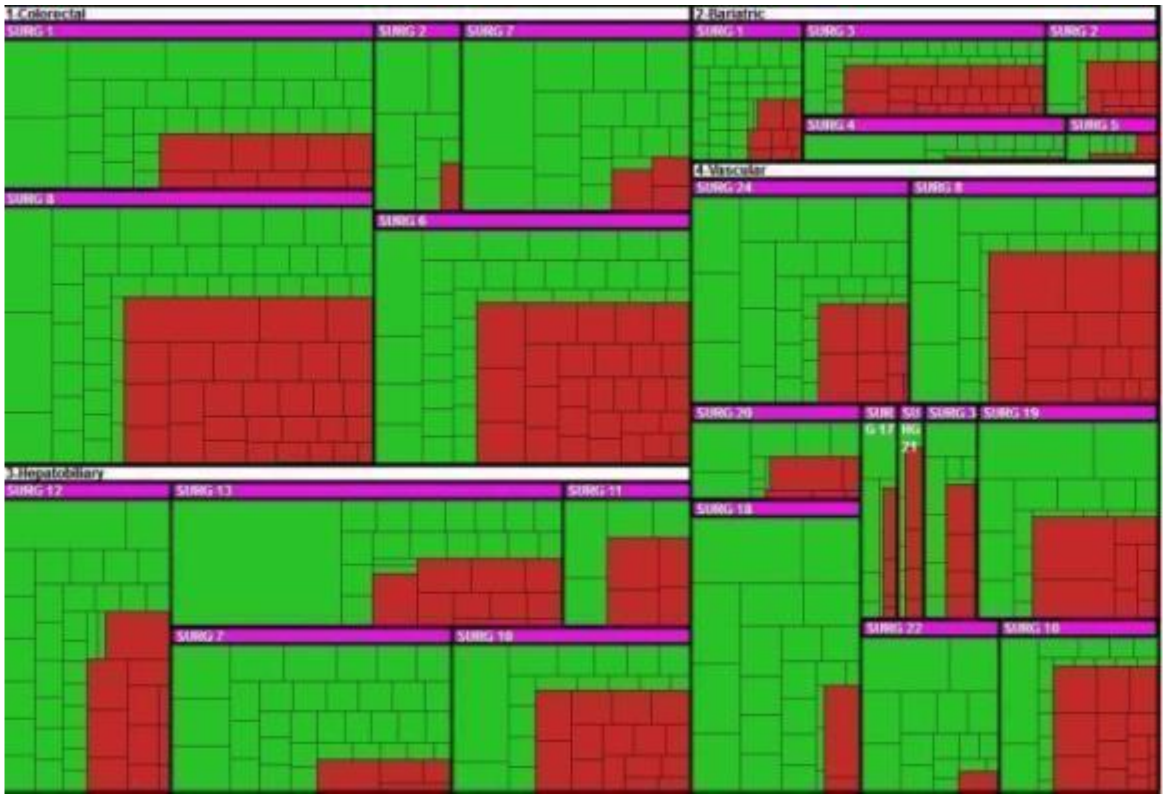


Figure 39. Grouped Treemap Layout Surgeon Hierarchy with Varied Nod Size

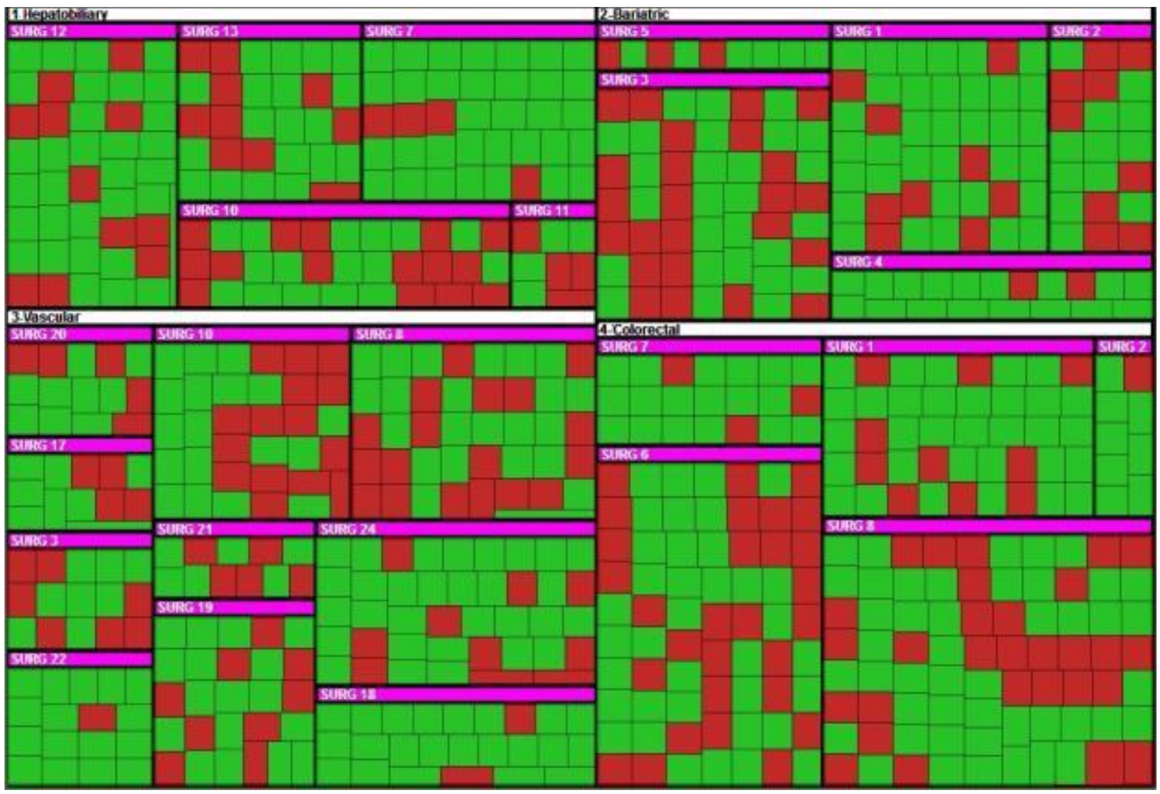


Figure 40. Squarified Treemap Layout Surgeon Hierarchy with Equal Node Size Represented

