Reading Numeric Data Tables: Viewer Behavior and the Effect of Visual Aids

by

YongFeng Ji
Bachelor of Science in Computer Science, University of Minnesota, Twin Cities, 2020

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

Reading Numeric Data Tables: Viewer Behavior and the Effect of Visual Aids

by

YongFeng Ji
Bachelor of Science in Computer Science, University of Minnesota, Twin Cities, 2020

Supervisory Committee

Dr. Miguel Nacenta, Supervisor
(Department of Computer Science)

Dr. Charles Perin, Co-Supervisor
(Department of Computer Science)

ABSTRACT

Data tables are one of the most common ways in which people encounter data. Although mostly built with text and numbers, data table representations have strong spatial components, and often exhibit visual elements meant to facilitate their reading. there is an empirical knowledge gap on how people read and use tables and how different visual aids affect people's ability to use them. In this work, I seek to address this gap through a series of empirical studies. I asked participants to repeatedly perform five different tasks with tables in four conditions where the table representations consisted of Plain tables, tables with Zebra striping, with cell background color encoding the value, and with background bar length in a cell encoding its value. I analyzed the gathered data in multiple ways to characterize human behavior when accessing tables and to assess the benefits of the different visual aids.

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

Supervisor Dr. Miguel Nacenta, for the invaluable guidance and encouragement in understanding academic thinking and reasoning. Your impact has been crucial in shaping my research path.

Co-Supervisor Dr. Charles Perin, for the thoughtful advice and consistent support in my academic journey. Your expertise and helpful critiques have been key in enhancing my research skills.

my parents, for their endless love and guidance. Your support has been the keystone of my life and accomplishments, always providing wisdom and support at every turn.

my girlfriend, Ellie Yang, for the endless love and support, offering a foundation of strength and encouragement. Your belief in me has been a significant source of motivation and resilience throughout this journey.

my cat, 98, for being a lovely companion during the extensive hours of research and writing. Your companionship has brought a sense of comfort to my academic journey.

Chapter 1

Introduction

Data plays a significant role in our daily lives, as highlighted by Cukier [21]. People strive to communicate and present information effectively through data visualization [78]. Given the importance of data presentation, it's essential to consider traditional methods, such as tables.

In the 1970s, tables were described as "nothing more than a systematic arrangement of information items" [88]. Initially, tables were primarily used to present numerical data in a matrix of rows and columns [47]. At that time, it was believed that "They are not recommended for communicating data to the general public; tables are most useful for fellow professionals. Both constructing and reading tables require a skill of a high order" [47]. However, fifty years later, our perception of tables has significantly changed, and their application has become much broader, now including the presentation of qualitative information. Tables today incorporate not only text but also numerical data and are utilized by professionals and millions of people daily. The widespread use of tables is exemplified by the democratization of spreadsheet software like MS Excel and Google Spreadsheets. This minimalist view of what tables are still holds true to some extent, especially in academic papers or the written press. For instance, researchers argue that tables must be space-efficient and convey a simple message [23]. Since tables are often used to simplify text verbosity, they can help reduce word count if designed efficiently and are not too large [30].

Some researchers argue that tables have limitations, such as difficulty in conveying large datasets [48, 82], displaying trends, and facilitating easy comparisons [47, 71, 54]. Therefore, when addressing questions involving comparison, trend identification, or mathematical analysis, a well-designed graph visualization might be a better choice [36, 37]. Graph visualization refers to the graphical representation of data using maps or other visual elements [49]. An effective graph visualization simplifies complex datasets, making patterns, trends, and insights more accessible and understandable [28]. Graph visualization aids in decision-making, storytelling, and communication by presenting information in a visually engaging and informative way [67]. Some graph visualization techniques include Bar, Color shade, Sparkline [75], and FatFont [56].

In summary, tables were initially seen as a structured way to present information and now serve a broader purpose, including the presentation of qualitative data. However, they do have certain drawbacks, as discussed earlier. On the other hand, graphs offer a different approach to data visualization, simplifying complex information and aiding in decision-making, storytelling, and value comparisons. To create effective visualizations, it's crucial to incorporate the strengths of both tables and graphs.

Encouraged by Hink's idea that a well-designed visualization, which combines features from tables and graphs, can greatly enhance visual perception [37], and drawing inspiration from Perin et al.'s Bertifier approach [59], which incorporates visual encodings into table cells, I examined the effects of incorporating various visual aids in table design. These visual aids are further categorized as **visual features** and **visual encodings**. My objective is to improve table readability by incorporating these visual aids into tables.

I categorize **visual features** as design choices that affect the table's appearance but do not directly encode data values. These include visual aids like zebra patterns, alternating shades or colors of rows to enhance visual distinction and aid in data interpretation [46]. Some of these visual features have been previously studied. For example, it has been found that using a zebra pattern or vertically aligning characters inside table cells reduces errors and eases comparison tasks [76]; a well-sized table, properly spaced within text, is beneficial [35]; design elements such as borders and shading are included [46]; dividing rows and columns by border lines or spaces is an important design aspect [30]; and that rows and columns should not be too far from each other [60].

On the other hand, **visual encodings** are visual aids that represent data graphically [70], inspired by Bertin's concept of "encoding table cells visually" [9]. For instance, MS Excel supports the application of color shading or the creation of data bars that directly encode values within cells by filling the cell's background with a color corresponding to the cell's value, or by adding a bar to the background whose length corresponds to the cell's value, respectively. Although such visual encodings have only recently appeared in mainstream software, they have been used in visualization for a long time. See [59] and visit https://aviz.fr/Bertifier/Bibliography for an annotated bibliography of the historical use of visual encodings for tables, including textures and patterns, Color shadings, Bars, and other visual marks and variables.

Figure 1.1: A: plain table (top left) B: table with zebra striping — visual feature (top right), C: table with Color data encoding (bottom left), D: table with bar data encodings (bottom right).

Surprisingly, though, despite how ubiquitous tables are, how visual aids are deemed important in helping people read tables, and how much research has been conducted on proposing table-specific visual encodings, little empirical evidence exists to characterize how people read tables and what the benefits of visual both visual features and visual encodings are.

I conducted an exploratory experiment using four distinct data tables; each had three different visual aids: Bar, Color, Zebra striping, and a Plain table as my control condition, as shown in Fig. 1.1. The principal goal of this experiment is to understand whether and how visual aids affect the reading of data tables by humans. A secondary goal is to achieve a better understanding of how people generally approach and read data tables.

1.1 Research Questions

To address these goals in Section 1, I formulate six main research questions, in two groups. In the first group, I consider a generic question:

Q0 How do people approach data tables when they first encounter them?

In the second group, I systematically address the principal goal by decomposing it into five questions (see Section 1 for the definitions of visual aids, features, and encodings)

Q1 Do visual encodings help read data tables?

Q2 Do visual features help read data tables?

Q3 Which of the two is more effective in facilitating the reading of data tables by reducing the completion time and minimizing errors: visually encoding data or using visual features like Zebra striping?

Q4 Which visual encoding provides the most benefit for reading data tables?

Q5 Does the movement of the mouse correlate with the ability to read data tables?

Each of these questions is cross-contextualized across two axes: the task (what the table is being used for) and the performance type (the outcome that I look at). I invited a total 27 participants (3 participants' data were discarded due to overtime and inaccurate eye tracking data) who did not report any visual impairments and had good eyesight to take part in my experiment. Participants wore eye-tracking glasses while completing various tasks using my self-developed software. After collecting all the data, I sought to address the questions Q0 and Q5 through qualitative analysis and questions Q1 to Q4 through quantitative analysis.

1.2 Scope

In this section, I define the research scope, which includes specifying the regulation of the table, data topic, data type, selection of visual aids, participant criteria, and limitations. More details can be found in section 3.

Table regulation To better replicate real-world conditions, each table was limited to 30 rows and 11 columns, resulting in a total of 333 cells. All tables maintained consistent dimensions, cell sizes, and font sizes. They were designed to be static, with no interactive features within the tables, in order to closely mimic real-world scenarios.

Data topic and type I chose topics related to cereal, candy, movie ratings, and animation ratings. These topics are not extremely unfamiliar but also not overly common. All data types in the tables were homogeneous, consisting of two-digit numeric values ranging from 10 to 99 with two significant digits, no decimal points, no negative numbers, and no textual or qualitative data. The data were intentionally randomized to prevent any bias due to familiarity. Each row name consisted of two words, and each column header attribute had one to three-word attribute names. This boundary was set to maintain consistent table dimensions and structures for all conditions.

Visual aid selection I chose to include Bar, Color shading (heatmap), and Zebra striping in the tables. Bar charts displayed variable-length horizontal bars in the background of each cell. Color shading, also known as a heatmap, encoded numerical values through the background color of each cell, with deeper colors indicating larger values. Zebra striping involves using a light gray color on the background of cells in alternate rows of data. These selections were made because these aids are commonly used in real-life scenarios, and testing additional aids would significantly complicate the design for a controlled lab experiment.

Tasks selection In terms of tasks, I carefully selected four:

- Vertical maximum value selection (TVertMax): Participants were asked to find the largest value in a column.
- Horizontal maximum value selection (THorMax): Participants were asked to finding the largest value in a row.
- Value retrieval (TValue): Participants had to find a specific value using given column attributes and row names.
- Multi-row attribute comparison (TDiff): Participants were required to compare two rows and find the column with the largest proportional difference among the last four columns. This task is considered complex.

These selections were made because they include two tasks involving comparison (to compare values), one task involving navigation (to search for a specific target), and one task that combines both comparison and navigation, which is considered complex. By choosing these four tasks, the experiment can be comprehensive and controllable. More detailed information can be found in Section 3.2.1

Measurements In terms of measurements, I considered completion time (where lower times indicate that a condition is more helpful) and error rate (where fewer errors indicate a more helpful condition). By measuring these two factors, I aimed to determine whether the visual aid can be beneficial in reducing response time and error rates. For more detailed information, please refer to Section 3.7.

1.3 Contributions

In this thesis, I make three main contributions:

- I explore how people read data tables and evaluate their common strategies when seeking answers within these tables.
- I investigate whether the utilization of visual encoding and visual features in data tables can help reduce completion time and minimize error rates by presenting an analysis of how visual encoding and visual features contribute to reading data tables and identifying which of these elements is the most useful.
- I explore the potential benefits of using a mouse as an external tool for reading data tables.

Overall, my work focuses on exploring how data tables with various visual aids can enhance human experiences and bridge the gap in existing research outcomes.

1.4 Organization of the thesis

The rest of the thesis is organized as follows: Chapter 2 delves into the history of data tables, discusses guidelines for effective table design, presents existing research on table enhancement, explores data visualization and various visual encodings, and delves into research on eye tracker analysis. Chapter 3 outlines my methodology, covering visual aid selection, task design, research hypotheses, experiment design, device and apparatus specifications, software and data collection methods, and my approach to data analysis which includes both quantitative and qualitative aspects. Chapter 4 reports the results and discussions for each specific task. In Chapter 5, I provide a comprehensive global discussion that encompasses findings across all tables and tasks, and addresses future works and limitations. Finally, Chapter 6 serves as the conclusion, summarizing the experiment procedure, highlighting key findings, and discussing contributions.

Chapter 2

Related Work

In this chapter, I explore the history and use of tables as a common data visualization method. I will cover the history of tables, their purpose, basic guidelines for creating effective tables, table enhancement related research, and briefly touch on graph visualization. Additionally, I will review existing research on the visual elements that I have incorporated into my tables. Finally, I will discuss the eye tracker analysis-related work.

2.1 Data Table

Modern data tables present information using text and numbers, employing visual elements like data alignment and spacing to facilitate readability. Datasets are organized in a row and column structure [29]. This format, known as a "spread-sheet structured table", has been in use for thousands of years, with its invention tracing back to the Ancient Sumerians [51]. Additionally, between 1900 BCE and 1300 CE, various types of tables were used, such as "Chronicles", which recorded events in chronological order, "Canon Tables", which aided in finding passages in the Gospels, and "Medieval Calendars", which organized dates and holidays in a tabular format [29].

In today's world, data tables are widely recognized as valuable tools for visualizing information in academic, business, and scientific settings. Unlike narrative text, which requires a fixed reading path, tables offer flexibility in reading direction due to their discontinuous, bounded, and gridded structures [52]. Wainer [80, 81] outlines four key purposes for table usage and design: exploration for answering questions, communication for effective storytelling with data, storage for archiving and data retrieval, and illustration as graphics to support narratives. To ensure a table's effectiveness, specific criteria need to be met.

Firstly, Wright [88] emphasizes that an ideal data table should systematically arrange information, meaning that data should be organized in a structured manner within the table to promote clarity and understanding.

Additionally, the choice of data type is crucial. Researchers suggest that numerical data is best suited for tables [47, 66], advising against presenting large sets of numeric data in tables due to potential complexity hindering effectiveness.

Furthermore, aesthetic design is also vital. Franzblau and Lauren [7] emphasize the importance of table size, borders, shading, and clarity, recommending consistency in the use of border lines or spacing to separate rows and columns [30].

In summary, effective data tables are characterized by a systematic arrangement, the choice of appropriate data types, and thoughtful aesthetic design considerations, including table size, spacing, and consistent formatting, to optimize data presentation and understanding.

2.2 Data table enhancements

Many researchers are striving to enhance data tables for better readability and performance. Wainer suggests that a table can also effectively visualize complex data [80]. Bertin proposes two key concepts to improve table readability: visually encoding table cells and grouping similar rows and columns to reveal patterns [9].

In 1994, Rao and Card introduced "Table Lens", a solution for large data tables. It utilizes the focus+context (fisheye) technique in tabular data visualization, allowing the presentation of crucial label information and accommodating multiple distant focal areas. This technique supports interaction with extensive information structures by dynamically distorting the layout based on the varying levels of interest in its components [62]. Inspired from Bertins' principles, Perin et al. have developed a web application that lets users interactively manipulate rows and columns while applying various visual encodings to table cells [59]. Their empirical study suggests that this application is valuable for data exploration and visualization. The other common method of interaction with a table is to have a tooltip that allows table to show additional information to increase the readability of a large table [16].

In addition to making tables interactive, Hink et al. proposed a novel visualization concept called "Grable". Grable combines features of graph visualization techniques with the precise quantities of a table in a single display [37]. Hink et al. conducted an empirical study comparing bar charts with bar Grables, line charts with line Grables, and pie charts with pie Grables. Subsequently, Bradstreet followed the guidelines from Hink and conducted several case studies to discuss the advantages and disadvantages of using Grable [11]. They found that Grable improved accuracy but could also slow down users' response time [38]. However, Bradstreet and Hink emphasize the importance of using both table and graph visualizations to enhance the effectiveness of data presentation [11, 38, 37].

2.3 Graph Visualization

The history of graph visualization can be traced back to ancient maps, which were early forms of visual representation of geographic data, helping people navigate and understand the world around them [15].

Early techniques like maps, charts, and diagrams served purposes in navigation, science, and business. In the 18th and 19th centuries, William Playfair introduced statistical graphs, and Florence Nightingale utilized diagrams for healthcare data, pioneering bar graphs and circular area charts [61, 58].

In the 20th century, Tufte emphasized clarity, simplicity, and precision in data graphics in his work "The Visual Display of Quantitative Information" [76], advocating for the minimization of unnecessary elements. In Cole Nussbaumer Knaflic's book [43] and Stephan Few's book [28], both authors explore the use of data visualization as a storytelling tool, discussing building narratives around data and guiding audiences through logical and compelling data-driven stories. This is particularly beneficial for business professionals and communicators aiming to convey insights effectively.

Therefore, effective graph visualizations should encompass clear and visually organized data, appropriate visual elements, comprehensive labeling and titling, and suitable choices of color, font, and style, making it easy for users to understand [74].

2.4 Visual Features

Visual features refer to the visual elements or characteristics used in data visualization to represent and convey information. These features are employed to make data more understandable and accessible to the audience [84]. Zebra-striping and font manipulation, including typeface, font size, font style, and font color, are common visual features used within tables.

Zebra-striping The Zebra-Striping tables have a common visual aid that is independent of the data contained in the table: a Zebra table has row-based darkening of the background of alternative rows (excluding headings). Many documents now are utilizing zebra-striping technique in order to improve the legibility [10]. Also, Wheeler et al. [85] discussed what types of table or graph could improve the reading of temporal data. Zebra striping was considered because it alternates the background color of each row, making it easier for users to distinguish rows when scanning tables. However, according to Tufte's data-ink principle [4], implementing zebra-striping could add unnecessary clutter that impedes reading performance. Lee et al. [46] conducted a small empirical study to find out if a zebra striping format could improve the accuracy and response time. The result shows plain format was the best choice in a task that asked participants to identify whether or not each grid contained the target; it also compares vertical shading and horizontal shading and concludes that vertical shading is better because it supports vertical scanning. However, my research focuses on the horizontal zebra striping table since I are assuming it could help viewer to make comparisons across different rows. Thus, Enders [24] also conducted an experiment to find out whether human's instinct is accurate or not. Unfortunately, from the result that she got, no evidence was found to support our stereotype; thus, zebra striping does not help to improve the accuracy of tasks, and the completion times were improved only when participants were answering difficult questions. After that, Enders [25] started to collect the survey from other viewers and suggested that even zebra striping is not assisting in reading, but it indeed was shown to be aesthetically preferred by viewer.