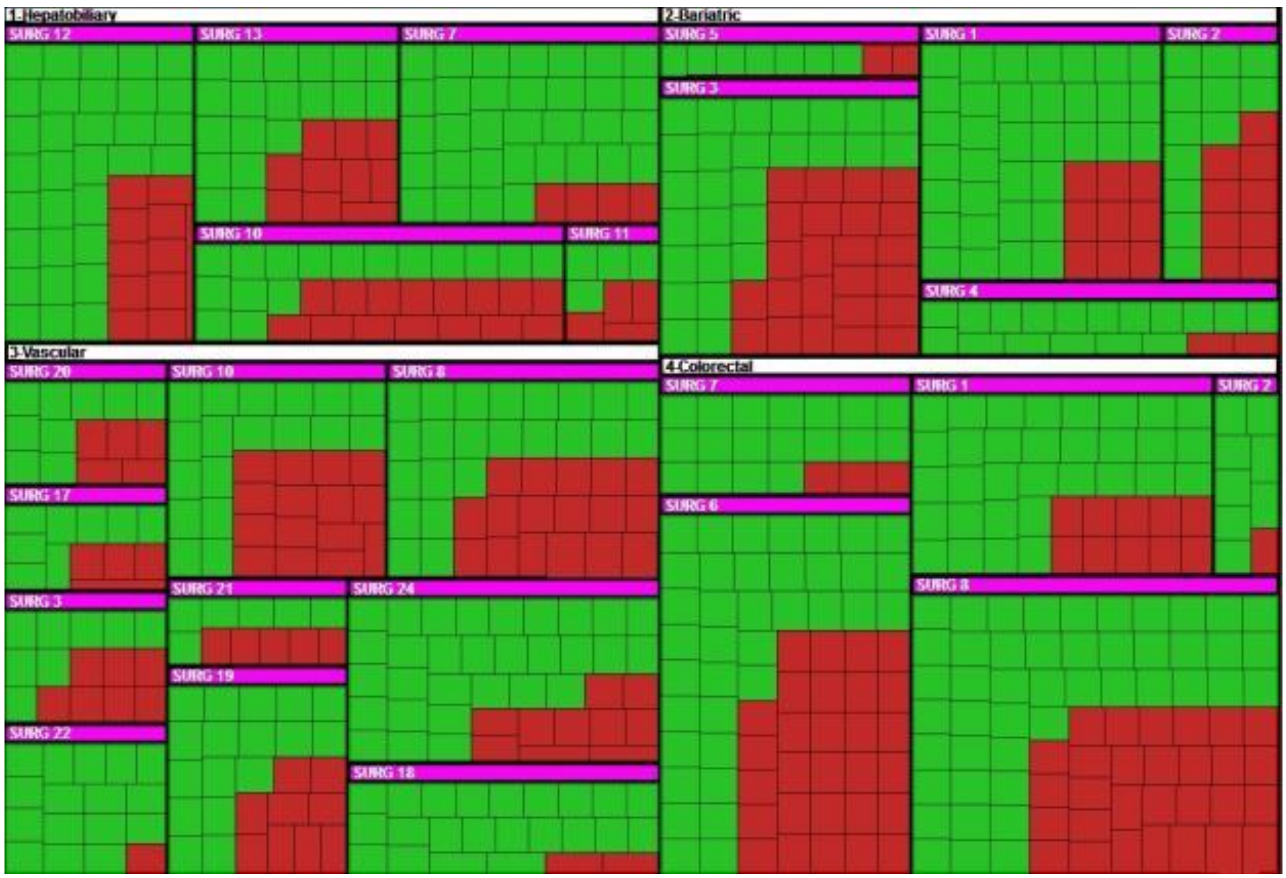


Figure 41. Grouped Treemap Layout Surgeon Hierarchy with Equal Node Size Represented

Table 12. Judgment Task for Surgery Data Experiment

Surgery Judgment Task
1. Which surgeon in the 2-Bariatric Group has the largest percentage of deaths?
2. Which surgeon in the 1-Colorectal Group has the largest percentage of deaths?
3. Which surgeon in the 3-Hepatobiliary Group has the largest percentage of deaths?
4. Which surgeon in the 4-Vascular Group has the largest percentage of deaths?
5. Does Surgeon 1 (“SURG 1”) have a larger percentage of deaths in the 2-Bariatric or 1-Colorectal Group?
6. Does Surgeon 10 (“SURG 10”) have a larger percentage of deaths in the 3-Hepatobiliary or 4-Vascular Group?
7. Does Surgeon 3 (“SURG 3”) have a larger percentage of deaths in the 2-Bariatric or 4-Vascular Group?
8. Does Surgeon 7 (“SURG 7”) have a larger percentage of deaths in the 1-Colorectal or 3-Hepatobiliary Group?
9. Does Surgeon 8 (“SURG 8”) have a larger percentage of deaths in the 1-Colorectal or 4-Vascular Group?
10. Which surgical group has the largest percentage of deaths?

The tasks completed in the experiment are shown in Table 12. Each task question was repeated for both layouts. The tasks were chosen as a representative sampling of the types of perceptual questions a user might encounter when analyzing the data. For all questions, the generic treemap interpretation required is to judge the total number of squares in one subsection that are of one color compared to the other color. This is a novel task compared to the tasks tested in previous treemap usability evaluations (Table 11). In this case, varying the size of the rectangles based on some attribute (in this case, LOS), is hypothesized to impede performance since it is systematically altering the total area seen of particular colors. We thus compared treemaps where the node size was varied (representing the changes in LOS) to those where the node size was equal (all sub-rectangles were of constant size).

The orders of the visualizations were counterbalanced, where participants were randomly assigned to two question orders. Half of the participants completed all of the squarified and grouped layouts with varied node size (Figures 38 & 39). The other half of the participants completed the trials with squarified and grouped layouts where node size was equal (Figures 40 & 41). All 120 participants completed 20 trial questions. Error and completion time was recorded for each trial. The error is coded as 0 if the question was answered incorrectly and 1 if the question was answered correctly. The completion time was measured from the time a question appeared on the screen to the time the subject pressed Enter on the keyboard.

6.3.4 Statistical analysis

All statistical analyses were performed using R version 2.15.2 [84]. Descriptive statistics were used to report measures of central tendency and variability for error and time on each independent variable. Main and interaction effects of the independent variables on the dependent measures were assessed. The effects of these within-subject and between-subject factors on the dependent measures were assessed using a generalized linear model (GLM) with the binomial variance function. There were a total of 20 trials per subject (n=120) for a total of 2,400 responses. Post-hoc analysis was used to identify significant differences between the levels of the effects. Results for the main and interaction effects are reported using $\alpha = .05$ for significance.

6.4 Results

Overall the average completion time per question was M = 24 sec, (SD = 17.7 sec). The main effects and the error percentages for correct responses and average completion for the different independent variables are shown in Table 13 and Table 14, respectively.

Table 13. ANOVA Table of Absolute Error and Time for Surgery Experiment ***Note*****: Significant at the p<0.05 Level in Bold.**

Effects	*Df*	Error Sig. (*p* =)	Time (sec) Sig. (*p* =)
Sex	1	0.1287	0.68479
Classifications	1	0.7139	0.19244
Order	1	0.2850	0.46578
Layout	1	**0.0505**	0.75087
Node Size	1	**0.0185**	**0.02445**

Table 14. Error Percentages and Average Completion Time for Independent Variables

	Squarified				Grouped			
	Varied Node Size		Equal Node Size		Varied Node Size		Equal Node Size	
Variable	% Correct	Time (µ,sec)	% Correct	Time (µ,sec)	% Correct	Time (µ,sec)	% Correct	Time (µ,sec)
Male	36%	24.1	55%	21.6	43%	26.1	66%	22.1
Female	36%	25.8	53%	21.4	47%	26.5	67%	21.8
Undergraduate	33%	26.8	49%	19.4	49%	27.4	67%	18.6
Graduate	39%	25.8	55%	23.2	41%	25.1	66%	22.4
Order 1	35%	24.8	45%	24.2	52%	25.5	65%	19.8
Order 2	33%	25.9	44%	26.1	50%	27.2	68%	20.1
Mean	35%	25.5	50%	22.7	47%	26.3	67%	20.8

The main effect of layout was found to be significant for absolute error, $p<0.0505$. The grouped layout (*M* = 57%) yielded better judgment results than the squarified layout (*M* = 43%). The main effect of node size was also significant for absolute error as well as time, $p<0.0185$ and $p<0.02445$, respectively. Participants made more accurate judgments for the equal node size layout (*M* = 59%) compared to the varied node size layout (*M* = 41%). Participants also recorded faster completion times for the equal node size layout (*M* = 21.75 sec, *SD* = 17.5 sec) than that of the varied node size layout (*M* = 25.9 sec, *SD* = 17.9 sec). No other main effects or interaction effects were significant on error and completion time. The trials which included the combination of the grouped layout and equal node size display overall had the highest judgment percentage and fastest completion time among all other combinations.