Font manipulation In a data table, "font" refers to the style and appearance of the data, including typeface, font size, style (like bold or italic), and color. These font attributes have a significant impact on the overall readability and aesthetics of the table [30]. While there's limited research on fonts in data tables, some studies focus on font design. Zhang explores how different fonts affect legibility [92], while Ali et al. suggest that fonts like Verdana and Georgia (for screens) and Times New Roman and Arial (for print) offer optimal readability [3]. Another study points out potential biases in using font size for data encoding in visualizations and recommends strategies, including font weight modification, typeface changes, or the use of monospaced fonts, to mitigate potential biases [2]. Moreover, research on tag clouds highlights the importance of font size and weight in grabbing user attention [5]. Wordle visualizations demonstrate the importance of font characteristics in text display and user engagement, collectively enhancing our understanding of font typeface, size, style, and color in various contexts [27].

This section discusses two Visual features in tables, encompass zebra-striping and font manipulation. Zebra-striping, alternating row colors for readability, is debated for its effectiveness. Font manipulation, involving typeface, size, style, and color, impacts table readability and aesthetics. Studies show mixed results on the readability benefits of these features, with considerations for task specificity and aesthetic preferences.

2.5 Visual Encodings

Nowadays, a growing range of visual encoding techniques such as bars, heatmaps [86], line charts, and pie charts are applied to graph visualization. Additionally, since the 19th century, people have started applying visual encoding to table cells to enhance readability [69]. I classify the following as visual encodings that can appear in tables: Color shading, Bar, Sparkline, Circle, Average Bar Chart, Dual Bar Chart, and Fatfonts [56].

Color Shading The Color shading technique is similar to a heatmap, which involves using a color scale-like data encoding on each cell according to its value [86]. Silva et al. [69] suggest that color scales are effective in visualization, but their effectiveness depends on the choice of color scales. Users should ensure that the color scale enhances readability, as indicated in surveyed literature. There are three main color schemes in data visualization: spectral, sequential, and diverging schemes [13]. The blue-white-red color scale is widely used in most color visualizations [55, 39] because it effectively encodes data through colors [72]. Also, Zeileis et al. [91] discussed that for large datasets, the difference in luminance should be quite obvious, and legends indicating the color scales should also be included.

Color-scale-like visualization can be helpful in certain tasks, such as comparing two or several values [64]. Nacenta et al. [56] conducted an empirical study indicating that the color technique can be beneficial in terms of speed. However, the color technique also has downsides. An experimental study by Han et al. [34] showed that accuracy did not improve with color scales alone. However, significant improvement was observed when color scales were combined with legends or digits.

Bar Besides Color Shading, the Bar is also an effective graphical tool for visualizing quantitative data [6]. Perin et al. designed an interactive table incorporating various visual encodings into table cells, including Bars [59]. The design of the bar is intuitive and understandable [42]. Cleveland et al. [18] conducted an experiment comparing pie charts, dot charts, and bars to determine which format is most effective for decoding by people. Siirtola et al. [68] also carried out empirical research comparing three common visualizations to determine which performed better for tasks related to proportions or proportional data. The results of both experiments show that the bar chart technique is the best because all bars start at the same horizontal level and are arranged parallel horizontally or vertically.

Circle Kitous et al. introduced a dot visual encoding method, varying dot sizes to represent data values [8], a precursor to modern scatter or bubble plots. This technique was recognized for its effectiveness by Cleveland et al. in 1984, who highlighted the scatter plot as a powerful tool for data analysis [18]. Wang et al. conducted an empirical study comparing scatter plots and line graphs, finding scatter plots especially effective for trends in time series data [83]. Perin et al. expanded this idea by incorporating dot visual encoding into table cells as Circles, enhancing table readability [59].

Average Bar Chart Kitous et al. introduced the Average Bar Chart, a visualization for comparing data against an average or central tendency, proving especially useful when the central measure is key [8]. This concept has been adapted into table cells by Perin et al. in their Bertifier. In these table cells, Average Bar Charts emphasize how individual data points deviate from an average value. This method is widely utilized in social sciences and educational research, where it's often necessary to compare individual performance against a group norm [59].

Dual Bar Chart Playfair introduced Dual Bar Charts, a bar chart variation designed for comparative studies, placing two distinct data sets side by side for easy comparison [61]. These charts are particularly effective in scenarios that require comparative analysis, such as before-and-after studies or comparisons between control and experimental groups. Perin et al. incorporated this encoding into table cells in their Bertifier. Dual Bar Charts in this context allow for the comparison of opposing datasets, providing a clear visual representation of comparative analyses [59]. This method is widely used in areas like clinical trials, environmental impact studies, and other fields where comparing two distinct sets of data is essential.

Sparkline The Sparkline was created by Tufte [76] and is a simple, word-sized graphic with typographic resolution [77]. As Brandes et al. explained, "sparklines, made from gestalt theory-informed glyphs, facilitate holistic recognition of patterns, trends, and outliers in multivariate sequences" [12]. Nowadays, Sparklines have become increasingly popular, with software like Excel allowing users to implement their own [41]. They are also being incorporated into a growing number of technical products, such as smartwatches [57] and in the stock market [31]. Compared to other techniques, Sparklines might not be as accurate, but their small size facilitates the arrangement of data text and graphics [89]. However, they are not ideal for visualizing large datasets, as they can confuse audiences. Frishberg et al. [31] introduced an interactive Sparkline that allows users to manipulate the visualization while observing it, thus improving usability and utility. Additionally, inspired by Sparkline design, Neshati et al. enhanced the line graph by condensing it along the x-axis to just one data point per pixel, calling it G-Spark, which makes it easier to quickly glance at the information and navigate through it [57].

FatFonts Although many visualization techniques exist, researchers continue to develop new visual encodings. Nacenta et al. [56] designed a novel numeric typeface for visualizing quantitative data: FatFonts, which falls under small contextual visualizations [34]. Not only did Nacenta et al. test it through an empirical study, but other researchers also conducted experiments to assess its usability. Manteau et al. [50] compared color scales, detail on demand, and FatFonts. The results empirically show that FatFonts has advantages in improving speed and accuracy for reading and value comparison, especially for tasks that require finding the correct coordinate of the extremes. Han et al. [34] conducted an experiment testing four types of visualizations: Plain digital table, heatmap, tooltip heatmap, and FatFonts. The results indicate that while the tooltip heatmap is the optimal solution, FatFonts also improves speed and accuracy compared to the plain table. However, it has drawbacks, as the size of the font might influence how people interpret data. Following this principle, FatFonts could mislead people into thinking the number is larger if the font size is bigger [2].

This section discusses various visual encoding that can be incorporated into table cells. Color shading, similar to heat maps, uses color scales to encode data, with blue-white-red being a popular choice. However, its effectiveness depends on the color scheme and clear legends. Bars are another effective tool, intuitively representing quantitative data. Circles or dots are used in scatter plots, with varying sizes representing data values. Average and Dual Bar Charts are useful for comparing data against an average or between two datasets, respectively. Sparklines, created by Tufte, are compact, typographic graphics for depicting trends, suitable for small datasets but less effective for larger ones. FatFonts, a novel numeric typeface, visually encodes quantitative data within the font itself, improving speed and accuracy for certain tasks but potentially misleading due to font size. These encoding methods enhance data readability and comprehension, each with unique advantages and suitable applications.

2.6 Gaze Tracking Analysis

Gaze tracking is a technique that tracks where a person's eyes go, how long they look at things, the path their eyes follow, and how they get there. Many fields, including psychology, computer science, medicine, and virtual reality, are beginning to use gaze tracking analysis in their studies [53, 40].

In 2013, Chen et al. conducted a comprehensive investigation into gaze movement patterns during reading from various display media, including computer displays, e-readers, and printed books. They found differences in reading behavior and collected valuable data on how long people look at things (Fixations) and the jumps their eyes make between them (Saccades) [90].

In the field of data visualization, some researchers believe that relying solely on quantitative analysis is insufficient to fully understand the impact of visualizing data. To truly understand how people interact with visualizations, they use gaze tracking data and analyze it. This method has become popular for understanding how users engage with visualizations [40].

Mackinlay and Hanrahan focused on critically examining information graphics through gaze tracking. Their study emphasized using eye tracker measurements to evaluate the effectiveness of visualizations, highlighting the importance of analyzing gaze patterns and dwell time. This kind of analysis is crucial for evaluating graphic design and data visualization [32].

When analyzing gaze tracking data, researchers consider various aspects, including Fixation, Saccades, Scan Path, and Region of Interest (or Area of Interest) [40]. Their analysis often begins with the thresholds they establish. Common methods for analyzing gaze tracking data include:

Fixation Analysis This involves identifying and analyzing fixations, which are periods during which the gaze is relatively stable and focused on a specific point on the screen [63]. Metrics related to fixations include the total number of fixations, the average duration of each fixation, and the distribution of fixations across different areas of interest.

Saccades analysis: Saccades are rapid gaze movements between fixations. Saccade analysis can include metrics such as saccade amplitude (the distance covered during a saccade) and saccade velocity [33].

Scan Path Analysis: This method involves tracking the sequence of fixations and saccades to understand the order in which users explore a user interface, website, or visualization [26].

Areas of Interest (AOI) Analysis: Areas of Interest, also known as Regions of Interest, are specific areas on the screen defined by researchers for analysis. The metrics include the total time spent on AOIs, the order of AOI visits, and the number of fixations within AOIs [87].

Heatmaps: Heatmaps are visual representations of eye-tracking data, showing areas on the screen that receive the most visual attention. They are often used to identify hotspots of user attention [22].

Chapter 3

Methodology

I designed a controlled laboratory experiment to address the research questions listed in the Introduction. The study is split into five within-subjects experiments testing five different tasks, with four different visual conditions for table presentation.

In this chapter, I introduce my visual aids selection, task design and research hypothesis and proceed to outline the methodology for experiment design, which includes the eye tracking device, data tables, as well as software and apparatus selection. Additionally, I will detail my approach to recruitment, data collection, and the data analysis process.

3.1 Visual Aids Selection

Since my principal goal of this experiment is to understand whether and how visual aids affect data table reading by humans, I have chosen four different conditions, which represent common table-related tasks based on three criteria: 1) the visual aids should be widely used to make the results relevant to real situations; 2) the visual aids should be static, suitable for printed and on-screen tables; 3) the selected visual aids should be as diverse as possible to cover a broad range of options.

Plain The *plain* table (Fig. 1.1.A) is my control condition. It shows text and data in black over a white background. The only visual feature besides the text and numbers are the horizontal and vertical grid lines that are commonplace in most tables (presumably to help the viewer navigate the table without skipping rows or columns). Because of its ubiquity and for comparability, I include lines separating rows and columns in all the conditions.

Zebra A *zebra* table (Fig. 1.1.B) uses a light gray color on the background of cells in alternative rows of data. This is a common design (e.g., is one of the template formatting options in popular spreadsheet and presentation software). Zebra is a visual feature (not a visual encoding) because the appearance of a Zebra-striped table does not vary when the values change.

Color A table where numerical values are encoded through the background color of each cell (Fig. 1.1.D) is easy to produce through spreadsheet software with a function usually called "conditional formatting". Values are mapped to color with a default MS Excel white to red scale. The mapping is done by column because different columns often represent different types of values, which have different scales (e.g., a column in a table about countries might show population in millions of people and another column might show GDP in trillions of dollars). There are three main color schemes in data visualization: Spectral, sequential, and diverging schemes [13]. Diverging schema is most used in visualization [13], but since my research doesn't include negative value, I used sequential schemes on my data encoding. Color condition is a visual encoding because the visual appearance of the table changes based on the values that it contains.

Bar A table with a *bar* visual encoding (Fig. 1.1.C) displays variable-length horizontal bars in the background of each numerical cell that are proportional in length to the value represented in the cell. The scaling of the bar length according to the value is selected per column, for the same reason as with color. This kind of data encoding is also common in existing software. The length of aligned bars has also repeatedly been shown as one of the most accurately perceived visual variable for differentiation of values [18, 73]. Although sometimes this kind of encoding uses an additional column to represent the bars [38] (instead of using the background of the cell), I avoided this because it consumes additional columns, and would therefore make the conditions not comparable to each other.

3.2 Task Design and Research Hypotheses

Tables are used for many purposes, and people might use them in many ways. To make an appropriate selection of tasks representative of the real world I followed a semi-formal procedure. Concurrently, I examine research hypotheses for each task.

3.2.1 Tasks

I gathered tasks from existing literature that empirically explore tables(e.g., [38, 24, 85]) and other grid-based visualizations (e.g., [50, 34]). I then reviewed and narrowed down these tasks to select ones that are representative of real-world scenarios and encompass a wide variety of task types. This includes both simple and more complex tasks, as well as tasks that are **value-dependent** which require participants to consider and work with the actual numerical values present in the data. For example, a value-dependent task might involve calculating averages, finding specific data points, or making comparisons based on the data values. These tasks directly depend on the value in the table. and **value-independent** which do not require participants to focus on the actual data values within the table. Instead, it might involve tasks like identifying patterns, detecting trends, or understanding the structure of the table without needing to perform calculations or comparisons based on specific data values.

The final set of tasks selected for the study consists of 5 tasks:

TDescr Describing the table (through a speaking aloud protocol, where participants express thoughts aloud while completing a task [14].). When participants saw a table for the first time, they had one minute to freely explore it. The prompt was: "*When you are done reading these instructions, you will see a data table. Please carefully look at the table and explore it. While you explore, describe aloud your thinking about what information this data table tries to convey. After this, describe any remarkable things that you find in this data. You will have 60 seconds in total for this task.*".

Rationale My goal is to understand how people initially perceive a table and what catches their attention. I'm interested in how they explore the data table and how various visual representations influence their gaze. This task was specifically designed to explore how individuals visually engage with tables and learn how to build an effective table.

TVertMax Finding the name of the item (row) that has the maximum value in a given column.

Participants got the name of an attribute (e.g., "*Find the name of the Candy which has the highest 'Pluribus' value. Press space as soon as you have found the answer, and say the name of the Candy aloud once the screen becomes white.*"). They had to find the name of the row that contained the highest value for that attribute within the table (e.g., “Reese pieces”).

Rationale My main goal is to understand how visual encoding aids in value comparison when the visual scales align with the values. In this task, since the color scales 3.1and bar length 3.1 directly correspond to the values, I can determine if visual encoding facilitates value comparison.

THorMax Finding the name of the attribute (column) that has the maximum value in a given row. Participants got the name of an object (e.g., "*For the Candy called ‘Pop Rocks’, find which attribute has the largest numerical value. Press space as soon as you have found the answer, and say the value aloud once the screen becomes white.*"). They had to find the name of the attribute that contained the highest value for that object (e.g., “fat content”).

Rationale As my table has color shade and bar length aligning vertically with column values, not row values, I aim to determine if horizontally comparing values with column-based encodings brings any impediments or not.

TValue Retrieving a value given the names of an item (row) and an attribute (column)(e.g., "*For the Anime called ‘Ultimate horror’, please find its ‘Favorites’ value. Press space as soon as you have found the answer, and say the value aloud once the screen becomes white.*").

Rationale Locating a specific value in a table is a common real-life task with data table [66]. I chose this task to assess how various visual representations produce different outcomes.

TDiff Finding the attribute (column), out of several possible attributes, that has the largest proportional difference between two items (rows) Participants saw the name of two objects (e.g., "*For the films called ‘Water Diviner’ and ‘Black White’, find in which column in the last four columns of the table they are most different (proportionally). Press space as soon as you have found the answer, and say the name of the film aloud once the screen becomes white.*") and had to find in which of the last four attributes (columns) the proportional difference between the values was largest. For example, if the value attribute “calories” was 20 for xxx and 25 for yyy (ratio of 1.25 for yyy), but 40 and 80 for attribute “fat content” respectively (ratio of 2 for yyy), then the name of the attribute with the highest ratio would be the correct answer. The rows were non-contiguous.

Rationale I aim to find out if the complexity of a task leads to varied outcomes when different visual aids are used within data tables.

3.2.2 Hypotheses

I have formulated hypotheses for the research questions outlined in the section 1.1, which were subjected to quantitative analysis. These hypotheses are based on comparisons of time and error across various conditions. However, for the second group of research questions that was subjected to qualitative analysis, I did not hypothesize because of lack of informed expectations or due to the nature of qualitative analysis.