The analysis also investigated the error percentages for each task question. Table 14, includes the correct response percentages for all 20 task questions divided by the grouped and squarified layout. For the grouped layout participants had an average percentage of correctness over 58%. Participants had better judgment percentages for task questions examining one surgeon group and finding which surgeon had the largest percentage of deaths. For the squarified layout participants had lower judgment accuracy overall. For the squarified layout there was lower accuracy for trial questions that involved the participants comparing the percentage of deaths for a particular surgeon between surgeon groups. For both layouts, participants had better judgment accuracy for displays with equal node size compared to when the rectangle size was varied. The highest judgment accuracy was the combination of the grouped layout with the equal node size display, which produced an accuracy percentage as high as 78%.

Table 15. Error Percentages for Each Trial Question

	Squarified		Grouped	
	Varied Node Size	Equal Node Size	Varied Node Size	Equal Node Size
Questions	% Correct	% Correct	% Correct	% Correct
1. Which surgeon in the 2-Bariatric Group has the largest percentage of deaths?	12%	18%	53%	72%
2. Which surgeon in the 1-Colorectal Group has the largest percentage of deaths?	49%	35%	69%	75%
3. Which surgeon in the 3-Hepatobiliary Group has the largest percentage of deaths?	8%	70%	63%	52%
4. Which surgeon in the 4-Vascular Group has the largest percentage of deaths?	44%	75%	69%	70%
5. Does Surgeon 1 ("SURG 1") have a larger percentage of deaths in the 2-Bariatric or 1-Colorectal Group?	59%	32%	29%	75%
6. Does Surgeon 10 ("SURG 10") have a larger percentage of deaths in the 3-Hepatobiliary or 4-Vascular Group?	61%	72%	36%	67%
7. Does Surgeon 3 ("SURG 3") have a larger percentage of deaths in the 2-Bariatric or 4-Vascular Group?	46%	77%	36%	48%
8. Does Surgeon 7 ("SURG 7") have a larger percentage of deaths in the 1-Colorectal or 3-Hepatobiliary Group?	22%	75%	61%	78%
9. Does Surgeon 8 ("SURG 8") have a larger percentage of deaths in the 1-Colorectal or 4-Vascular Group?	20%	62%	34%	68%
10. Which surgical group has the largest percentage of deaths?	10%	37%	27%	70%
Mean (*M*)	33%	55%	48%	68%
Standard Deviation (*SD*)	0.21	0.22	0.17	0.10

6.5 Discussion

This chapter explored the use of treemap visualizations to help analyze NSQIP quality data. Although the final target audience would be surgeons or hospital quality personnel, this study utilized students as test subjects as it was a preliminary study that required a large group of subjects. It is presumed that knowledge of the domain is not a dominant requirement when making perceptual judgments as required by the experimental tasks. The study found that factors about the layout design affect judgment performance. One would typically think that mapping LOS to the size of the rectangles would be a logical approach when mapping surgery data characteristics onto the features of a treemap display, but this actually hindered performance when making other judgments where the size of the rectangles distort the visual perspective of how many people had a good or poor surgical outcome. When the areas are held constant, individuals can make a more accurate assessment of the data. Grouping nodes by color also helped improve accuracy. However, even utilizing these two features of treemap data, accuracy varied depending on question type and was not consistently high. The average judgment time to complete a task question was 24 seconds with an accuracy rate of approximately 68%. These findings imply that treemap visualizations may not be the most useful method to analyze surgical quality data or, more generically, the kinds of questions that require proportion judgments of the type studied here.

To date there are not any studies that have investigated utilizing treemaps to examine large surgical data sets as those found in the NSQIP database. This study served as a proof of concept to determine if treemaps could be beneficial in assessing surgical data retrospectively by allowing surgeons and healthcare administrators to make quick visual judgments. The treemap could be incorporated into a dashboard system that links key surgical data within the visualizations. Healthcare organizations are increasingly using dashboards to provide at-a-glance views of quality performance and decision-making [8]. Implementing the treemap within a dashboard system may improve upon judgment accuracy for quality questions in that the treemap would not have to be a silo system and can be part of an integrated system. This aspect is beneficial in that if a user finds a significant result from the treemap system they can utilize the integrated system to further investigate the phenomenon that was found.

Surgeons have a limited amount of time to spend reviewing data. Insights from our study illustrates that our application of treemap visualization may not be the right approach for a surgeons ability to summarize, synthesize, and prioritize large amounts of data available to them in a timely manner for questions involving comparing proportional data. Further research is required to find the optimal use cases that would benefit from treemap visualizations.

CHAPTER 7. CONCLUSIONS

The usefulness of data visualizations depend on the decision makers' ability to comprehend and interpret what information is being displayed graphically. In order for a data visualization to be effective developers must follow design principles that are derived from an understanding of human perception. This dissertation extends the literature on human perception and visualizations by investigating the impact of relative area and proportional judgments on hierarchical data visualizations.

The first aim of this study was to investigate relative rectangular area judgments. A comparative study experiment was conducted that investigated the effect of relative size difference (aspect ratio and area difference) and relative placement (distance between rectangles and offset angle) on the speed and accuracy of rectangular area judgments. It also considered the effect of combinations of aspect ratio and area difference where rectangles are contained within each other (observable fit) as compared to cases where rectangles do not fit (observable non-fit).

Contribution 1. Area judgments. In terms of relative size difference, the results validated prior perceptual research [48-49] that comparing 1:1 aspect ratio squares results in higher error and that people tend to underestimate area differences. Results from this study provide new insights that in rectangular area judgments, size (area difference) and location (offset angle) of rectangular stimuli affect judgment accuracy and speed. The results also suggest that the lowest error and shortest judgment time occurs for comparing pairs of rectangles when the smaller rectangle fits inside the larger.

Contribution 2. New factors to consider. Based on these results, we find that particular geometries of rectangular stimuli in area judgment result in people making systematic judgment errors. In general, relative area judgments of rectangular area comparisons result in low accuracy and do not support fast judgments. Results from this research give the visualization community additional factors to consider in designing treemap visualization to account for human perception issues. When creating visualizations where the user has to judge the area between two rectangular stimuli, the stimuli being compared must have a close spatial proximity, in terms of relative size and relative placement.

Contribution 3. Display groupings. Another finding of this research was exploring what type of display layout works best for making relative proportional judgments in hierarchical data visualizations. From the results, we can infer that the layout would be most effective if the stimuli being compared would have close spatial proximity. To ascertain whether this proposition improved visualization, a recursive attribute grouping algorithm based on lexicographical theory was developed, that clusters like attribute together in a treemap display to test this theory. An analysis was conducted to assess the impact of the attribute grouping layout to the traditional “squarified” treemap layout when making proportional judgments. The results showed that the attribute grouping layout increased accuracy and judgment time compared to the squarified layout. As such, the results support insights from human perceptual research that indicate that grouping items with like attributes aid in successful completion of perceptual grouping and proportion tasks in hierarchical data visualizations and provides a technical solution to provide such visualizations.

Contribution 4. Application to healthcare. Given the improved results from utilizing the grouping algorithm in accuracy and time among individuals in making relative proportional judgments, the alternative layout was applied to healthcare data to validate the effectiveness of the layout when making assessments about real world data. A controlled human subjects experiment was conducted to assess the ability of decision makers to make quick and accurate judgments on surgery data by visualizing a squarified treemap and attribute grouping treemap, with data hierarchically displayed by surgeon group, surgeon, and patient comparing each surgeons group of patients based on outcomes (dead or alive) and length of stay (LOS) days.

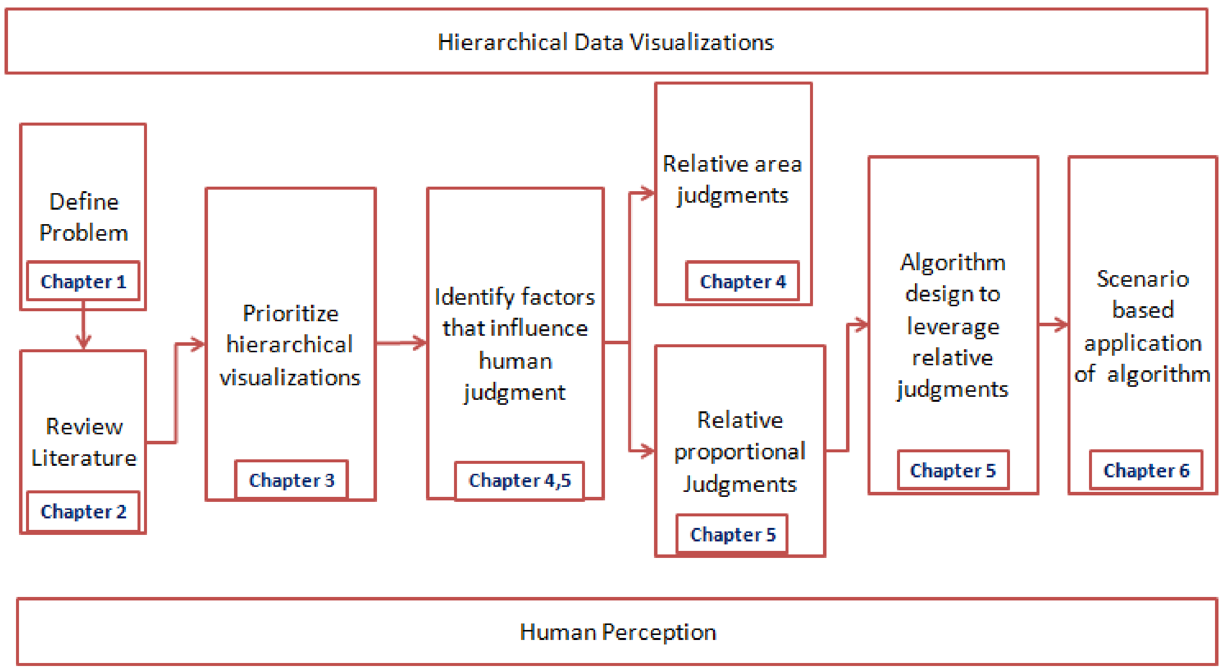
Contribution 5. Implications for surgery application. Judgment accuracy was not as high as expected, but was higher for the attribute grouping treemap compared to the traditional (“squarified”) layout. Performance was found to be impeded when utilizing the space-varying capability of treemaps when the size of large sub-rectangles impeded the ability to accurately judge the total percentage of rectangles of a certain attribute (represented by its color). This type of proportional judgment task is not restricted to surgery applications, and is likely a general problem with treemap interpretations, that they are not well-suited for this type of proportional judgment task.

The results suggest several avenues of future effort as follows.

1. Developing a perceptual model of the cognitive steps individuals utilize when making rectangular area judgments, including the heuristics people use when comparing two geometric stimuli that do not fit within each other.
2. Examining relative proportional judgments of hierarchical data sets using dynamic treemaps that include features such as drill down capabilities, corresponding detail displays, and the utilization of tool tips to help with accuracy of the proportional judgments.
3. Designing alternative hierarchical data visualizations that better support the types of proportional judgment tasks examined here.

APPENDICES

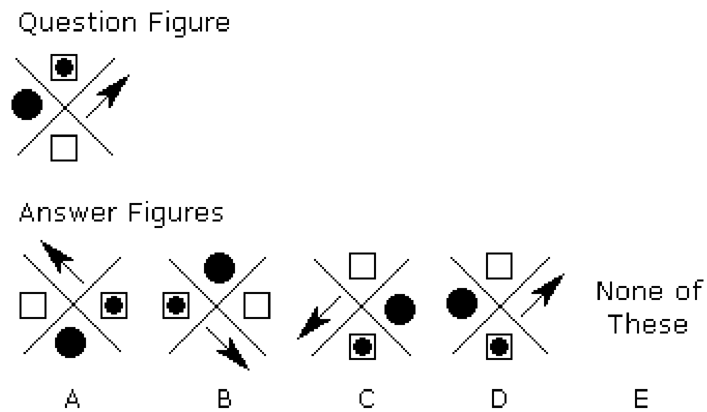
Appendix A. Conceptual Approach to Assessing Hierarchical Visualizations and Human Judgment



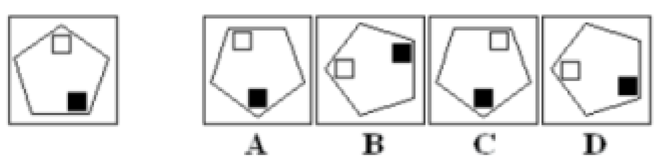
Appendix B. Spatial Ability Assessment

Instructions: Please complete each question to the best of your ability.