I use the $Measurement_{Technique}$ notation below to simplify reading the comparisons. For example $\tau_{Bar} > \tau_{Plain}$ refers to completion time for the given task is greater with the Bar condition compared to the Plain condition. There are a total of 32 possible hypotheses, based on 4 research questions $\times$ 4 tasks (TVertmax, THormax, TDiff and TValue—I exclude TDescr here because it is qualitative) $\times$ 2 measures (completion time: τ and error: ε).

I refer here to each individual hypothesis as a combination of the research question (e.g., RQ1), the task (e.g., THormax), and the measurement (e.g.,τ). For example, "RQ1 • THorMax • τ" refers to a hypothesis about whether visual data encoding is helpful (RQ1) with the task of finding a maximum within a row (THorMax), in terms of completion time (τ). For conciseness and space, I use list of tasks or research questions (between { and }) in the notation when a hypothesis applies to several tasks or measurements.

H1: RQ1 • {TVertMax, THorMax, TDiff} • {τ,ε} Visual Data Encoding will help people complete TVertmax, THormax, and TDiff faster and with fewer errors than a plain table. Therefore I expect the following individual comparisons to hold: $\tau_{Color} < \tau_{Plain}$, $\tau_{Bar} < \tau_{Plain}$, $\varepsilon_{Color} < \varepsilon_{Plain}$ and $\varepsilon_{Bar} < \varepsilon_{Plain}$. **Rationale:** In tasks that require comparing values, visual encoding allows people to find and access the locations of values of interest (e.g., larger values) in a more efficient and accurate way than a plain table, which requires reading and interpreting digits one-by-one.

H2: RQ2 • {TVertMax, THorMax, TDiff, TValue} • {τ,ε} The visual features of the Zebra condition will help people complete all tasks faster and with fewer errors than a plain table: $\tau_{Zebra} < \tau_{Plain}$ and $\varepsilon_{Zebra} < \varepsilon_{Plain}$. **Rationale:** Zebra striping is designed to help people navigate tables and has been shown in some previous research that it can enhance data table usability [24].

H3a: RQ3 • {TVertMax, THorMax, TDiff} • {τ,ε} Visual encoding of data will be more useful than the Zebra striping for completing TVertmax, THormax, and TDiff faster and with fewer errors: $\tau_{Color} < \tau_{Zebra}$, $\tau_{Bar} < \tau_{Zebra}$, $\varepsilon_{Color} < \varepsilon_{Zebra}$ and $\varepsilon_{Bar} < \varepsilon_{Zebra}$. **Rationale:** I think that the value comparison subtasks of table-reading are harder and more time consuming than the table visual navigation subtasks. Since data encoding techniques are likely to help with the harder tasks, I expect participants to finish them faster with data-encoding techniques.

H3b: RQ3 • TValue • {τ,ε} Zebra striping will be faster and more accurate for TValue than the data-encoding conditions: $\tau_{Zebra} < \tau_{Color}$ and $\varepsilon_{Zebra} < \varepsilon_{Bar}$. **Rationale:** The TValue task does not require decoding and comparison of values, therefore data-encoding tasks will not be useful here, while the visual features offered by Zebra striping should still help navigate the table.

H4: RQ4 • {TVertmax, THormax, TDiff, TValue} • {τ,ε} Color will be faster than Bar and Bar will be more accurate than Color: $\tau_{Color} < \tau_{Bar}$ and $\varepsilon_{Bar} < \varepsilon_{Color}$. **Rationale:** I expect Color to be faster than Bar because Bar creates more visual interference in a cell (sharp transitions below numbers) than Color. In contrast, I expect Bar to result in fewer errors because previous literature presents length as a more accurate visual variable than Color [56].

3.3 Experiment Setups

This section covers the experiment setups, including the use of an eye-tracking device, the choice of data tables, the design of custom software, and the equipment used in the experiment. These setups together enhance the precision and reliability of the experiment.

3.3.1 Eye tracking device

In my experiment design, I used the Tobii Pro Glasses 2 eye tracking device to collect real-time gaze tracking data from participants, enabling me to have an objective understanding of their gaze behavior. The components of the eye tracking device that I utilized in the experiments include:

Head Unit This is the glasses worn by participants. It allows for precise calibration and captures participants' eye movements, including gaze points, saccades, smooth pursuit movements, and also records their voice.

Recording Unit This unit records and processes the eye tracking data collected by the Head Unit. It serves as the central hub for data storage and analysis.

Controller software The software provides control and configuration for the Tobii Pro Glasses 2 system. It enables experimenters to initiate calibration, start and stop recording, and save data.

Tobii Pro software This software is used for qualitative data analysis. It allows experimenters to view the recorded video with all eye movement data collected during the experiment. The software facilitates data analysis by individual participants, tasks, or conditions. It provides the option to slow down the video for more detailed analysis and set breakpoints to separate different trials.

3.3.2 Data Tables Selection

Based on the main condition levels described in Section 3.1, I created eight tables for presentation to participants under different conditions (see Figure 1.1). All tables are based on four data sets, inspired by real data, but modified to avoid unnecessary noise in performance due to variations in familiarity, data ranges, number of attributes and other factors. I selected datasets where: 1) the subject and language are familiar to most people (avoiding technical terms or acronyms); 2) the data itself is not widely recognized (participants wouldn't have prior knowledge or expectations about specific attribute values); and 3) the data is plausible and realistic. The chosen topics include: A) candy brands and their nutritional values; B) ratings, viewership, and other details of animation series; C) nutritional values of cereal brands; and D) movie ratings from various websites.

The final datasets each comprised 30 items (rows), with two-word names (e.g., "Reese Pieces" in the Candy dataset), and 11 attributes (columns) with one to three-word names (e.g., "Sugar Percent"). This resulted in tables containing a total of 333 cells. The item and attribute names were modified to ensure they were not overly similar between items or attributes, preventing them from becoming too distinctive or acting as landmarks.

All the data in the study was numeric, consisting of two-digit values between 10 and 99, with two significant digits, no decimals, and no negative numbers. While textual and categorical data are common in tables, in this study focus on numerical data. This approach ensures consistency in the dimensions and structure of the tables across all conditions.

In the Color and Bar conditions, the mappings were applied on a *per column* basis, This means that cells with the most saturated red or longest bar represented the highest values in each column. This is consistent with the design of most existing tables, where columns often represent different attributes with varying scales, sometimes differing by orders of magnitude. Using a global color or length mapping wouldn't be appropriate in such cases because the differences between values in columns with scale of tens wouldn't be perceivable if another column had values in the scale of millions.

This approach ensured that the visual encoding of numerical values in each cell remained consistent across all conditions. It allowed me to accurately assess the impact of each visual design on participants' completion time and errors in interpreting the data. I also adjusted values based on specific tasks to standardize the difficulty of different task trials. For example, I deliberately modified the numbers to avoid consecutive trials with the same answer, which aided in result accuracy. Additionally, for the TDiff task, I ensured that all numerical values in each trial had similar ratios, and I made the answer relatively easy to detect by ensuring that the answer ratios were consistently greater than 2, while the other ratios were noticeably smaller than 2. These adjustments reduced difficulty and maintained consistency across the trials, enabled us to more effectively compare the effects of different visual designs on participants' performance.

I generated high-resolution bitmap images that cover the entire screen of the monitor used in the experiment pixel-by-pixel without interpolation (see 3.3.4). These tables of all conditions had a thin-line grid around all cells, as well as boldface font for the text in the row and column headers. Example tables in each condition are available in the Appendix.

3.3.3 Software Design

I made custom experimental software in Windows for the experiment using Python's Trinket package [1]. The software has two interfaces: one for the experimenter (Admin Interface) and one for the participant (User Interface).

This software was created to read a CSV file containing task descriptions, corresponding table images, correct answers, task ID, and correct answer's row and column.

During the experiment, participants viewed table images and task descriptions in the

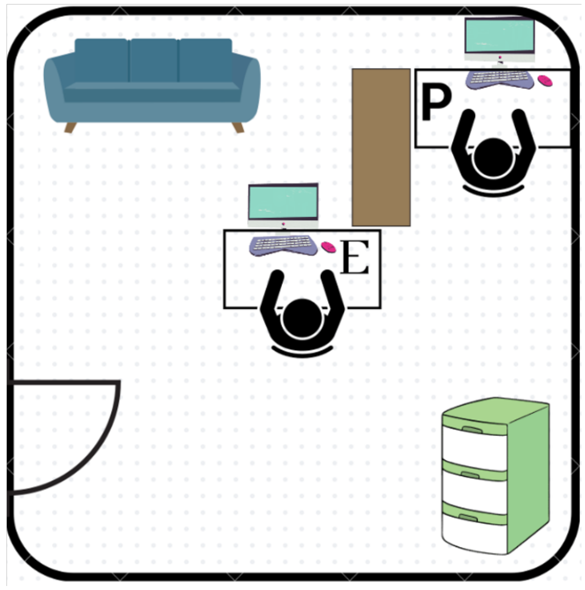


Figure 3.1: Experiment room layout, P indicates Participant, E indicates Experimenter

user interface, while the experimenter used the admin interface. The tables displayed in the admin interface were the same as those seen by the participants, with the only difference being that the experimenter's view of the tables was created using Excel rather than a high-resolution image.

In the admin interface, the experimenter had the ability to select specific cells within the tables. When a participant verbally provided their answer, the experimenter could select the relevant cell and click the "save" button. This action allowed the experimenter to record various data, such as the time spent by the participant on that particular task, the task's start and end times, task ID, the table's name, participant's answer, and the column and row of participant's answer.

The software allowed two mice and two keyboards to be connected, one for the experimenter and one for the participants. Users could use the Space bar to move to the next page and the Tab key to go back to previous page if needed.

Additionally, the software records mouse movements when it detected that participants began and finished a trial.

3.3.4 Apparatus

Participants were sat at a table in a quiet room in front of a 32-inch 4:3 display with a resolution of 3840×2160 and a 60Hz refresh rate. They had access to a mouse and a keyboard.

Additionally, participants wore a head-mounted gaze tracking device(see Section 3.3.1) that was lightweight and unthethered, allowing them to move their head and body freely. The experimenter sat to the right of the participant with an additional mouse and an extra screen positioned at an angle to ensure that the participant could not view its content(see Figure 3.1). This setup enabled the experimenter to input the participant's answers once each trial was completed. The experiment was conducted on a Windows Intel Core i7 PC.

3.4 Participants

Because there are 24 different orders in which the experiment can take place for this study, I need to have at least 24 participants. I recruited 27 participants (ages 19 to 48, average 27, 14 female, 13 male). Participants had to be able to see a computer screen at a regular sitting distance ($\sim$ 80 cm) without glasses or contact lenses (due to gaze-tracker constraints), not have a visual disability, come across data tables in their regular activities (e.g., as part of their study or work), and not have photo-sensitive epilepsy. I discarded the data of three participants. Two because although they indicated having good eyesight, the results showed that they had to sit very close to the screen in order to complete the tasks, which made the eye tracking data inaccurate. The data of the third participant was discarded because they exceeded the time limit by 30 minutes, which is over 50% of the scheduled time (60 minutes)

3.5 Post-Task Questionnaire

At the end of each task, except for the TDescr task, participants were given a post-task questionnaire and asked to provide feedback for each condition in terms of *preference*, *speed* and *accuracy*. This feedback was collected using a ranking scale, requiring participants to rank the conditions from 1 to 4. In this ranking scale, 1 indicated their most preferred condition, while 4 signified their least preferred one. Importantly, participants were not allowed to assign the same rank to multiple conditions. Furthermore, participants had the opportunity to provide additional comments, if any, as part of the process. The questionnaire form can be accessed in the Appendix.

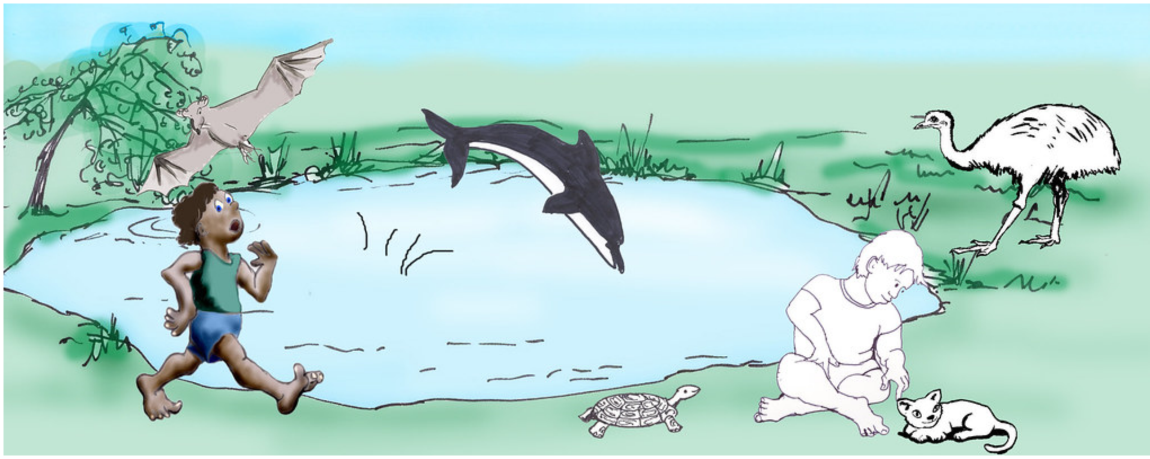


Figure 3.2: Sample picture used for TDescr training

3.6 Procedures

Participants agreed to a consent form and completed a demographic questionnaire, both included in the Appendix. Before starting the experiment, the experimenter assisted participants in putting on the eye-tracking glasses, setting up the calibration, and explained how to use the software for the initial task, known as TDescr (see Section 3.2.1).

In this task, participants encountered a prompt asking them to describe a table on the next page. Once they understood the prompt, they could press the space bar to proceed to the next page, where the table was displayed. They were instructed to describe whatever they observed on the table and then press the space bar to move to the following page. If they did not press the space bar, the page would automatically transition after 60 seconds. After explaining the software instructions for the first task, the experimenter launched a training exercise to familiarize participants with the task and the software usage. The software used in the training procedure worked similarly to the actual task, but it displayed different picture instead of tables (see Figure 3.2). In the training task, participants were asked to answer questions that required them to describe what was happening in the picture.

After the initial task, participants received additional training for the remaining tasks such as TVertMax, THorMax, TValue, and TDiff, with a focus on using the software. Understanding the software was crucial for effectively completing these tasks. The training involved showing them questions like "What's the number in that picture?" or "What's the result in the picture?" on a prompt page. Participants pressed the space bar to advance to the picture page once they felt they had a good understanding of the question. It is important to note that the training pictures were unrelated to tables (see Figure 3.3). This approach

Figure 3.3: Sample picture used for TVertMax, THorMax, TValue, and TDiff training

ensured that participants were familiar with the software, and once their understanding was confirmed, we proceeded with the rest of the experiment.

Before moving on to the next task, the experimenter presented participants with a similar sample task related to the real task but using a smaller sample table (see Figure 3.4). Participants needed to fully understand this task before the experimenter continued with the experiment. Afterward, participants carried out four trials of that task under each condition. The first trial of each task-condition combination was considered training and not included in the analysis. If a participant misunderstood the task (e.g., providing a numeric value instead of the required name of a row), the trial was marked as invalid, repeated, and the repetition also marked as invalid. The only exception to this pattern was the first descriptive task (TDescr), which had only one trial and could not result in an invalid trial. Participants always saw the tasks in the same order (TDescr, TVertmax, THormax, TValue, TDiff), and the conditions within the same task were presented in the same order. The order of the conditions was balanced across participants using all possible orders. For each condition, all participants worked with the same dataset (e.g., the Candy dataset for Plain table trials).

After completing each task, participants were required to fill out a questionnaire (see Section 3.5) in which they ranked the conditions in terms of *preference*, *speed* and *accuracy*. They were also asked to provide comments.

Once all the tasks were finished, participants were free to leave their email addresses so that I could send them a 15 Canadian dollar e-transfer. The research procedure was approved by the local research ethics board.

Group	People Number	Money owned
A	8	1323
B	10	1300

Figure 3.4: Sample table used for training

3.7 Data collection and measurement

I collected the following measurements:

τ **— Trial Completion Time.** This is the time from when the table becomes visible to when the participant presses the space bar because they have an answer ($t_1 - t_0$).

ε **— Error Rate.** The proportion of trials not answered correctly.

Gaze Tracking Data. The gaze tracker recorded a video from the participant's viewpoint, which includes the gaze points.

Mouse Movement. The experimental software recorded the movement coordinates of the mouse over time. I calculated the distance traveled by the mouse to obtain a measure of the total use of the mouse in each trial.

Subjective Performance. After they had completed each task, I asked participants to rank the visual representations in terms of *preference*, *speed* and *accuracy*(see Section 3.5). They assigned a rank from 1 to 4 to each visual representation, with no duplicate rank allowed.

Additional Error Measurements. To explain different sources of error, I captured several additional measurements. For TVertMax and THorMax, I recorded the error magnitude: the difference between the correct maximum value and the value for the column (for TVertMax) or row (for THorMax) indicated as an answer by the participant. For TVertMax/THorMax, I also recorded the error offset, which is the difference between the row/column number of the correct answer and the row/column number of the participant's answer. Finally, for TValue, I recorded both horizontal and vertical offsets.

3.8 Analysis Approach

In order to address the research questions in section 1.1, three researchers, the author, Miguel Nacenta, and Charles Perin analyzed the measurements described in Section 3.7 both quantitatively and qualitatively.

3.8.1 Quantitative Analyses

To address the hypotheses outlined in Section 3.2.2 (completion time and error data) I used a Bayesian statistical approach using Markov Chain Monte Carlo (MCMC) simulations [79, 45]. Completion time, which was log-transformed, was modeled using a Student-t distribution(robust to outliers). The distribution's average was influenced by parameters representing the condition (Plain, Zebra, Color, Bar) and the participant. I avoided modeling interactions between conditions and participants to prevent overfitting. When reporting average times, I back-transformed values from the log-transformed domain. It's important to note that training trials (the first of each four repetitions) and invalid trials(see Section 3.6) were excluded from the analysis. These statistical analysis was built upon Miguel Nacenta's previous study [34].

For errors, I modeled the count of incorrect answers for each participant in each condition-task combination using a binomial distribution. The chance of making an error was determined by a logistic function based on the conditions and the participant. As with completion time, I also excluded training trials (the first of each four repetitions) and invalid trials (see Section 3.6) from the analysis.

The quantitative analysis for RQ5 (i.e., Does the movement of the mouse correlate with the ability to read data tables?) were variations of the models above which incorporate a mouse movement variable. The statistical tests that we run aimed to answer three main sub-questions: 1) Do participants tend to use the mouse differently for various techniques? 2) Does the use of the mouse result in better or worse performance for participants? 3) Does the use of the mouse affect the number of errors made by participants?

I did not pre-register these analyses on OSF because I couldn't predict in advance how mouse movement might meaningfully impact completion times or errors. Mouse use was determined on a binary basis per trial, with trials counting as "mouse trials" if a participant moved the mouse more than 2000 pixels. The 2000-pixel threshold corresponds to approximately half the horizontal size of a 4K screen and effectively distinguishes trials with minimal or no mouse activity from those with mouse activity. It's important to note that the analyses for questions 2 and 3 are correlational, as I did not enforce or discourage mouse use (participants had the freedom to use the mouse). I used these analyses to provide insights into RQ5 and occasionally to offer explanations for overall time and error results.

It's important to note that for the error analysis, I had to modify the MCMC model compared to what was pre-registered. This adjustment was necessary due to my initial failure to anticipate the low level of errors, rendering the previous model inadequate for modeling the collected data. The final approach aligns better with the standard modeling of errors as recommended by Kruschke [44, p. 621]. I used suitably uninformative priors for all models and analyses. Additionally, I ensured that each MCMC simulation demonstrated a good mix of the chains and achieved an Equivalent Sample Size (ESS) of at least 10,000.

Qualitative Analysis

I conducted a qualitative analysis of the gaze tracking data due to the limitations of the high-end head-mounted gaze tracker, which provided measurements that were not precise or frequent enough for automated analysis. My goal was to utilize gaze data to provide a comprehensive description of gaze behaviors (TDescr), and strategies to complete tasks (TVertmax, THormax, TDiff and TValue). As well as to explore potential explanations for performance variations between techniques. This requires analysis at levels of granularity that need to adapt depending on the observed behavior (i.e.,*a posteriori*).

I therefore carried out an iterative systematic visual analysis of the trial video recordings captured by the glasses-mounted camera that were overlaid with gaze data by the gaze-tracker software. For the TDescr task, my focus was on extracting the sequence in which participants looked at different regions of interest (ROI). I conducted a comprehensive analysis of the TDescr videos, with assistance from Miguel Nacenta and Charles Perin. The ROIs I considered included the Table Caption, Row Header, Column Header, and Cells.

For the rest of the tasks, I conducted a more detailed multiple-pass analysis. I performed the analysis on the last two trials of each participant for every condition (Bar, Color, Zebra, plain) across all tasks except TDescr, which had a separate analysis as previously described. By excluding the first two trials, I eliminated some transient behaviors observed as participants familiarized themselves with each visual aid or developed their strategies.

The total number of videos analyzed is thus $24(participants) \times 4(tasks) \times 4(conditions) \times 2(repetitions) = 768$.

In the initial two iterations, I watched the videos with overlaid gaze positions, identifying the observable types of movements (e.g., Smooth Scans, Jumps) and their characteristics of interest (e.g., direction, presence of regressions). During each iteration, we all came together to validate the accuracy of these observations and to ensure that the categories were suitable for my purposes. Subsequently, in the next two iterations, I introduced subcategories and decided to create a structured language to describe movement types and sequences, instead of individually coding each combination due to the large number of combinations of movements and characteristics. In the final iteration, I ensured that all movements were adequately describable and collectively validated a subset of the trials. Although there is some degree of subjective judgment in the categorization of movements, errors in categorization have minimal impact on the analysis. Table 3.1 summarizes the final dimensions of the coding language.

Table 3.1: Gaze Analysis Coding Book

Dimension	Category	Description
MovementType	Jump	Saccade directly to a destination.
	Jump Scan	Reach destination through several saccades skipping cells.
	Smooth Scan	Cell-by-cell traversal of (part of) a row or column.
	Super Smooth Scan	Like smooth scan but with substantial fixation on cells.
	Gaze to Mouse	Saccade to location of the mouse cursor.
	Mouse to Gaze	Moves cursor to currently fixated cell.
Mouse Use	With Mouse	Modifier when mouse and gaze move synchronously.
ROI	Column Header	Horizontal movement in column header (attribute names).
	Row Header	Vertical movement in Row Header (object names).
	Row	Horizontal movement in non-header row.
	Column	Vertical movement in non-header column.
Regression	With Regression	Modifies gaze movement with occassional back and forths.
Direction	Left to Right	
	Right to Left	
	Top to Bottom	
	Bottom to Top	

Chapter 4

Results

In this section, I present results for the following tasks: TVertMax, THorMax, TDiff, and TValue. I conducted a quantitative analysis of completion time and errors, while also examining mouse movement and eye tracker data to observe participants' gaze behavior. Additionally, I considered participant subjective feedback for each task and discussed the task results.

The primary focus for TDescr was on analyzing the eye tracker data, specifically the gaze path. I reviewed the gaze tracking videos to observe the order in which participants viewed the key table elements.

Table 4.1 summarizes how the data supports the hypotheses across different tasks and conditions, corresponding to each hypothesis in Section 3.2.2.

Table 4.1: Summary of hypotheses support. Cells with saturated green background and white text indicate strongly supported hypotheses (prob > 95%), cells with non-saturated green background indicate weaker (prob > 90%) or partial support (e.g., only one of the comparisons hold). The cell with pink background indicates evidence against the hypothesis.

		TvertMax		THorMax		Tdiff		Tvalue	
		τ	ε	τ	ε	τ	ε	τ	ε
H1	Encoding helps	**Color < Plain Bar < Plain**	**Color < Plain Bar < Plain**	Color < Plain	No Support	No Support	No Support	No Support	No Support
H2	Zebra helps	No Support	No Support	No Support	No Support	**Zebra < Plain**	No Support	**Zebra < Plain**	No Support
H3a	Encoding better than Zebra	**Color < Zebra Bar < Zebra**	**Color < Zebra Bar < Zebra**	Color < Zebra	No Support	Zebra < Color Zebra < Bar	No Support	N/A	N/A
H3b	Zebra better than Encoding	N/A	N/A	N/A	N/A	N/A	N/A	**Zebra < Color Zebra < Bar**	No Support
H4	Color faster than Bar (τ) Bar more accurate than Color (ε)	**Color < Bar**	No Support	**Color < Bar**	No Support	Color < Bar	No Support	No Support	No Support

4.1 Subjective data from questionnaire

I used a ranking system to collect participant feedback after each task. Participants ranked their preferences on a scale from 1 to 4 for accuracy, speed, and condition preference. Figure 4.1 displays the average rankings, with darker cells indicating higher ranks (1 being the highest) and lighter cells representing lower ranks (4 being the lowest). These rankings will be discussed in more detail in the task-specific sections.

		TVertMax	THorMax	TDiff	TValue
Plain	Preference	3.8	3.7	3.7	2.8
	Accuracy	3.8	3.7	3.7	3.0
	Speed	3.8	3.7	3.7	2.7
Zebra	Preference	2.9	2.4	1.7	1.3
	Accuracy	2.8	2.5	1.8	1.3
	Speed	3.0	2.5	1.5	1.4
Color	Preference	2.0	1.6	2.4	2.9
	Accuracy	2.1	1.6	2.4	2.9
	Speed	1.9	1.7	2.4	3.0
Bar	Preference	1.3	2.3	2.3	2.9
	Accuracy	1.3	2.3	2.1	2.9
	Speed	1.3	2.1	2.4	2.9

Figure 4.1: Average subjective ranking (between 1 and 4) in terms of preference, accuracy and speed for the four conditions, per task. Darker cells (and smaller value) mean higher ranks and lighter cells (and larger values) mean lower ranks.

4.2 TDescr Results

During the description task (TDescr), participants had the opportunity to familiarize themselves with the tables in each condition. I analyzed the sequence in which participants looked at the main components of the table during this familiarization period. The Sankey diagram in Figure 4.2 shows, for each condition, the number of participants who looked at any of: the table caption (which appeared in the top left of the table, see Figure 1.1), the row headers (leftmost column), the column headers (topmost row), and the cells of the table (the grid containing the data). After inspecting the videos, the results show that, across all conditions, the most common sequence is Caption-Row Headers-Column Headers-Cells, and the second most common sequence inverts the order of Row Headers and Column Headers. However, there are differences between conditions; approximately twice as many people looked first at the cells in the Color and Bar conditions as in the Plain and Zebra

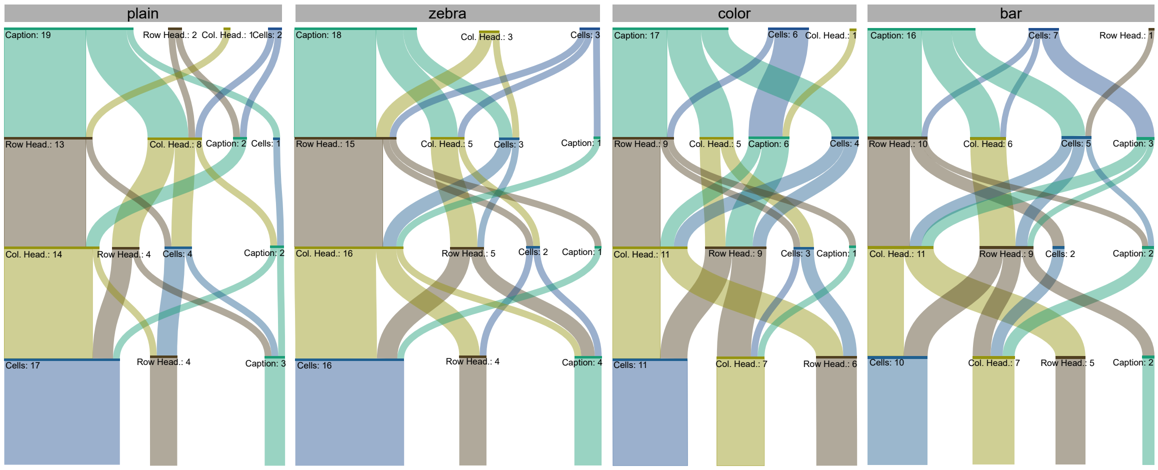


Figure 4.2: Gaze sequences at the table objects during the description task (TDesc), by condition. The width of the bars indicates number of participants who looked at a particular object (e.g., the caption), and bars on top are the beginning of the sequence.

conditions, resulting in a more heterogeneous collection of sequences for Color and Bar (i.e., the most common sequence was less frequent: from 12 of 24 participants in Plain and Zebra striping to 7 and 6 for Color and Bar respectively).

I speculate that the difference is caused by the visual encodings in the Bar and Color conditions, which created a visual stimulus that was more tempting, attractive, unusual, or difficult to ignore than in the other conditions.

4.3 TVertMax Results

TVertMax is about finding the name of the row that has the maximum value in a given column.

4.3.1 Completion Time and Error

Figure 4.3 shows that participants took less time to complete the task with the Color condition, 89% faster than Plain and 90% faster than Zebra, and definitely faster than Bar (16% faster, a second on average). This data supports H1, (visual encodings help) H3a (visual encodings are better than Zebra) and H4 (participants are faster with Color than Bar), but not H2, since participants did not have an advantage with Zebra over Plain.

Overall, error levels were low (a total of 18 errors or 6.2%, see Figure 4.4). The evidence indicates that participants were more likely to make errors with Plain and Zebra

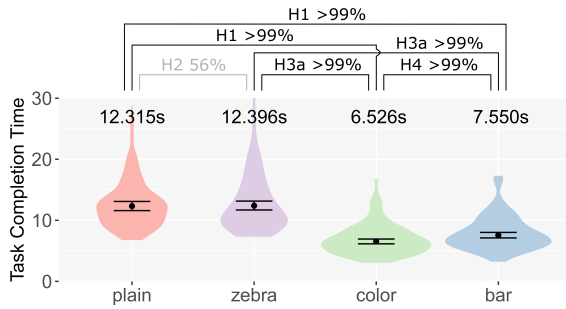


Figure 4.3: Task completion time for task TVertMax. Violin plots indicate density of measurements. Error bars are 95% High-Density Intervals of the log-untransformed mean estimation (numbers in seconds). Pairwise comparisons (lines at the top) are black if the estimated probability of a condition (number) is $>95\%$ or $<5\%$ and grey otherwise.

conditions compared to Color and Bar, although the evidence is slightly weaker for Bar compared to Color. Therefore H1 and H3a are supported. H2 was not supported since the tests do not find conclusive evidence that Zebra results in fewer errors than any other visual feature.

TVertMax	Plain	Zebra	Color	Bar
Plain		53.0	4.6	7.5
Zebra	47.0		3.4	5.5
Color	95.4	96.6		69.5
Bar	92.5	94.5	30.5	

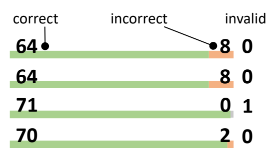


Figure 4.4: Errors for task TVertMax. Pairwise comparisons cells (test of row condition being more likely to produce errors than column condition, in percentages) have black backgrounds if the estimated probability of a condition (number) is $>95\%$ or $<5\%$ (the conjugate), grey if $>90\%$ or $<10\%$ and white otherwise. The horizontal stacked Bar on the right half indicates the number of correct, incorrect, and invalid trials per condition (row).

4.3.2 Effects of Mouse Use

After I examined the mouse movement data for TVertMax, I found that participants were more likely to use the mouse with Plain and Zebra (40.3% and 41.7% of the time, respectively), than with Color and Bar (32.4% and 34.7%). Trials where the mouse was used in Zebra took almost certainly longer, by an estimated 3.5 seconds on average; for the other techniques the differences are not conclusive. There is also no evidence that using the mouse affected the number of committed errors with in any of the conditions.

4.3.3 Gaze Behavior

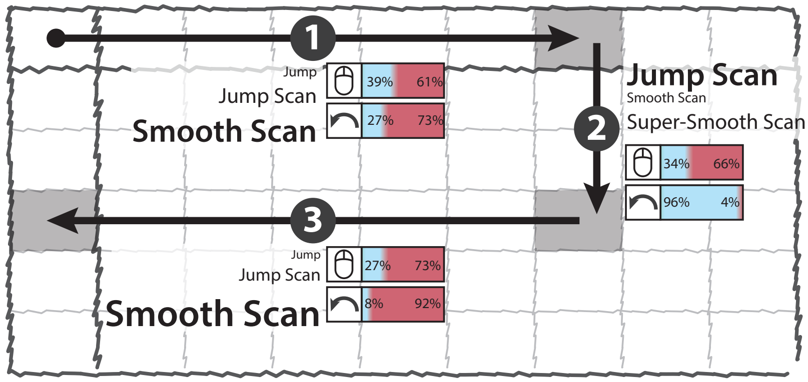


Figure 4.5: Canonical strategy for TvertMax. For each step, the labels indicate with gaze movement occurred, and the font size and font weight communicate the number of occurrences of each type of gaze movement. The mouse and back-arrow charts indicate the percentage of trials where the mouse was used or not, and the number of trials where visual regression was used or not, respectively.