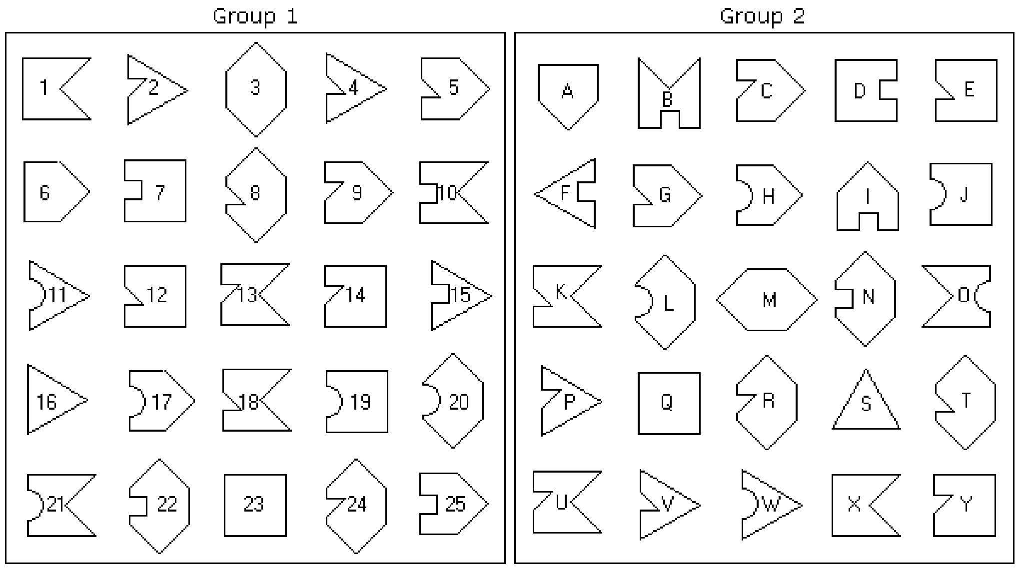
Which of the Answer Figures is a rotation of the Question Figure?



Which figure is identical to the first?



. Which shape in Group 2 corresponds to the shape in Group 1?



1)	2)	3)	4)	5)
6)	7)	8)	9)	10)
11)	12)	13)	14)	15)
16)	17)	18)	19)	20)
21)	22)	23)	24)	25)

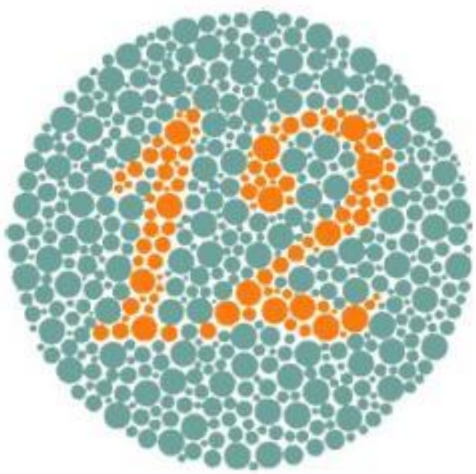
Appendix C. Attribute Grouping Code Scheme

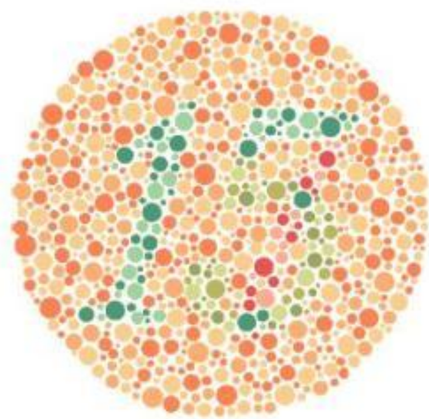
```
<fx:Script>
	<![CDATA[
		import com.ruwan.data.xml.TreemapLeafDO;
		import com.ruwan.data.xml.XMLLoader;
		import com.ruwan.data.xml.XMLProcessor;
		import com.ruwan.event.CustomItemEvent;
		import com.ruwan.treemap.AbstractDataFromatter;
		import com.ruwan.treemap.NoGroupDataFormatter;
		import com.ruwan.treemap.SingaleGroupDataFormatter;
		import com.ruwan.treemap.TreeMap;
		import com.ruwan.treemap.TreeMapItem;
		import com.ruwan.treemap.TwoGroupDataFormatter;
		import com.ruwan.treemap.interfaces.IDataFormatter;

		import customitems.ToolTipComp;

		import mx.collections.ArrayCollection;
		import mx.controls.Alert;
		import mx.core.FlexGlobals;
		import mx.events.FlexEvent;

		import spark.events.IndexChangeEvent;

		private			var map:TreeMap;
						//Parent of all Nodes
		private			var dataTooltip:ToolTipComp ;
					//Data Tooltip

		public static		var	GRADIENT_ENABLED:Boolean	=
		false;				//Gradient
		public static		var	SHOW_LABELS:Boolean =
		false;				//Gradient
		public static		var	NUMBER_OF_GROUPS:int = 2;
		//Number of Hierarchy Level
```

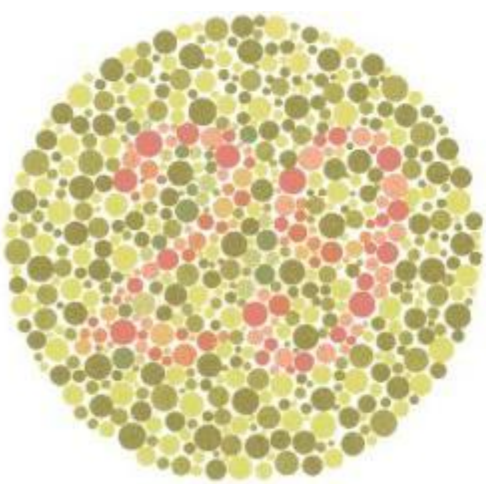
```
public static      const  LAYOUT_SQUARIFIED:String  =
       'layoutSquarified'; //Layout Algorithm, Squarified
or Slice and Dice
public static      var    GROUP1:String
       = 'surgeonGroups';        //Hierarchi level 1, Can
set to any field
public static      var    GROUP2:String
       = 'surgeonId';            //Hierarchi level 2, Can
set to any field
public static      var    WEIGHT:String
       = 'los';                         //Weight Field,
Must be a Numeric field
public static      var    CURRENT_LAYOUT:String
       = 'layoutSquarified';     //Current Active Layout
public static      var    COLOR_FIELD:String
       = 'outcome';              //Filed name use to define
the color
public static      var    SORT_ON:String
       = 'sortOn1';              //Field name use to define
the Sorting order
public static      var    ELEMENT_DICT:Dictionary;


protected function creationComplete(event:FlexEvent):void
{
       ELEMENT_DICT               = new Dictionary()
       dataTooltip                =  new ToolTipComp();
       dataTooltip.visible        = false;
       var xmlLoader:XMLLoader = new XMLLoader();

       cmbLayout.dataProvider     = new ArrayCollection([
              {label:'Square',           data:'sortOn1'},
              {label:'Clustered',        data:'sortOn2'}
       ])
       cmbHierarchy.dataProvider = new ArrayCollection([
              {label:'Surg. Group->Surg.ID', data:2},
              {label:'Surg. Group',      data:1},
              {label:'No Hierarchy',     data:0}
       ]);

       cmbColorField.dataProvider = new ArrayCollection([
              {label:'Outcome',          data:'outcome'},
              {label:'Sex',                     data:'sex'},
              {label:'State',            data:'state'}
       ]);

       cmbLayout.selectedIndex    = 0;
       cmbHierarchy.selectedIndex       = 0;
       cmbColorField.selectedIndex = 0;

       xmlLoader.load('data.xml', dataHandler);
              // Data Loading;
       labelHolder2.addElement(dataTooltip);
              // Add Tooltip
```