Analysis of gaze behavior showed that most participants followed a 3-step strategy to complete TVertMax. This strategy is illustrated in Figure 4.5 and consists of the following steps:

Step 1: Find the column header indicated in the prompt.

Step 2: Find the cell within the corresponding column that has the maximum value.

Step 3: Find the row header that corresponds to that cell.

I found that 60% of participants started with Step 1, while 40% started directly with Step 2; 99% of participants then completed Step 2, and 99% also finished Step 3. All three steps were performed using Jumps, Jump Scans, Smooth Scans, and Super-Smooth Scans, sometimes involving the mouse. I also found the following differences between conditions:

- Participants used two to three times as many Jump Scans with encodings (28 for Color and 29 for Bar) compared to without encodings (10 for Plain and 10 for Zebra).
- Participants almost never used Super-Smooth Scans with encodings (0 for Color and 2 for Bar) but used them often without encodings (25 for Plain and 25 for Zebra).
- Participants used Jumps (Jump, Jump Scan, and Mouse Jump combined) fewer times with Plain (19) than with Zebra (29), as well as than with Color and Bar (44 with both).

Step-by-step analysis reveals further gaze patterns, which include eye movements and fixations on objects or areas of interest in their field of view [65], based on the subtask involved in each step, and these patterns are the same for all conditions.

Step 1 Analysis. To complete Step 1, participants had to locate the column heading provided in the prompt. There is one notable high-level finding. There is also one notable difference between conditions. There were 7 instances of Mouse Jump (No Regression) with encodings, compared to none with non-encoding conditions.

Step 2 Analysis. To complete Step 2, participants had to read candidate values in the cells of the column corresponding to the found header, and determine which was largest. There are two notable high-level findings. First, participants had to slow down compared to Step 1 because there were 0 Jumps used in this step. Second, participants often had to use regression. They used Jump Scans extensively (in 75/183 trials without mouse, and in 28/183 trials with mouse), but only with regression. They also sometimes used Smooth Scans (in 19/183 trials without mouse, and in 6/183 with mouse) and Super-Smooth Scans (in 22/183 trials without mouse, and in 25/183 with mouse), again always with regression. There are also two notable difference between conditions. First, I found that encodings allow participants to use Jump Scan. Indeed, participants used Jump Scan on the column from top to bottom 29 times with Color and 31 times with Bar, much more than the 5 times with Plain and 10 times with Zebra. Similarly, they used Mouse Jump (Jump Scan) 13 times with Color and 12 times with Bar, but only 2 times with Plain and 1 time with Zebra. Second, I found that conditions without encoding force people to (Super) Smooth Scan. Indeed, participants used Smooth Scan on the column from top to bottom more with Plain (8) and Zebra (8) than with Color (2) and Bar (1). Similarly, they used Super-Smooth Scan more with Plain (11) and Zebra (10) than with Color (0) and Bar (1), and Mouse Scan (Smooth or Super-Smooth) more with Plain (16) and Zebra (14) than with Color (0) and Bar (2).

Step 3 Analysis. There are two notable high-level finding. First, the dominant gaze behavior is clearly Smooth Scan (with and without mouse), without regression. They used Smooth Scans in 89/183 trials but only 5 times with regression. They used Mouse Scans in 42/183 trials but only 4 times with regression. The second preferred gaze behaviors are Jumps (in 21/183 trials) and Jump Scans (in 22/183 trials). The second notable high-level finding is that there were 0 instances of Super-Smooth Scan for this step. There are no notable difference between conditions for this step.

4.3.4 Subjective Results

The subjective ratings for TVertMax (see Figure 4.1) indicate that participants strongly prefer to choose Bar condition in terms of preference, accuracy, and speed, followed by Color, Zebra, and Plain. These findings contrast with the objective time measurements, where Color was clearly better than the Bar.

4.3.5 TVertMax Discussion

The results for the TVertMax task generally align with my expectation that visual encodings help (both in time and error—H1). Some of this is explained by participants choosing to use the mouse more often in the Plain and Zebra conditions. Moving the mouse to serve as a bookmark or as a moving indicator to help keep gaze in straight lines might have some benefits, but these seem to be outweighed by the cost of having to control the mouse itself, leading to slower performance in these conditions. It's worth noting that the mouse analysis is correlational, and the use of the mouse can be both the cause and the consequence.

Nevertheless, differences between conditions can be better explained by gaze behavior: the value encodings allowed participants to jump to potential maximum values in a column, allowing them to jump to these cells without needing to focus on the specific numbers. In contrast, in Plain and Zebra conditions, participants had to scan each vertical column one by one.

Participants did not do better with Zebra than with Plain (H2 not supported). Surprisingly, Zebra, which I expected to facilitate faster horizontal movements with less risk of "changing lanes," did not provide an advantage. I also did not observe significant differences in behavior when gaze had to move to the left to find the row header.

Furthermore, I found a clear difference between Color and Bar in completion time. I speculate that Color is simply more noticeable than the more subtle bars, although I have no strong evidence to differentiate them in terms of mouse use or gaze behavior.

Lastly, an interesting observation is that many trials began directly at the appropriate column. This doesn't mean that participants instinctively knew which column to go to; it's probably the consequence of them already being familiar with the column location that has been mentioned in the prompt from a previous trial (the gaze analysis only covered the last two trials in each condition).

4.4 THormax Results

THorMax requires finding the name of the column that has the maximum value in a given row. It is, therefore, a transposed version of TVertMax.

4.4.1 Completion Time and Error

Figure 4.6 shows that participants took less time to complete the task with Color than with any of the other conditions (about 800 ms faster than Zebra, the second). For this task, participants did not seem to benefit from the Bar encoding. Therefore, H1 is only partially supported (not all encodings produced benefit, only Color), H2 is not supported (Zebra does not help), H3a is only partially supported (only the Color encoding is better than Zebra) and H4 is supported because participants took less time with the Color condition than Bar condition.

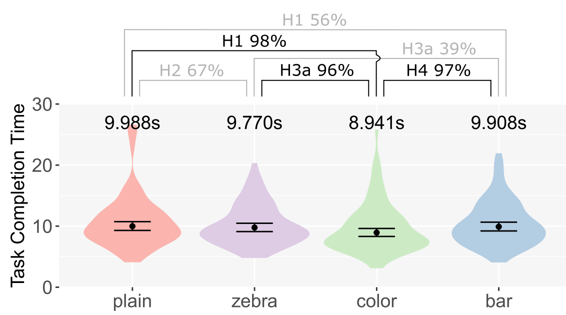


Figure 4.6: Task completion time for THorMax. Refer also to the caption of Fig. 4.3.

Overall, there were also few errors in this task. Participants made more errors with the Color encoding than other conditions (89% probability of Color being more likely to result in errors than with Bar, and 86% with respect to Plain and Zebra). The reduced accuracy with Color is likely due to the task, which I explain in more detail in Section 4.4.5.

THorMax	Plain	Zebra	Color	Bar
Plain		50.8	86.5	43.8
Zebra	49.2		86.4	42.9
Color	13.5	13.6		10.9
Bar	56.2	57.1	89.1	

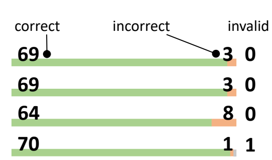


Figure 4.7: Errors for task THorMax. Refer also to the caption of Fig. 4.4

4.4.2 Effects of Mouse Use

After I examined the mouse movement data for THorMax, I found that participants were more likely to use the mouse with Plain than with any other aids (52.7% vs $< 44\%$ in all others, all with probability $> 96\%$). There are no other significant differences in mouse use between conditions. There is also clear evidence that using the mouse is associated with slower completion times. When participants used the mouse, they did 2.9 s worse on average for Color, 2.5 s worse for Plain, 1.2 s for Zebra and 1.0 s for Bar. There is also no evidence that using the mouse helped or impeded accuracy.

4.4.3 Gaze Behavior

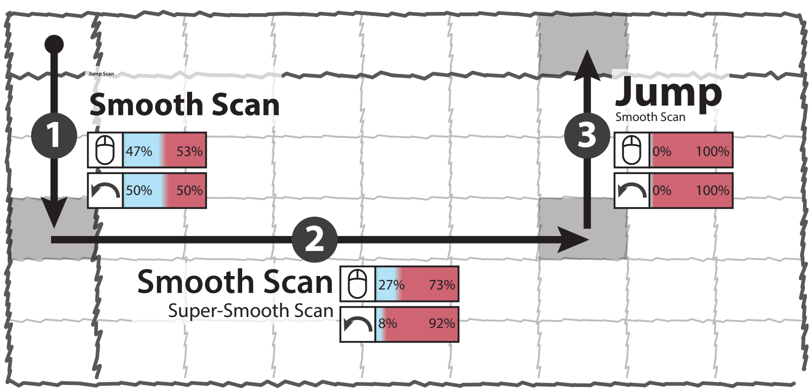


Figure 4.8: Canonical strategy for ThorMax. For each step, the labels indicate with gaze movement occurred, and the font size and font weight communicate the number of occurrences of each type of gaze movement. The mouse and back-arrow charts indicate the percentage of trials where the mouse was used or not, and the number of trials where visual regression was used or not, respectively.

Gaze behavior analysis revealed that a large majority of participants used the following 3-step canonical strategy to complete THorMax. Basically, participants used the same strategy than for TvertMax, except they started from the row header, then found the cell, then found the column header. This strategy is illustrated in Figure 4.8 and consists of the following steps:

Step 1: Find the row header indicated in the prompt.

Step 2: Find the cell within the corresponding row that has the maximum value.

Step 3: Find the column header that corresponds to that cell.

I found that 98% of participants started with Step 1; 98% continued with Step 2; and 94% continued with Step 3. All three steps were performed using Jumps, Jump Scans, Smooth Scans and Super-Smooth Scans, and sometimes involved using the mouse. Unlike TvertMax, I did not find overall differences between conditions.

Step-by-step analysis, however, reveals further gaze patterns based on the subtask involved in each step, and these patterns are the same for all conditions.

Step 1 Analysis. To complete Step 1, participants had to locate the row header that corresponded to the one given in the prompt. The predominant gaze movement was Smooth Scan (in 172/184 trials), with or without using the mouse. The data shows that when participants did not use the mouse, more would do so with regression (54/184 trials were Smooth Scans with regression against 35/184 without regression); but when they did use the mouse, more would do so without regression (53/184 trials without regression against 30/184 with regression). There were no instances of Jumps (0/184), very few Jump Scans (6/184), and 0 Super-Smooth Scans.

Step 2 Analysis. To complete Step 2, participants had to read candidate values in the cells of the row corresponding to the found header, and determine which was largest. Like in Step 1, the largely predominant gaze movement was Smooth Scan (in 140/184 trials), but this time there were no differences in terms of regression used based on whether they used the mouse or not. The second preferred gaze movement was Super-Smooth Scan (in 47/184 trials), with or without using the mouse equally. All instances of Super-Smooth Scans but two were accompanied with regression. In this step, there were 0 instances of Jumps or Jump Scans.

Step 3 Analysis. The largely predominant strategy was to use Jump, without regression, and without the mouse (in 145/184 trials). The second preferred strategy was to use Smooth Scan, again without regression and without the mouse (in 18/184 trials).

4.4.4 Subjective Results

The subjective ratings for THorMax (see Figure 4.1) indicate that participants strongly prefer the Color for this task and also thought it was faster and more accurate. Participants rated Bar slightly above Zebra also in all three criteria, and clearly above Plain. These findings align with the results of time measurements, where Color was clearly better than other conditions.

4.4.5 THorMax Discussion

I included THorMax as a task in this study mostly as a counterpoint to TVertMax. However, unlike TVertMax, THorMax is not a natural task for most tables because columns usually represent attributes that are often unrelated to each other in terms of measurement units and orders of magnitude (e.g., one column could count Potassium milligrams—0.01 to 0.2, and the next the sugar content— 3 to 50 g). This means that, unless columns happen to be in same measurement units and range (e.g., the same measurement for a series of years as columns), it is usually meaningless to find the attribute with the largest value for a row. As a side consequence, value encodings make more sense *in columns*, this implies that comparing horizontal encodings can result in errors when a value is the maximum in one column but appears lower when compared to values in other columns. This was the case in my stimuli and it explains the higher number of errors in the Color condition. All errors belonged to the group of trials where the encoding "contradicts" the values (7 out of 16 trial types).

Despite these challenges, most participants completed the task correctly, even in the Color condition. And they tended to perform the task more quickly when using Color encoding. This indicates that errors in the Color condition are primarily related to the table's configuration. In contrast, there were no similar errors observed with Bar encoding, suggesting that bars that are not aligned vertically are recognized as useless. Interestingly, the use of a mouse did not appear to reduce completion time or reduce error rates.

4.5 TDiff Results

TDiff requires finding the column, out of several possible columns, that has the largest proportional difference between two specified rows.

4.5.1 Completion Time and Error

As expected, this task takes longer to complete due to its complexity, with averages between 30 and 34 seconds depending on the condition. Participants completed the TDiff task fastest with Zebra (30.46 s average), which is nearly two seconds faster on average than with Color, three seconds faster than with Plain and four seconds with Bar, which was slowest (see Figure 4.9). The data offers no conclusive evidence that the encodings help (H1), it clearly supports the value of Zebra (H2), and it contradicts H3a, since participants were clearly faster with Zebra than with Bar and with Color (this last, with less certainty). There is also moderate evidence that Color speeds up this task with respect to Bar (H4).

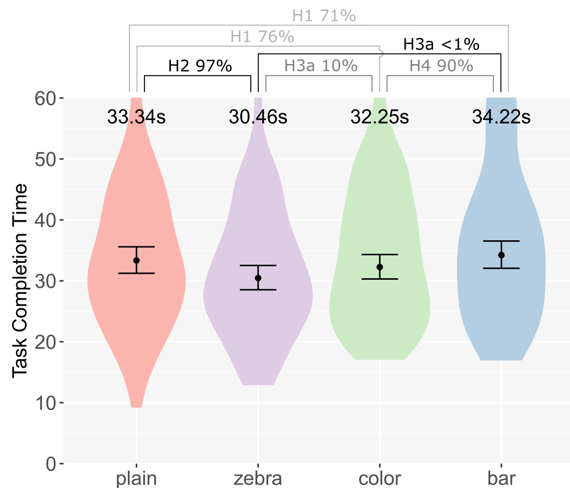


Figure 4.9: Task completion time for TDiff. Refer also to Fig. 4.3.

Participants made more errors in this task (17% of non-training trials reported a wrong answer), but the errors are quite homogeneously distributed among the conditions, without evidence indicating an advantage for any of the conditions.

TDiff	Plain	Zebra	Color	Bar
Plain		35.8	26.8	31.9
Zebra	64.2		38.1	44.9
Color	73.2	61.9		55.7
Bar	68.1	55.1	44.3	

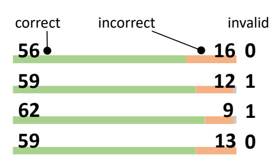


Figure 4.10: Errors for task TDiff. Refer also to the caption of Fig. 4.4

4.5.2 Effects of Mouse Use

For this task most participants (91%) used the mouse, similarly across all conditions. There is also no conclusive evidence that using the mouse shortened or lengthened completion times. Similarly, mouse use does not seem to affect the likelihood of errors.

4.5.3 Gaze Behavior

For the TDiff condition, we initially conducted 184 trials. However, only 160 of these trials were accessible through gaze tracking videos. The videos for 24 trials, which involved 3 different participants, were unavailable for review. This unavailability was due to two main reasons: either the eye-tracking device failed to record these trials, or the trials exceeded the device's recording time limit, resulting in the video stopping recording at some point during the TDiff trials. As a consequence, these were discarded due to incomplete data. Among the 160 observable trials, 15 presented significant challenges for analysis. These trials involved rapid and complex eye movements that were difficult to categorize accurately. Considering the high likelihood that including these data would introduce more noise than useful information into our analysis, we decided to exclude these 15 trials. As a result, 145 trials from the TDiff condition were deemed suitable and were included in our detailed analysis.

Analysis of gaze behavior showed that participants did not have a step strategy from start to finish that was common across participants to complete TDiff. However, I have outlined Analysis of gaze behavior showed that participants did not have a dominant step strategy to complete TDiff. However, I have outlined several steps that were frequent among many participants:

Step 1: Find two row headers indicated in the prompt.

Step 2: Find the horizontal location of the values to compare by identifying the corresponding columns based on the column header.

Step 3: Diverse steps for Row Comparison

I found that 99% of participants began with Step 1. Subsequently, 78% completed Step 2. However, 42% of participants jumped back to the row header after completing Step 2. From Step 3 onwards, gaze behavior became chaotic, as participants followed several different paths, Step 3 involved a diverse range of actions, with participants carrying out multiple different actions in various orders. There were no clear discernible patterns in order to compare different values in different columns across the two rows. Therefore, Step 3 can be considered as comprising a multiplicity of steps. Minimal differences were noticeable between conditions. Additionally, more Mouse Scans were used in the TDiff condition. Notably, 87% of participants used the mouse simultaneously while moving their gaze in Step 1, and 63.5% did the same in Step 2.

Step-by-step analysis, however, reveals further gaze patterns based on the subtask involved in each step, and these patterns are the same for all conditions.

Step 1 Analysis. To complete Step 1, participants had to locate two row headers specified in the prompt. The predominant gaze movement observed was the Smooth Scan (in

143/145 trails), executed with or without using the mouse. The data indicates that the majority of participants used the mouse while performing a Smooth Scan (in 124/145 trails), while only 12% did not use the mouse (in 19/145 trails). Furthermore, when participants did not use the mouse, a higher proportion resorted to regression (91%) compared to when they used the mouse (65%). There were no instances of Jumps, Jump Scans, or Super-Smooth Scans.

Step 2 Analysis. To complete Step 2, participants had to follow the row that corresponds to one of the two row headers, moving from left to right. As in Step 1, the predominant gaze movement was the Smooth Scan (in 113/145 trails). Notably, 63.5% of participants used the mouse while performing a Smooth Scan, with or without regression (37/145 trials without regression against 54/145 with regression). Participants who used the mouse tended to hover the cursor over the other row they were scanning, then moved their gaze and the mouse simultaneously. In this step, there were no instances of Jumps or Jump Scans. After completing Step 2, 42% of participants performed a back-and-forth Jump to the row header, without using the mouse and without regression (in 61/145 trails).

Step 3 Analysis. Step 3 in our study was notably chaotic, encompassing a variety of movements and steps rather than a single, uniform strategy. Within this context, a specific gaze movement named 'Compare' was identified as a key element. This 'meta movement' was coded separately from those discussed in Table 3.1 due to its distinct nature. 'Compare' was observed in 95% of participants (137 out of 145 trials). It involved examining values across two rows and was not a singular, one-time step but a multiplicity of steps within each trial. The exact frequency and timing of this step varied widely, making it difficult to count the exact number of times it occurred or to determine when it happened, which was not very helpful for my analysis. Additionally, a common pattern observed in completing Step 3, other than comparing values belonging to the two rows, involved participants performing jumps between row and column headers, and making some back-and-forth jumps between the row headers and the values. This was likely done to ensure they were comparing the correct rows.

4.5.4 Subjective Results

The subjective ratings for TDiff (see Figure 4.1) indicate that participants strongly prefer to choose Zebra in terms of preference, accuracy, and speed, followed by Bar, Color, and Plain. These findings align with the results for time measurements, where Zebra was clearly better than other conditions.

4.5.5 TDiff Discussion

I included the TDiff task in my study because it's a complex task. Participants need to scan two rows to the rightmost side and compare values in the last four columns. In a prior study [24], a task involving scanning a single row within Zebra striping was conducted, and the results indicated that Zebra striping reduced completion time. This aligns with my findings, suggesting that Zebra striping helps in reducing completion time. This is likely because participants don't get lost easily when comparing two rows. Surprisingly, while I initially thought that value encodings might assist in comparison tasks, my results didn't show any clear evidence that they improve completion time or reduce errors. This might be because this task involves more scanning than direct comparison.

Notably, 91% of participants chose to use a mouse for the task in all conditions. Even though there's no strong evidence that using a mouse improves completion time or reduces errors, it seems that, when the task becomes sufficiently complex, participants instinctively rely on external aids like the mouse to complete it.

An observation from eye tracker videos is that participants tended to find the first target row, place the mouse cursor as a marker on that row, and then locate the other target row. Once they identified both rows, they would use the mouse to scan one row and their eyes to scan the other simultaneously, facilitating accurate column comparison. Additionally, participants were likely to jump back to the row header, which I speculate was to prevent straying from the rows.

Overall, due to the task's complexity, participants made more mistakes compared to other tasks. However, in most cases, this complexity led them to use the mouse as a marker. When a task requires extensive scanning along rows or columns and comparison, participants tend to prefer visual aids that keep their gaze from straying from on the row or column, rather than the visual aids that assist with the comparison.

4.6 TValue Results

TValue requires retrieving a value contained in the cell determined by the names of its row and column.

4.6.1 Completion Time and Error

Trials completed with Zebra were, on average, almost a second faster than with Color or Bar, and about 1.5 second faster than with the Plain tables (see Figure 4.11). H1 is not supported, since the differences between Plain and Bar and Color are small and not statistically reliable. H2 is clearly supported by the evidence (people are faster with Zebra than with Plain). H3b, which is different for this task because we did not expect the encoding to help, is supported, although the evidence for Zebra being faster than Bar is much stronger than with Color.

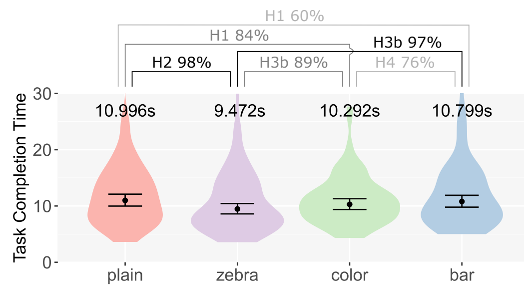


Figure 4.11: Task completion time for TValue. Refer also to Fig. 4.3.

In terms of errors, TValue saw only six errors overall, and the differences between conditions are too small to extract any conclusion about correctness (see Figure 4.12).

Tvalue	Plain	Zebra	Color	Bar
Plain		47.4	49.7	35.9
Zebra	52.6		52.5	38.5
Color	50.3	47.5		36.0
Bar	64.1	61.5	64.0	

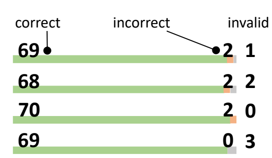


Figure 4.12: Errors for task TDiff. Refer also to the caption of Fig. 4.4

4.6.2 Effects of Mouse Use

Participants used the mouse between 70% and 76% of the time, depending on the condition. There is no evidence that they were more likely to use the mouse in one condition than another. There is some statistically reliable evidence (98% probability) that, when in the Bar condition, mouse use resulted in faster trial completion, by an average of about 1.7 seconds. The evidence is inconclusive for all other conditions.

4.6.3 Gaze Behavior

The analysis of gaze behavior revealed that, like in TDiff, participants did not follow a specific step strategy to complete TValue. However, I have outlined several common steps:

Step 1: Find either the row header or column header as indicated in the prompt

Step 2: Find the remaining header, either row or column, that was not identified in Step 1

Step 3a: Find the cell within the corresponding column that aligns with the specified row header

Step 3b: Find the cell within the corresponding row that aligns with the specified column header

I found that 99% of participants began with Step 1, with 79% of them finding the row header first and 21% finding the column header first. Subsequently, 99% of participants completed Step 2. From Step 3 onwards, gaze behavior was separated into two different scenarios, Step 3a and Step 3b, based on their choice in Step 1. Consequently, 79% of participants completed Step 3a, while 21% completed Step 3b. No differences were found between conditions for individual actions or when aggregated (Regression/No Regression, Jumps, Smooth Scans).

Step 1 and Step 2 Analysis. Some participants chose to start Step 1 by locating the row header first, followed by Step 2 where they found the column header, while others opted to find the column header first and then the row header. Consequently, I have analyzed both steps together. To complete Step 1 or Step 2, participants needed to locate either the row header or column header as specified in the prompt. The predominant gaze movement observed in both steps was the Smooth Scan (166 of the 184 trials). Participants were more likely to use the mouse while scanning the row header compared to the column header (with mouse usage noted in 40 out of 184 trials for row scanning against only 4 out of 184 trials for column scanning).

Due to the different sequences in which participants completed Step 1 and Step 2, Step 3 had to be divided into two distinct parts: Step 3a for the scenario where they located the row header first and then the column header, and Step 3b for the scenario where they located the column header first and then the row header.

Step 3a Analysis. To complete Step 3a, participants navigated down the corresponding column header and then moved left to check the row header, repeating this process until Step 3a was completed. This step involved various gaze movements, including Jumps and Smooth Scans, both with and without mouse use. The frequency and direction of gaze movements varied in each trial. In Step 3a, 146 out of 184 trials included Jumps within the row, with 93 being from right to left and 53 from left to right. There were 129 out of 184 trials that included Smooth Scans within the row, with 40 from right to left and 89 from left to right. Additionally, Mouse Scans within the row were observed in 164 out of 184 trials, with 11 from right to left and 153 from left to right.

Step 3b Analysis. To complete Step 3b, participants navigated down the corresponding row header and then moved from bottom to top to check the column header, repeating this process until Step 3b was completed. This step exclusively involved Jump movements without mouse usage. The frequency and direction of gaze movements varied in each trial. A total of 180 Jump movements within the column from top to bottom were observed, along with 293 Jump movements from bottom to top. There were no instances of Mouse Scan and Smooth Scan in this step.

4.6.4 Subjective Results

Participants preferred Zebra to all other conditions and judged it fastest and most accurate. All other conditions had similar scores. These findings align with the results for time measurements, where Zebra was clearly better than other conditions.

4.6.5 TValue Discussion

I included the TValue task in the study as a data retrieval task, which is a common real-life task. The results showed that using Zebra striping reduced completion time by one second compared to the other visual aids. This aligns with my initial hypothesis that Zebra striping can be beneficial for tasks that do not require value comparison. Similar to TDiff, TValue primarily involves extensive scanning along rows or columns, and the row-based darkening of the background in alternate rows helps participants prevent their gaze from straying from the rows they are scanning.

Analyzing mouse movement data, I found that 70% to 76% of participants used the mouse as a tool for TValue, which is less than TDiff but higher than comparison tasks like TVerMax and THorMax. Combining this with eye tracker video, it seems that for tasks that involve substantial scanning along rows and columns, participants prefer to use the mouse as a marker. They first locate either the target row name or column attribute and hover the mouse cursor there. Then, they use their eyes to find the other target, moving the cursor and eyes simultaneously until they find the result.

Overall, because the task doesn't involve value comparison, participants prefer the Zebra striping condition that assists with scanning. Visual encoding may not introduce significant visual obstacles, as Plain, Bar, and Color shading had similar results in terms of errors and completion time. However, for tasks that involve extensive scanning along rows and columns, participants prefer the Zebra striping condition which helps them prevent straying from rows and columns.

Chapter 5

Discussion

In this chapter, I interpret the findings from the experiments and evaluations. I begin by discussing the value of Zebra Striping, followed by a reflection on its benefits compared to previous research. Further, I delve into the value of visual encodings, such as Color and Bar, including a comparative discussion of these two encodings. Additionally, I compare visual encodings with Zebra Striping. Then, I discuss the value of mouse use and the novelty of my study. Finally, I address the limitations of my research and suggest areas for future work to expand our understanding of how other visual aids may impact the reading of data tables.

5.1 The Value of Zebra Striping

My study examines the effectiveness of Zebra striping in data tables. A key finding, particularly for tasks that require extensive scanning along rows and columns or are complex, such as TValue and TDiff, is the significant impact of Zebra striping in reducing completion time. This suggests that Zebra striping may help users scan rows and columns in tables more efficiently.

While Zebra striping demonstrated a clear advantage in terms of time efficiency, its effect on error reduction was less supported. The study did not find conclusive evidence to assert that Zebra striping significantly reduces error rates. This aligns with previous research by Enders [24] and Lee [46], who observed similar trends where Zebra striping reduced completion time for complex tasks but did not notably affect error rates.

Furthermore, the subjective results and gaze tracking analysis in this study revealed a strong user preference for Zebra striping in complex tasks or tasks that involve substantial scanning along rows and columns, such as TDiff and TValue. Participants seemed to suggest that the alternating row colors in Zebra striping might act as a helpful visual guide, potentially aiding them in preventing straying from specific elements they were scanning. This feature was particularly beneficial in tasks that required extensive scanning along rows and columns, as it helped users keep track of their position in the table.

The implications of these findings for table design are significant. They suggest that incorporating Zebra striping can be a straightforward yet effective strategy for improving completion time in tasks that are complex or involve extensive scanning along rows and columns. This insight can be leveraged to enhance the usability of data tables. Additionally, the study opens up possibilities for further research into other visual aids that could complement Zebra striping in various scenarios.

5.2 The Value of Visual Encodings

The results from my empirical study show that incorporating visual encodings in table cells, like Color and Bar, can indeed help to facilitate value comparison within tables, and also suggest that Color is more effective than Bar in terms of completion time, depending on different scenarios.

For TVertMax, a task that involves vertical comparison, the findings revealed that visual encodings significantly helped participants quickly identify maximum values within columns. This was evident as participants could jump to potential maximum values using these encodings, rather than scanning each column value individually, as in the Plain and Zebra conditions, observed from their gaze behavior.

In the THorMax task, where columns in tables usually show unrelated attributes, identifying the maximum value across them can be more challenging. This was evident in the higher error rates, especially in the Color condition, where the visual encodings occasionally contradicted the actual values. Despite these challenges, participants generally did well in this task, even when using the Color encoding. This implies that the errors were likely due to the layout of the table rather than the encoding itself.

The TDiff task was complex, involving extensive scanning along rows and columns and comparing values across multiple rows and columns. In this task, the study did not find clear evidence that visual encodings like Color and Bar significantly improved completion times or reduced errors. This might be due to the task's nature, which required extensive scanning along rows and columns rather than direct comparison.

Similarly, in the TValue task, which solely involved scanning along rows and columns, the use of Color and Bar did not markedly enhance performance in reducing time and errors. This again points to the possibility that the tasks involving scanning along rows and columns rather than comparison might not align well with the strengths of these particular visual encodings.

When comparing Color and Bar directly, for instance, in TVertMax, participants completed the task faster with Color than with Bar. I speculate that this difference was caused by Color being more immediately noticeable than Bar, and bars are difficult to compare when placed horizontally. However, in THorMax, this advantage was mitigated due to the potential for misinterpretation of values, especially when encodings contradicted the actual data. This finding suggests that while both encodings have their strengths, their effectiveness can vary based on the task context, the nature of the data being presented, and how the visual encoding is incorporated.

The findings from this study both align with and extend the existing literature on data table design and visualization techniques. While researchers like Mahon [48], Wainer [82], MacDonald-Ross [47], Spence [71], and Midway [54] have highlighted the limitations of traditional tables in facilitating comparisons, this study confirms that visual encodings, specifically Color and Bar, can indeed enhance table readability in tasks involving direct value comparison, such as TVertMax and THorMax. This aligns with what Henry [36] and Hink [37] find in their empirical study, where well-designed graph visualizations, which are a kind of visual encoding, are more effective for comparison tasks than relying solely on tables. However, when a task requires extensive scanning along rows or columns and comparison, as in TDiff, participants tend to prefer visual aids like Zebra striping, which help keep their gaze from straying from the rows or columns, rather than visual encodings that aid in comparison.

The implications of these findings are significant for the design of data tables. They highlight the potential of combining traditional tables and visual encodings to optimize the usability and effectiveness of data tables, particularly in scenarios requiring quick comparison.

5.3 Visual Encoding vs. Zebra Striping

Both visual encoding and Zebra Striping have shown benefits in reducing response time. However, when comparing these two approaches, it's essential to discuss their performance within the specific context of tasks. The findings show that for tasks involving comparison, incorporating visual encodings like Color and Bar in table cells can significantly facilitate value comparison. Meanwhile, for tasks that substantially involve row or column scanning, a visual feature like alternating row background colors, known as Zebra Striping, might be helpful in concentrating on scanning, thus potentially reducing completion time.

In TVerMax, using visual encoding has been shown to significantly reduce completion times when dealing with tasks that involve a lot of comparisons. This improvement in completion time is because participants can quickly identify potential maximum values in a column without having to focus on specific numbers, as seen in gaze behavior. On the other hand, Plain requires participants to scan columns one by one, leading to more mouse usage, which can act as a guide but also slows down the task. In summary, in tasks involving comparisons, visual encoding outperforms visual features because it reduces the need for scanning and frequent mouse usage.

In the TValue and TDiff tasks, a visual feature like Zebra striping is more effective than Color and Bar visual encodings in reducing completion time. This might be because these tasks require extensive scanning along rows or columns. While TDiff primarily involves comparisons, the scanning aspect is crucial due to its complexity. The results show that the visual feature aids in both tasks by reducing completion time. Participants' gaze behavior indicates that they minimize back-and-forth checking to prevent straying from rows or columns in TDiff. Most participants used the mouse as a marker or tool for scanning, and this usage was consistent across different visual features and encodings. The use of visual features like Zebra striping reduces back-and-forth checking and helps prevent straying from the rows or columns that participants were scanning, leading to faster task completion compared to using other visual encodings.