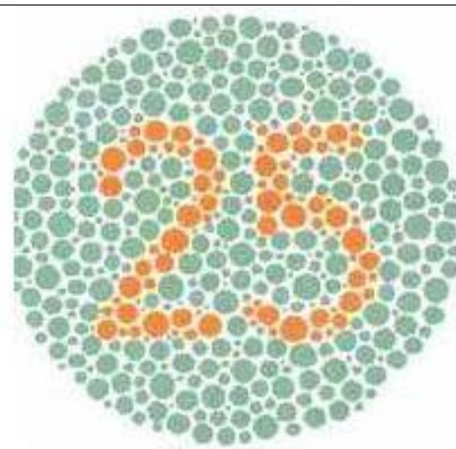
```
        FlexGlobals.topLevelApplication.addEventListener(
               MouseEvent.MOUSE_MOVE, mouseHandler, false );
        this.addEventListener('CUSTOM_ITEM_ADDED',
               customItemAdded);
}
```

Appendix D. Color Blind Assessment

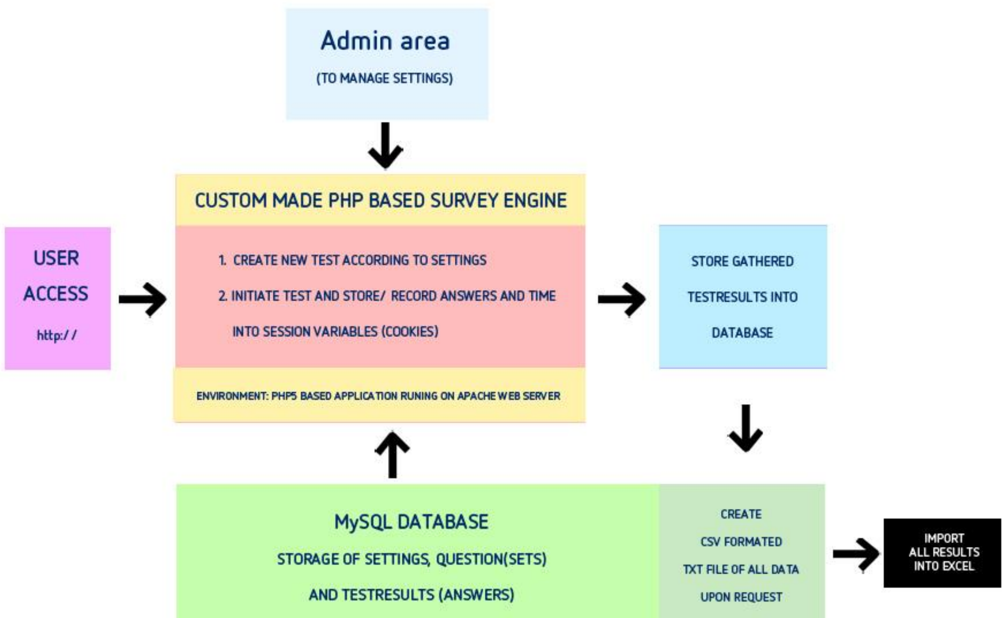








Appendix E. Schema for Experimentation Apparatus



Appendix F. NSQIP Data Tables

<table>
<tr><th>Variable</th><th>Classification in NSQIP Database</th></tr>
<tr><td>Case number/Patient Number</td><td>Studies Table → CASENUM</td></tr>
<tr><td>Length of Stay (LOS)</td><td>Studies Table → Discharge_Date-Admit_Date</td></tr>
<tr><td>Outcomes</td><td>PostOcc Table → OutCome</td></tr>
<tr><td>Complications</td><td>PostOcc Table → Occurrence
* Focusing on occurrence 27 (SUPERFICIAL INCISIONAL SSI) and 30(URINARY TRACT INFECTION)</td></tr>
<tr><td>Surgeons</td><td>NCOP Table → Surgeon_ID</td></tr>
<tr><td>Surgery Type</td><td>Studies Table → CPT
<table>
<tr><th>Surgery Group</th><th>CPT Codes</th></tr>
<tr><td>Vascular</td><td>34001-35907</td></tr>
<tr><td>Hepatobiliary/Pancreas</td><td>Liver: 47120-47130
Pancreas: 48105-48999</td></tr>
<tr><td>Endocrine</td><td>60210; 60220-60650</td></tr>
<tr><td>Bariatric</td><td>43644; 43770; 43846; 43847; 43775</td></tr>
<tr><td>Colorectal</td><td>44204, 44205, 44206, 44207, 44208, 44210 ;44140-44160
* Do not include 44125 or 44130</td></tr>
</table>
</td></tr>
</table>

Appendix G. Related Scholarly Activities

Below is a synopsis of research projects and experiences that I was involved in during my graduate school tenure.

Usability Testing for a Compact Epinephrine Auto-Injector

One of my past research foci examines human factors and medical devices. Medical device misuse is an important cause of medical error and therefore, incorporating human factors methodology into the design of medical devices has assumed an important role in ensuring patient safety. The research involved usability testing of an FDA investigational epinephrine auto-injector sponsored by Intelliject, LLC. The goal of the research was to measure the usability and user-based assessment in a head-to-head comparison between two different versions of the investigational product and competing devices; EpiPen® & TwinJect®. This was part of an FDA Clinical Trial that evaluated the disparities between epinephrine auto-injector products and received feedback on the usability and design characteristics of the products using a scenario-based usability test. Two research publications resulted from this work:

Guerlain, S., Wang, L., Hugine, A. (2010) Intelliject's novel epinephrine autoinjector: sharps injury prevention validation and comparable analysis with

EpiPen and Twinject. *Annals of Allergy, Asthma & Immunology* , 105:6, 480-484. PMID: 21130387

Guerlain, S., Hugine, A., Wang, L. (2010) A comparison of 4 epinephrine autoinjector delivery systems: usability and patient preference. *Annals of Allergy, Asthma & Immunology*, 104: 2, 172-177. PMID: 20306821

Institute for Collaborative Innovation (ICI)

The Institute for Collaborative Innovation (ICI) is a consortium of researchers that pools expertise in cognitive systems engineering, political science, design, cognitive science, field research, perception, and computer science to solve problems. I had the opportunity to participate in this externship opportunity which was sponsored by the Cognitive Systems Engineering Laboratory at Ohio State University in Columbus, OH. The research team that I was involved in focused on uncertainty and predictive analysis. Specifically, my focus was on visualizing and representing uncertainty in disaster relief efforts. During the summer, I developed design seeds to help visualize task reallocation with the emergence of side effects (unanticipated or unwarranted events) in disaster relief efforts. Also, I created tools and approaches that aid analysts in evaluating the uncertainty associated with multiple sources of data, including instituting a Ranking, Assessment, & Confidence (RAC) framework to help analyst visualize uncertainty in sources. This work was presented at the Converging Perspectives on Data (CPoD) consortium at the Ohio State University.

Usability Analysis of an Innovative Digital Reading Room

Reading room designs can have a major impact on radiologists' health, productivity and accuracy in reading. Several factors must be taken into account in order to optimize the work environment for radiologist. This research evaluated alternative workstations in a real-world testbed space, using qualitative data (users' perspectives) and quantitative (musculoskeletal perspective) data to measure satisfaction with the lighting, ergonomics, furniture, collaborative spaces, and radiologist workstations. In addition, we investigate the impact that the added collaboration components of the future reading room design, has on increasing team communication and coordination efforts in the reading room environment. By evaluating a fully functional testbed, health care administrators can examine potential problems with the testbed design before actually implementing the designs on a larger planning infrastructure. Results of this work led to conference presentations as well as a journal publication:

Hugine, A., Guerlain, S. (2011) User evaluation of an innovative digital reading room. *SIIM Conference*, Washington, DC.

Hugine, A. (2010). Usability Evaluation of Reading Room Testbed. *Radiology QA Meeting*, University of Virginia Health Systems, Charlottesville, VA.

Hugine, A., Guerlain, S., Hedge, A. (2011) User evaluation of an innovative digital reading room. *Journal of Digital Imaging*, PMID: 22080291.

REFERENCES

[1] "Open Framework, Information Management Strategy & Collaborative Governance | Data & Social Methodology - MIKE2.0 Methodology." Web Blog post. *MIKE2 Methodology RSS*. N.p., n.d. Web. 12 Feb. 2013.

[2] Rooney, B. "Big Data Is on the Rise, Bringing Big Questions", *The Wall Street Journal*. 29-Nov-2012.

[3] NSF, "Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)", *NSF Program Solicitation* 12-499.

[4] Tufte, E.R. *The Visual Display of Quantitative Information.* Cheshire: Graphics Press, 1983.

[5] Shneiderman, B. "Tree visualization with tree-maps: 2-d space-filling approach," *ACM TOG*, 11 (1992): 92–99.

[6] Bruls, D.M., Huizing C., and Van Wijk, J. "Squarified treemaps," *In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization*, (1999): 33–42.

[7] Vliegen, R., Van Wijk, J., and Der Linden, E. "Visualizing Business Data with Generalized Treemaps." *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006): 789-96.

[8] "SmartMoney - The Wall Street Journal." *SmartMoney*.12-Oct-2011.

[9] Wang, J., Qizhen, H., Yongxin, L., Chunfeng, Z. "Visualizing Air Traffic Flow Management Alert Information Using Squarified Treemaps." *Computer Graphics, Imaging and Visualization* 11.14 (2009): 419-422.

[10] Mansmann, F., Keim D.A., North, S.C., Rexroad, B., and Sheleheda, D. “Visual analysis of network traffic for resource planning, interactive monitoring, and interpretation of security threats.” *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007): 1105–1112.

[11] Humphrey, G., “The psychology of the gestalt.” *Journal of Educational Psychology*, 15.7 (1924): 401-412.

[12] Chintalapani, G., Plaisant, C., Shneiderman, B. “Extending the Utility of Treemaps with Flexible Hierarchy”, *Proc. International Conference on Information Visualization,* (2004): 335-344.

[13] Schamber, L. “A Content-Driven Approach to Visual Literacy: Gestalt Rediscovered”, *Paper presented at Association for Education in Journalism and Mass Communications*, Visual Communication Division, Norman, OK, 1986.

[14] Wickens, C. and Andre, A. “Proximity Compatibility and Information Display: Effects of Color, Space, and Objectness on Information Integration.” *Human Factors*, 32 (1990): 61-77.

[15] Cleveland, W.S. and McGill, R. “Graphical perception: Theory, experimentation, and application to the development of graphical methods.” *J. Am. Statistical Assoc.*, 79 (1984): 531–554.

[16] Stevens, S. S., “On the psychophysical law.” *Psychological Review* 64.3 (1957):153–181.

[17] Borg, I. and Leutner, D. “Dimensional models for the perception of rectangles.” *Perception and Psychophysics* 34 (1983): 257-269.

[18] Krantz, D.H. and Tversky, A. “Similarity of rectangles: an analysis of subjective dimensions”, *Journal of Mathematical Psychology* 12 (1975): 4- 34.

[19] Feldmann, J. “Mapping the Perceptual Space of rectangles’, *Perception* 27 (1998): 1191-1202.

[20] Regan, D. and Hamstra, S. “Shape discrimination and the judgment of perfect symmetry: Dissociation of shape from size.” *Vision Research* 32 (1992): 1845–1864.

[21] Morgan, M. J. “The visual computation of 2-D area by human observers.” *Vision Research* 45 (2005): 2564–2570.

[22] Nachmias, J. “Judging spatial properties of simple figures.” *Vision Research* 48 (2008): 1290-1296.

[23] Singer, P.A. and Feinstein, A.R. “Graphical display of categorical data.” *Journal of clinical epidemiology* (1993): 46.

[24] Menon, A. and Nerella, N. “Effective graphical displays.” *Pharm-Dev-Technol* 6.3 (2001): 477-84.

[25] Spriestersbach, A., Röhrig, B., du Prel, J.B., Gerhold-Ay, A., Blettner, M. “Descriptive statistics: the specification of statistical measures and their presentation in tables and graphs.” *Dtsch Arztebl Int.* 106.36 (2009):578-83.

[26] Kosara, R., Bendix, F., Hauser, H. “Parallel sets: interactive exploration and visual analysis of categorical data.” *IEEE Trans Vis Comput Graph* 12.4 (2006): 558-68.

[27] Whitley, E. and Ball, J. “Statistics review 1: Presenting and summarizing.” *Critical Care* 6 (2002): 66-71.

[28] Dugas, M., Kuhn, K., Kaiser, N., Uberla, K. “XML-based visualization of design and completeness in medical databases.” *Med Inform Internet Med* 26.4 (2001): 237-50.

[29] Ette, E. “Statistical graphics in pharmacokinetics and pharmacodynamics: a tutorial.” *The Annals of Pharmacotherapy* 32.7 (1998): 818-828.

[30] Pavlidis, P. and Noble, W.S. “Matrix2png: A Utility for Visualizing Matrix Data.” *Bioinformatics* 19 (2003): 295-296.

[31] Becker, R.A. and Cleveland, W.S. “Brushing Scatterplots.” *Dynamic Graphics for Statistics*, 1988.

[32] Andrews, K., “Visual exploration of large hierarchies with information pyramids.” *Proceedings of Sixth International Conference on Information Visualisation*. (2002): 793-798.

[33] Teoh, S.T. and Ma, L. “RINGS: A technique for visualizing large hierarchies.” *International Symposium on Graph Drawing* (2002): 268-275.

[34] Lane, D. *Hierarchy*, *Complexity*, *Society*. In Pumain, Denise. Hierarchy in Natural and Social Sciences. New York: Springer-Verlag, 2006.

[35] Agresti, A. *Categorical Data Analysis (2nd ed)*. New York: Wiley, 2001.

[36] Berman, J. “Maximizing project value: defining, managing, and measuring for optimal return”, *AMACOM Div American Mgmt Assn*, 2007.

[37] Shneiderman, B. and Wattenberg, M. “Ordered treemap layouts.” *In Proceedings of IEEE Information Visualization (InfoVis 2001)* (2001): 73–78.

[38] Johnson, B. and Shneiderman, B., “Treemaps: A space filling approach to the visualization of hierarchical information structures.” *Proceeding of the 2nd Conference on Visualizations,* (1991): 284-291.

[39] Coulom, R., “Treemaps for Search-Tree Visualization”, *The Seventh Computer Olympiad Computer-Games Workshop Proceedings*, Maasfricht. 2002.