Limited research exists on the comparison between Zebra Striping and visual encodings in tables. However, an empirical study by Han and Nacenta compared tables of digits with and without Color encoding, including tasks like finding extremes [34]. Their findings indicate that incorporating Color in table cell backgrounds significantly reduces completion time compared to using tables with digits only. My result aligns with their findings, showing that visual encodings like Color notably aid in reducing completion time for tasks requiring value comparison.

The research findings on the use of visual encoding and Zebra Striping in tables have important implications for data visualization. For tasks requiring comparison, like in TVerMax, visual encodings (e.g., Color and Bar) enhance efficiency by enabling quick identification of key values, thereby reducing the need for detailed comparison. However, in tasks involving extensive scanning along rows and columns, such as TValue and TDiff, visual feature like Zebra Striping are more effective for scanning. This suggests a need for task-specific design in data tables, where the choice of visual aids should align with the user's task requirements.

5.4 The Value of Mouse Use

My study also investigated mouse usage while participants performed various tasks, revealing key findings. Participants were more inclined to use the mouse in tasks with Plain and Zebra striping conditions during the TVertMax task. This tendency correlates with my finding that visual encodings, such as Bar and Color in table cells, substantially reduce TVertMax completion time, suggesting a potential reduction in visual complexity. Therefore, there appears to be a preference for mouse use in environments that are more visually complex or less distinct, as seen in the Plain and Zebra conditions. Additionally, it was found that using the mouse tends to slow down task completion time, though its effect on task accuracy was not definitively determined. Notably, in more complex tasks such as TDiff, a substantial majority of participants (91%) chose to use the mouse. This indicates that the frequency of mouse usage may correlate with the complexity of the task.

Moreover, The specific mouse behaviors observed in the TDiff and TValue tasks offer insightful details into how participants interact with tables using the mouse.

In the TValue task, participants frequently used the mouse cursor in conjunction with their gaze to scan along rows and columns. This behavior suggests that the mouse cursor often aligns with the participants' gaze, serving as a dynamic tool for navigation through the table. Additionally, in TValue, participants were observed using the mouse cursor to hover over row headers or column headers as markers. This technique appears to aid in maintaining a reference point, thereby reducing the cognitive effort needed to locate additional specific data points within the table.

In the TDiff task, a notable behavior was the use of the mouse cursor to track one row while simultaneously using gaze to track another. This dual-tracking method implies that participants might use the mouse cursor as an auxiliary gaze point, allowing them to keep track of different information while their gaze is directed elsewhere. This approach could indicate a sophisticated use of the mouse as a complementary tool to the gaze, enhancing the ability to process and compare information in complex tasks.

Limited research exists on investigating mouse usage in data tables. However, researchers have made great efforts to investigate mouse usage in other fields. Chen and Sohn conducted an empirical study, suggesting a strong alignment between gaze position and mouse cursor position [17]. This aligns with my findings that participants frequently used the mouse cursor along with gaze to navigate through columns and rows. Extending Chen and Sohn's research, Cooke conducted an exploratory experiment to further study the relationship between gaze position and mouse cursor position [19]. He suggested that the mouse could be used as a "poor man's eye tracker", meaning it could potentially serve as an alternative to an eye tracker. However, my findings introduce more perspective to Cooke's suggestion, indicating his suggestion can be task-dependent. For tasks that involve multi-row comparisons, like in TDiff, participants often use the mouse as an auxiliary gaze point, rather than completely aligning it with the actual gaze point. Thus, in this scenario, the mouse can not serve as an alternative to an eye tracker. Furthermore, Cox and Silva's empirical study on the interaction between mouse movements and gaze movement [20], while not explicitly stating that more complex tasks lead to more mouse use, does indicate that the mouse plays a crucial role in managing and navigating through complex information. This suggests that in more complex tasks, effective mouse usage becomes increasingly important, which aligns with my findings that mouse usage might increase when tasks become more complex.

The findings from my study highlight the significant role of mouse usage in navigating data tables. Participants demonstrated a preference for using the mouse in environments perceived as more visually complex, such as those with Plain and Zebra conditions. This suggests that reducing visual complexity might decrease mouse usage, as mouse use was also associated with slower task completion times. Additionally, the observed alignment of gaze and mouse cursor positions among participants suggests the potential for using the mouse cursor as an alternative to eye-tracking in certain contexts, particularly in understanding how audiences explore data tables.

5.5 Novelty

While the concept of enhancing data table readability and performance by incorporating non-interactive visual encodings into the background of table cells is not new, my study introduces a novel aspect through its systematic investigation of this approach's effects. Diverging from previous studies, it combines graphical elements with numeric data within the same cell and offers a detailed and structured analysis of how this approach impacts table readability. This systematic approach provides new insights into the effectiveness of various visual encodings in data tables.

Compared to the existing study Bertifier by Perin et al. [59], which incorporates visual encoding in table cells without presenting actual digits, this study offers a different approach. According to Han and Nacenta's findings, the absence of digits can make locating specific values more difficult than when visual encoding is presented alongside digits [34]. My study addresses these challenges identified in Bertifier. Additionally, my approach differs from Hink's Grable [37], which combines graphs with tables by placing numbers within graphs, rather than using graphs as visual encodings in the background of table cells. This deviation from the typical table structure sets my approach apart from Hink's.

5.6 Limitations and Future Work

Even though I've carefully planned my study design with confirmation from my supervisory committee, it's important to acknowledge that there are limitations in my study, as well as opportunities for future work.

- The monitor screen has a limited size, which restricts the visibility of the table when employing a large table. The current table consists of only 30 rows and 11 attributes. In future work, there is a plan to expand the table's size to accommodate larger datasets. This expansion will help validate whether effective solutions can be found for handling large data tables and explore which visual aids can assist with larger datasets.

- The actual data in my tables only includes two digits (10-99). This was done to ensure consistency and avoid introducing any noise into the results. However, in real-life scenarios, data can vary greatly in range, so I may need to reconsider the table's configuration and incorporate larger numbers into the table.

- Even though I designed the study carefully and filtered out tasks that my supervisory committee deemed inappropriate for the experiment, There still be additional tasks that could be included. In future work, I plan to design tasks that are either commonly used in real-life scenarios or specifically tailored for result-driven experiments.

- In my study, I focused on only two visual encodings: Bars and Colors. However, to gain a more comprehensive understanding of how visual encodings affect table readability, it may be insufficient to examine only these two conditions. In future research, I plan to incorporate additional visual encodings into the table cells, such as sparklines [89], and FatFonts [56], to provide a more comprehensive analysis of the impact of various visual encodings on table readability.

- The device I used in the experiment for gaze tracking was the Tobii Glass 2, which has been discontinued. While the gaze behavior tracking was relatively accurate while watching the video, the precision was insufficient to support automated analysis. Furthermore, participants had to wear the eye tracker throughout the entire experiment, which could cause discomfort and potentially affect the results. Additionally, due to the requirement of wearing glasses, there were strict participant recruitment criteria, such as good vision without the use of external tools.

- In the future, I plan to use a higher-level, more advanced gaze tracker that doesn't require wearing glasses. This change should enhance result reliability and make a wider range of participants eligible, thus expanding the target group for the study.

- Another factor that may need improvement is controlling for the participants' learning background. Some participants may be good at calculations and have experience with tables, while others may not have such experience. This variance can introduce noise into the results. In the future, I intend to create specific criteria when recruiting participants to ensure the fairness of the experiment and obtain more reliable results.

Chapter 6

Conclusions

Tables are often used for data presentation, and while some research focuses on enhancing their readability through interactivity like adjusting rows and columns [62] and adding tooltips [16], limited attention has been given to improving table readability using visual aids. Perin et al. introduced Bertifier [59], which incorporates visual elements into table cells but doesn't display the actual digits, potentially making it more error-prone for users to retrieve values from the table [34]. Hink's Grable combines graphs with tables but places numbers in graphs instead of using graphs as visual encodings in the background of table cells [37], which deviates from the typical table structure.

To address these concerns, I incorporated visual encodings such as bars and colors into the background of table cells. This approach facilitates value comparison and trend identification, while still allowing users to view the digits easily. Additionally, I included a visual feature, Zebra Striping, in my study to evaluate how both visual features and visual encodings contribute to improving table readability from various perspectives.

I conducted a study with 24 participants and presented the results. The study highlights the advantages of using visual elements like color and bars in table cells for tasks that require comparisons. Additionally, it highlights the benefits of employing visual features, such as Zebra Striping, for tasks primarily involving scanning along rows and columns. The results also reveal how people interact with tables through their gaze behavior and mouse usage, strengthening the study's conclusions and offering insights for creating more effective tables. Given the growing use of data tables in real-life, it becomes increasingly crucial to help people improve their understanding of tables and enhance table readability.

Bibliography

[1] Trinket, “help documentation,” 2023. https://trinket.io/help. Accessed: 2023-11-10.

[2] Eric Alexander, Chih-Ching Chang, Mariana Shimabukuro, Steven Franconeri, Christopher Collins, and Michael Gleicher. Perceptual biases in font size as a data encoding. *IEEE transactions on visualization and computer graphics*, 24(8):2397–2410, 2017.

[3] Ahmad Zamzuri Mohamad Ali, Rahani Wahid, Khairulanuar Samsudin, and Muhammad Zaffwan Idris. Reading on the computer screen: Does font type have effects on web text readability?. *International Education Studies*, 6(3):26–35, 2013.

[4] c Arbogast. Tufte, e. r. envisioning information. cheshire, ct: Graphics press, 1990. *Perceptual and motor skills*, 71:351–351, 1990.

[5] Scott Bateman, Carl Gutwin, and Miguel Nacenta. Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 193–202, 2008.

[6] James R. Beniger and Dorothy L. Robyn. Quantitative graphics in statistics: A brief history. *The American statistician*, 32(1):1–, 1978.

[7] Chris Bennett, Jody Ryall, Leo Spalteholz, and Amy Gooch. The aesthetics of graph visualization. In *CAe*, pages 57–64, 2007.

[8] William H Benson and Bernard Kitous. Interactive analysis and display of tabular data. In *Proceedings of the 4th annual conference on Computer graphics and interactive techniques*, pages 48–53, 1977.

[9] Jacques Bertin. *Graphics and Graphic Information Processing*. De Gruyter, Berlin, Boston, 1981.

[10] Alison Black and Karen L Stanbridge. Documents as 'critical incidents' in organization to consumer communication. *Visible language*, 46(3):246–249, 2012.

[11] Thomas E Bradstreet. Grables: Visual displays that combine the best attributes of graphs and tables. In *A picture is worth a thousand tables: graphics in life sciences*, pages 41–69. Springer, 2012.

[12] Ulrik Brandes, Bobo Nick, Brigitte Rockstroh, and Astrid Steffen. Gestaltlines. *Computer graphics forum*, 32(3pt2):171–180, 2013.

[13] Cynthia A. Brewer, Alan M. MacEachren, Linda W. Pickle, and Douglas Herrmann. Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers*, 87(3):411–438, 1997.

[14] Elizabeth Charters. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, 12(2), 2003.

[15] Chun-houh Chen, Wolfgang Härdle, Antony Unwin, and Michael Friendly. A brief history of data visualization. *Handbook of data visualization*, pages 15–56, 2008.

[16] Lianjun Chen and Hongbo Zhou. Research and application of dynamic and interactive data visualization based on d3. pages 150–155, 07 2016.

[17] Mon Chu Chen, John R Anderson, and Myeong Ho Sohn. What can a mouse cursor tell us more? correlation of eye/mouse movements on web browsing. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 281–282, 2001.

[18] William S. Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[19] Lynne Cooke. Is the mouse a" poor man's eye tracker"? In *Annual Conference-Society for Technical Communication*, volume 53, page 252, 2006.

[20] Anna L Cox and Mariana M Silva. The role of mouse movements in interactive search. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 28, 2006.

[21] Kenneth Cukier and Viktor Mayer-Schoenberger. The rise of big data: How it's changing the way we think about the world. *Foreign Aff.*, 92:28, 2013.

[22] Gautier Drusch, JC Bastien, and Stéfane Paris. Analysing eye-tracking data: From scanpaths and heatmaps to the dynamic visualisation of areas of interest. *Advances in science, technology, higher education and society in the conceptual age: STHESCA*, 20(205):25, 2014.

[23] Charles Durbin. Effective use of tables and figures in abstracts, presentations, and papers. *Respiratory care*, 49:1233–7, 11 2004.

[24] Jessica Enders. Zebra striping: does it really help? In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, pages 319–322, 2007.

[25] Jessica Enders. Zebra striping: more data for the case. *A LIST APART*, 2008.

[26] Sukru Eraslan, Yeliz Yesilada, and Simon Harper. Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research*, 9(1), 2016.

[27] J. Feinberg. Wordle. http://www.wordle.net. Accessed: 2023-11-10.

[28] Stephen Few. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, Oakland, CA, USA, 1st edition, 2009.

[29] Stephen Few and Perceptual Edge. Data visualization: past, present, and future. *IBM Cognos Innovation Center*, pages 1–12, 2007.

[30] Lauren E Franzblau and Kevin C. Chung. Graphs, tables, and figures in scientific publications: The good, the bad, and how not to be the latter. *The Journal of hand surgery (American ed.)*, 37(3):591–596, 2012.

[31] Leo Frishberg. Interactive sparklines: a dynamic display of quantitative information. In *CHI '11 Extended Abstracts on human factors in computing systems*, CHI EA '11, pages 589–604. ACM, 2011.

[32] Joseph H Goldberg and Jonathan I Helfman. Comparing information graphics: a critical look at eye tracking. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization*, pages 71–78, 2010.

[33] Ziad M Hafed. Mechanisms for generating and compensating for the smallest possible saccades. *European Journal of Neuroscience*, 33(11):2101–2113, 2011.

[34] Han L. Han and Miguel A. Nacenta. The effect of visual and interactive representations on human performance and preference with scalar data fields. In *Proceedings of Graphics Interface 2020*, GI 2020, pages 225 – 235. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2020.

[35] James Hartley. Tabling information. *The American psychologist*, 46(6):655–656, 1991.

[36] Gary T Henry. Using graphical displays for evaluation data. *Evaluation Review*, 17(1):60–78, 1993.

[37] Jessica K Hink, Jason K Eustace, and Michael S Wogalter. Do grables enable the extraction of quantitative information better than pure graphs or tables? *International journal of industrial ergonomics*, 22(6):439–447, 1998.

[38] Jessica K Hink, Michael S Wogalter, and Jason K Eustace. Display of quantitative information: are grables better than plain graphs or tables? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 40, pages 1155–1159. SAGE Publications Sage CA: Los Angeles, CA, 1996.

[39] T Höllt, J Beyer, F Gschwantner, P Muigg, H Doleisch, G Heinemann, and M Hadwiger. Interactive seismic interpretation with piecewise global energy minimization. In *2011 IEEE Pacific Visualization Symposium*, pages 59–66. IEEE, 2011.

[40] Petr Benda Pavel Šimek Jakub Štěpán Novák, Jan Masner and Vojtěch Merunka. Eye tracking, usability, and user experience: A systematic review. *International Journal of Human–Computer Interaction*, 0(0):1–17, 2023.

[41] Bill Jelen. Using sparklines to visualize your data. *Strategic finance (Montvale, N.J.)*, 92(12):62–, 2011.

[42] Daniel A. Keim, Ming C. Hao, Umesh Dayal, and Meichun Hsu. Pixel bar charts: A visualization technique for very large multi-attribute data sets. *Information visualization*, 1(1):20–34, 2002.

[43] C.N. Knaflic. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley, 2015.

[44] John Kruschke. Doing bayesian data analysis: A tutorial with r, jags, and stan. page 621, 2014.

[45] John K Kruschke. Bayesian analysis reporting guidelines. *Nature Human Behaviour*, 5(10):1282–1291, 2021.

[46] Michael Lee, Travis Kent, C. Melody Carswell, Will Seidelman, and Michelle Sublette. Zebra-striping: Visual flow in grid-based graphic design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1):1318–1322, 2014.

[47] Michael MacDonald-Ross. How numbers are shown: A review of research on the presentation of quantitative data in texts. *AV communication review*, 25(4):359–409, 1977.

[48] BH Mahon. Statistics and decisions: the importance of communication and the power of graphical presentation. *Journal of the Royal Statistical Society: Series A (General)*, 140(3):298–307, 1977.

[49] Melissa Mallon. Data visualization. *Public Services Quarterly*, 11(3):183–192, 2015.

[50] Constant Manteau, Miguel Nacenta, and Michael Mauderer. Reading small scalar data fields: Color scales vs. detail on demand vs. fatfonts. In *Proceedings of Graphics Interface 2017*, GI 2017, pages 50 – 56. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2017.

[51] Francis T Marchese. Exploring the origins of tables for information visualization. In *2011 15th International Conference on Information Visualisation*, pages 395–402. IEEE, 2011.