[40] Asahi, T., Turo, D., Shneiderman, B. “Using treemaps to visualize the analytic hierarchy process.” *Technical Report CAR-TR-719*, Department of Computer Science, University of Maryland, College Park, MD, 1994.

[41] Freeman, L. “Visualizing Social Networks.” *Journal of Social Structure* 1 (2000).

[42] Day, L. and Dickinson, R. “Treemap Presentation as a Corporate Dashboard”, Corporate Performance Measures” *BNSF Railway,* 2007.

[43] Thian-Huat, O., Hsinchun, C., Wai-ki, S., Bin, Z. “Newsmap: a knowledge map for online news”, *Decis. Support Syst.* 39.4 (2005): 583-597.

[44] Shneiderman, B. and Plaisant, C. “Treemaps for space-constrained visualization of hierarchies.” Retrieved from http://www.cs.umd.edu/hcil/treemap-history/, 2008.

[45] Frantz, T.L. and Carley, K.M. “Treemaps as a Tool for Social Network Analysis”, *CASOS Technical Report*, September 2005

[46] Bederson, B. and Shneiderman, B. “Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies.” *ACM Transactions on Graphics* 21.4 (2002): 833-854.

[47] Van Wijk, J.J. and Van de Wetering, H. “Cushion Treemaps: Visualization of Hierarchical Information” *Proc. IEEE Symp. Information Visualization (InfoViz ',99)* (1999): 73-78.

[48] Heer, J. and Bostock, M. “Crowdsourcing graphical perception: Using mechanical turk to assess visualization design.” *Proceedings of the 28th International conference on Human factors in computing systems*. (2010): 203-212.

[49] Kong, N., Heer, J., Agrawala, M. “Perceptual Guidelines for Creating Rectangular Treemaps” *Proc. International Conference on Information Visualization*, 2010.

[50] Amazon’s Mechanical Turk [Online] Available: https://www.mturk.com/

[51] Hunter Lab, “CIE L*a*b* Color Scale”. Applications Notes, 8, 1-4, 2008

[52] Shneiderman, B. (n.d.) Treemaps for space-constrained visualization of hierarchies. Retrieved from http://www.cs.umd.edu/hcil/treemap-history/

[53] Cleveland, W.S. and McGill, R. ”Graphical perception and graphical methods for analyzing scientific data.” *Science*, 229 (1985): 828-833.

[54] Longjas, A., Legara, E.F., Monterola, C. "Power law mapping in human area perception", Int *J of Mod Phys C* 22.5 (2011): 495-503.

[55] Stevens, S.S. “The psychophysics of sensory function.” *American Scientist, 48* (1960): 226-253.

[56] Holmberg, L. and Holmberg, I. “The perception of the area of rectangles as a function of the ratio between height and width.” *Psychological Research Bulletin*, 9.3 (1969) 1-6.

[57] Shepard, R.N. and Metzler, J. “Mental rotation of three-dimensional objects.” *Science* 171 (1971): 701–703.

[58] Cooper, L.A. “Demonstration of a mental analog of an external rotation.” *Perception & Psychophysics*, 19 (1976):296–302.

[59] Johnson, A.M. “Speed of mental rotation as a function of problem solving strategies.” *Perceptual and Motor Skills*, 71 (1990): 803-806.

[60] Johnson, A.M. “Speed of mental rotation as a function of problem solving strategies.” *Perceptual and Motor Skills* 71 (1990): 803-806.

[61] Cave, K. R. and Kosslyn, S. M. “Varieties of Size-specific Visual Selection. Journal of Experimental Psychology.” *General* 2.18 (1989): 148-64.

[62] Anderson, N. H. *Foundations of information integration theory*. New York: Academic Press, 1981.

[63] Bosco, A., Longoni, A. M., Vecchi, T. “Gender Effects in Spatial Orientation: Cognitive Profiles and Mental Strategies” *Applied Cognitive Psychology*, 18.5 (2004): 519-532.

[64] Newton, P. and Bristoll, H. “Spatial Ability-Practice Test 1.” Internet: http://www.psychometric-success.com/. [May 5, 2011]

[65] Simkin, D. and Hastie, R. “An information processing analysis of graph perception.” *Journal of the American Statistical Association* 82 (1987): 454-465.

[66] Reuter, L., Tukey, P., Maloney, L.T., et al. “Human perception and visualization”, *Proceedings of the 1st conference on Visualization '90*, October 23-26, 1990, San Francisco, California

[67] Goldstone, R. “Feature Distribution and Biased Estimation of Visual Displays”, *Journal of Experimental Psychology*, 19.3 (1993): 567-579.

[68] Kaufman, E. L., Lord, M.W., Reese, T.W., Volkman, J. “The Discrimination of Visual Number,” *American Journal of Psychology*, 62 (1949): 498–525.

[69] Mandler, G. and Shebo, B.J. “Subitizing: An Analysis of Its Component Processes,” *Journal of Experimental Psychology: General*, 111.1 (1982): 1–22.

[70] Krueger, L. “Perceived Numerosity”, *Perception and Psychophysics* 11 (1972): 5-9.

[71] Wicken, C. and Carswell, C. “The Proximity Compatibility Principle: Its Psychological Foundation and Its Relevance to Display Design.” *Human Factors* 37.2 (1995): 473-494.

[72] Séroul, R. *Programming for Mathematicians*. Berlin: Springer-Verlag, 2000.

[73] Hanbury, A. *The Taming of the Hue, Saturation and Brightness Colour Space*. Rapport Technique N-28/02/MM, Paris, 2002.

[74] Agoston, M.K. *Computer Graphics and Geometric Modeling: Implementation and Algorithms*. London: Springer, 2005.

[75] Tu, Y., and Shen, H.W. “Visualizing changes of hierarchical data using treemaps.” *IEEE Trans. Visual Comput. Graphics* 13.6 (2007): 1286–1293.

[76] Ishihara, *Tests for color-blindness*, Hongo Harukicho, Handaya, Tokyo, 1917.

[77] Baehrecke, E. H., Dang, N., Babaria, K., and Shneiderman, B. “Visualization and analysis of microarray and gene ontology data with treemaps.” *BMC Bioinformatics*. 2004.

[78] Institute of Medicine I. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, D.C.: National Academy Press, 2001.

[79] ACS National Surgical Quality Improvement Program,[Online] Available: https://acsnsqip.org/main/default.asp

[80] North, C. “A Taxonomy of Information Visualization User-Interfaces” http://www.lirmm.fr/~infoviz/ASEval/References/shneiderman.php

[81] Bladh, T., Carr, D. A., Scholl, J. “Extending Tree-Maps to Three Dimensions: A Comparative Study.” *Proceedings of the Asia Pacific Conference on Computer-Human Interaction* (2004): 50--59.

[82] Kobsa, A. “User experiments with tree visualization systems” *In INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, (2004): 9-16.

[83] Toshiyuki A, et al. *Usability Evaluation Method Applying AHP and Treemap Techniques Human-Computer Interaction*. Berlin: Springer, 2009.

[84] R Development Core Team. “R: A language and environment for statistical computing.” *R Foundation for Statistical Computing*, Vienna, Austria.