[52] Francis T Marchese. Tables and early information visualization. In *Knowledge visualization currents: From text to art to culture*, pages 35–61. Springer, 2012.

[53] Ann McNamara and Ranjana Mehta. Additional insights: Using eye tracking and brain sensing in virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2020.

[54] Stephen R Midway. Principles of effective data visualization. *Patterns*, 1(9), 2020.

[55] Kenneth Moreland. Diverging color maps for scientific visualization. In *Advances in Visual Computing*, Lecture Notes in Computer Science, pages 92–103, Berlin, Heidelberg. Springer Berlin Heidelberg.

[56] Miguel Nacenta, Uta Hinrichs, and Sheelagh Carpendale. Fatfonts: Combining the symbolic and visual aspects of numbers. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, page 407–414, New York, NY, USA, 2012. Association for Computing Machinery.

[57] Ali Neshati, Yumiko Sakamoto, Launa C Leboe-McGowan, Jason Leboe-McGowan, Marcos Serrano, and Pourang Irani. G-sparks: Glanceable sparklines on smartwatches. In *Proceedings of the 45th Graphics Interface Conference on Proceedings of Graphics Interface 2019*, GI'19, Waterloo, CAN, 2019. Canadian Human-Computer Communications Society.

[58] F. Nightingale. *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army: Founded Chiefly on the Experience of the Late War*. Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army, Founded Chiefly on the Experience of the Late War. Harrison and Sons, 1858.

[59] Charles Perin, Pierre Dragicevic, and Jean-Daniel Fekete. Revisiting bertin matrices: New interactions for crafting tabular visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2082–2091, 2014.

[60] Rune Pettersson. *Information Design An introduction*. Document design companion series, v. 3. John Benjamins Publishing Company, Philadelphia, PA, USA, 2002.

[61] W. Playfair. *The Statistical Breviary*. Wallis, 1801.

[62] G N V Ramana Rao and Stuart Card. The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. *Proc. CHI*, 06 2002.

[63] Keith Rayner. Eye movements in reading and information processing. *Psychological bulletin*, 85(3):618, 1978.

[64] Marc L Resnick and Carlos Fares. Visualizations to facilitate online tabular presentation of product data. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(13):1498–1502, 2004.

[65] Shane L Rogers, Craig P Speelman, Oliver Guidetti, and Melissa Longmuir. Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific reports*, 8(1):4271, 2018.

[66] Milo Schield. Statistical literacy: Reading tables of rates and percentages. In *ASA Proceedings of Statistical Education Section*. Citeseer, 2001.

[67] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16:1139–48, 01 2011.

[68] Harri Siirtola. Bars, pies, doughnuts & tables – visualization of proportions. 09 2014.

[69] Samuel Silva, Beatriz Sousa Santos, and Joaquim Madeira. Using color in visualization: A survey. *Computers & graphics*, 35(2):320–333, 2011.

[70] Jaime Snyder, Elizabeth Murnane, Caitie Lustig, and Stephen Voida. Visually encoding the lived experience of bipolar disorder. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[71] Ian Spence and Stephan Lewandowsky. Displaying proportions and percentages. *Applied Cognitive Psychology*, 5(1):61–77, 1991.

[72] Reto Stauffer, Georg J. Mayr, Markus Dabernig, and Achim Zeileis. Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations. *Bulletin of the American Meteorological Society*, 96(2):203–216, 2015.

[73] Justin Talbot, Vidya Setlur, and Anushka Anand. Four experiments on the perception of bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2152–2160, 2014.

[74] A.C. Telea. *Data Visualization: Principles and Practice, Second Edition*. Taylor & Francis, 2014.

[75] Torbjørn Torsvik, Børge Lillebo, and Gustav Mikkelsen. Presentation of clinical laboratory results: an experimental comparison of four visualization techniques. *Journal of the American medical Informatics association*, 20(2):325–331, 2013.

[76] Edward R Tufte. The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3):15, 1985.

[77] Edward R. Tufte. *Beautiful evidence*. Graphics Press, Cheshire, CT, 2006 - 2006.

[78] Antony Unwin. Why is data visualization important? what is important in data visualization? *Harvard Data Science Review*, 2(1):1, 2020.

[79] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved r for assessing convergence of mcmc (with discussion). *Bayesian analysis*, 16(2):667–718, 2021.

[80] Howard Wainer. Understanding graphs and tables. *Educational researcher*, 21(1):14–23, 1992.

[81] Howard Wainer. Improving tabular displays, with naep tables as examples and inspirations. *Journal of Educational and Behavioral Statistics*, 22(1):1–30, 1997.

[82] Howard Wainer and D Thissen. Plotting in the modern world: Statistics packages and good graphics. *Chance*, 1(1):10–20, 1988.

[83] Yunhai Wang, Fubo Han, Lifeng Zhu, Oliver Deussen, and Baoquan Chen. Line graph or scatter plot? automatic selection of methods for visualizing trends in time series. *IEEE transactions on visualization and computer graphics*, 24(2):1141–1154, 2017.

[84] Martin Weinmann. Visual features—from early concepts to modern computer vision. In *Advanced topics in computer vision*, pages 1–34. Springer, 2013.

[85] Andrew P Wheeler. Tables and graphs for monitoring temporal crime trends: Translating theory into practical crime analysis advice. *International journal of police science & management*, 18(3):159–172, 2016.

[86] Leland Wilkinson and Michael Friendly. The history of the cluster heat map. *The American statistician*, 63(2):179–184, 2009.

[87] Julian Wolf, Stephan Hess, David Bachmann, Quentin Lohmeyer, and Mirko Meboldt. Automating areas of interest analysis in mobile eye tracking experiments based on machine learning. *Journal of Eye Movement Research*, 11(6), 2018.

[88] Patricia Wright and Kathryn Fox. Presenting information in tables. *Applied Ergonomics*, 1(4):234–242, 1970.

[89] Tom Yokum. Sparklines: The Tom Thumb of Statistical Graphs. *Foresight: The International Journal of Applied Forecasting*, (14):48–50, Summer 2009.

[90] Daniela Zambarbieri and Elena Carniglia. Eye movement analysis of reading from computer displays, ereaders and printed books. *Ophthalmic and Physiological Optics*, 32(5):390–396, 2012.

[91] Achim Zeileis and Kurt Hornik. Choosing color palettes for statistical graphics. WorkingPaper 41, Department of Statistics and Mathematics, WU Vienna University of Economics and Business, 2006.

[92] Yan Zhang. *The effect of font design characteristics on font legibility*. PhD thesis, Concordia University, 2006.

Appendix A

Additional Information

I include the sample experiment table images in each conditions, the study script the post study questionnaire, and the demographic survey.

Experiment Table Images: There are four sample images of the experiment tables for each condition.

Study Script: This is the full script for the experiment. Words highlighted in orange indicate the parts to be spoken by the experimenter.

Demographic Survey: This is the survey completed by participants before starting the experiment.

Post-Study Questionnaire: A ranking scale questionnaire filled out by participants at the end of each task.

Start of Experiment Table Images

Anime Bar Table

R T

Table : Anime related data

Name	Score	Episodes	Ranked	Popularity	Members	Favorites	Watching	Completed	On-Hold	Dropped	Plan to Watch
Hungry Heart	17	29	12	57	36	60	15	43	29	84	29
Trinity Blood	15	21	34	86	36	30	27	18	25	80	53
Neon Genesis	13	26	18	33	46	38	52	25	50	40	64
Fourth Stage	21	24	11	74	19	20	28	23	68	98	55
Monster Hunter	20	26	35	38	24	89	24	36	28	52	73
Kidou Tenshi	19	22	52	46	12	38	66	51	75	57	67
Ultimate horror	23	15	46	37	19	40	40	14	26	45	22
Fighting Animal	29	12	47	65	17	61	55	14	60	27	71
Black Cats	30	25	48	52	42	55	21	38	18	42	30
Queen Kingdom	13	17	55	54	26	51	47	47	51	69	65
Fafnar Weapon	21	23	65	22	43	55	51	46	57	71	54
Different Life	27	24	31	24	25	57	43	10	16	18	19
Attacking Titan	17	12	30	30	19	54	50	15	12	35	24
Metal Panic	11	24	46	27	21	44	58	37	61	20	25
Helping Alice	25	26	30	41	48	61	24	41	28	57	53
Connect Angel	37	16	27	58	16	23	41	40	50	34	78
Mobile Phone	31	30	61	42	47	78	37	52	54	37	37
Legendary Forks	27	26	27	47	31	47	23	43	35	32	58
Ultra Generation	29	17	22	41	28	41	21	58	64	59	66
First Stage	21	13	31	56	22	69	42	49	49	15	64
Beating Music	19	24	32	40	56	46	16	28	20	19	52
Great Teacher	20	32	47	84	24	18	43	24	27	20	23
Travel World	21	24	44	35	45	59	36	33	65	42	53
Magical Warfare	17	13	28	58	38	36	32	73	29	62	23
Dragon Ball	38	41	35	20	35	45	55	38	31	39	53
Best Humanity	40	26	58	90	28	28	69	27	14	65	45
Assassian World	33	13	22	37	51	23	47	25	58	39	35
King Ranking	24	37	40	34	28	61	42	51	48	73	30
Cursing Notebook	35	26	19	49	35	37	25	23	37	50	55

V M

Canada Color Table

R T

Table : Candy Key nutrition facts

Candy Name	Chocolate	Fruity	Caramel	Almond	Nougat	Hard	Bar	Pluribus	Sugar Percent	Rating	Price
Laffy Taffy	10	29	24	49	15	51	34	63	22	41	16
Boston Beans	17	17	21	24	14	12	27	34	32	37	25
One Dime	14	33	30	31	17	34	24	17	60	46	32
Reese Pieces	27	28	40	57	13	29	34	21	55	23	43
Air Heads	18	21	51	24	30	17	37	46	33	61	41
Lemon Head	24	24	41	24	17	42	31	54	27	53	27
Crunch Time	29	16	37	45	17	58	23	28	70	44	53
Berry Chews	15	13	20	36	16	30	20	17	31	48	17
Now Later	16	26	26	64	33	47	45	36	84	60	75
Good Bar	16	28	28	17	20	63	24	19	70	55	17
Junior Mints	18	24	21	38	19	30	21	12	35	31	37
Gob Stopper	14	37	44	17	17	47	34	20	28	40	33
Cheese Bears	14	15	37	35	27	29	41	24	70	48	30
Haribo Cola	15	27	24	40	21	53	31	21	50	22	41
Milk Duds	26	14	39	33	15	27	14	26	30	55	29
Fruit Chews	27	27	24	44	17	40	32	17	20	96	39
Root Barrels	14	12	37	53	27	53	19	32	27	46	23
Hersheys Kisses	19	14	37	30	21	15	13	22	17	90	16
Pixie Sticks	29	26	24	43	19	58	13	64	55	38	23
Pop Rocks	25	21	39	45	21	53	19	12	29	41	34
Life Savers	12	19	41	47	24	35	27	10	14	40	26
Rolo Chiclets	31	21	40	67	16	18	33	12	13	38	10
Red Vines	15	23	47	37	21	43	17	45	38	37	46
Quarter World	17	18	39	47	25	27	22	27	32	54	19
Caramel Apple	13	23	23	47	26	24	44	37	39	23	43
Jaw Busters	19	14	49	37	12	53	21	45	32	31	35
Baby Ruth	27	31	37	48	17	30	37	54	23	35	25
Fun Dips	28	15	37	25	17	24	23	53	33	30	41
Gold Bears	11	27	34	39	19	53	16	57	49	52	36

V M

Cereal Plain Table

R

Table : Cereal Key nutrition facts

T

Cereal Name	Fat	Protein	Calories	Sodium	Fiber	Carbo	Sugars	Potassium	Score	Iron	Vitamins
Natural Bran	30	33	42	15	24	18	29	35	34	13	37
Cap Crunch	27	17	32	22	23	42	22	15	18	22	53
Post Nat	13	18	72	12	21	22	13	25	23	56	45
Fruity Pebbles	19	32	62	12	55	12	22	29	30	31	63
Mini Wheats	26	33	72	28	31	14	27	22	58	39	42
Bran Chex	46	42	51	23	40	25	46	15	37	21	42
Cinnamon Toast	19	33	56	14	30	43	17	80	42	66	47
Froot Loops	30	33	92	21	50	23	15	42	53	27	25
Almond Delight	19	32	32	35	22	34	18	56	40	24	40
Golden Crisp	37	26	42	29	21	37	21	25	53	19	64
Sweet Combo	27	51	42	22	22	47	59	35	20	39	14
Honey Nut	37	31	52	32	19	34	21	50	30	60	19
Corn Favorite	28	24	51	32	23	18	12	46	16	19	32
Smile Maker	27	42	23	28	41	12	53	25	11	28	34
Wonderful Chex	18	23	37	24	51	24	22	32	41	43	24
Almond Raisin	34	44	45	26	60	17	35	65	13	18	20
Colorful Bran	29	44	72	33	47	15	26	32	68	28	27
Wheat Cream	19	32	32	27	53	21	23	26	32	55	24
Frosted Flakes	21	42	42	22	21	21	32	35	37	39	29
Lucky Charms	47	41	32	37	33	42	23	45	22	34	14
Raisin Bran	21	14	22	22	20	38	38	26	10	12	11
Apple Jacks	13	52	22	25	41	43	33	20	18	22	15
Count Chocula	33	33	62	41	31	12	37	52	42	27	46
Grape Flakes	10	23	42	33	21	31	12	32	65	36	70
Oatmeal Crisp	27	42	22	32	39	31	22	22	36	24	69
Puffed Rice	24	32	52	42	10	18	45	22	44	11	19
Double Chex	56	42	32	37	16	21	24	32	33	17	55
Fruit Fibre	47	63	52	38	38	52	42	22	43	32	46
Quaker Squares	28	21	62	51	32	43	32	22	36	24	37

V

M

Movie Zebra Striping Table

R T

Table : Rating of films in websites

Film	Rotten Tomatoes	MrQE	Metacritic	Netflix	IMDB	Hulu	Film Comment	Roger Ebert	PluggedIn	Naver	Fandango
End Tour	92	59	84	75	79	46	42	40	39	63	15
Red Army	96	36	82	74	77	48	41	39	51	52	35
Black Sea	89	49	71	64	78	44	36	39	29	41	25
Ricki Flash	92	52	77	78	75	46	39	38	46	11	35
Still Alice	19	31	40	44	55	25	20	28	67	52	35
Irrational Man	13	40	24	24	63	35	12	32	75	46	20
My Dreams	94	50	75	40	69	47	38	35	21	33	40
Flights Up	52	47	55	68	61	26	28	31	55	51	31
Hot Pursuit	95	30	50	63	23	25	32	25	21	19	20
Welcome Home	71	47	67	69	59	55	34	30	33	43	14
Running Night	60	46	67	58	60	30	46	55	46	24	35
Tomorrow land	50	53	60	31	66	25	30	30	15	13	25
Mapping Stars	27	61	33	48	52	15	17	10	13	15	16
Crazy Adventure	93	38	91	69	72	45	27	36	63	52	34
About Elly	97	36	87	71	41	45	50	41	35	26	34
Saint Laurent	51	45	23	68	63	42	30	45	26	22	43
Love Mercy	75	44	58	79	78	35	29	39	54	32	45
Water Diviner	32	58	38	52	71	46	47	52	28	64	48
Divergent Series	30	61	42	54	64	35	21	32	61	46	45
Fantastic Four	19	20	27	35	40	65	14	20	31	38	23
Next Door	10	35	32	55	46	35	15	34	18	13	17
Black White	39	15	45	79	49	43	43	40	11	36	30
South Paw	26	60	38	64	69	13	19	35	79	45	45
Pitch Perfect	67	28	63	57	67	35	32	34	69	53	25
Adaline Age	54	38	51	74	73	27	26	37	57	45	25
Teenage Girl	13	37	31	37	49	14	16	45	78	40	24
Wild Tales	71	56	56	64	66	52	28	33	69	35	35
Project Almanac	34	46	47	54	64	17	24	32	51	40	34
Hunting Ground	64	53	54	47	50	32	27	31	37	47	24

V M

End of Experiment Table Images

Start of Study Script

Investigation of data table Script

Pre-Experiment

-Prepare forms
-Open experiment device
-Retrieve table combination for the participant
-test eye tracker (Including Calibrate)
-Start software and test it
-Welcome participant

Experiment Proper

Experiment play (Darker color) (training goes to paper and show to p before each task, and think about training software with different description, after the first task)

<Participant walks into the room>

<Experimenter says to participants>
Hello and welcome to the VIXI lab, thank you for participating in this experiment!, My name is YongFeng Ji, and I will guide you through the experiment. Feel free to ask me any question at any time. Before we start, I need to let you know about your rights as a participant.
If you feel uncomfortable you may quit at any time, and if you do so then I will delete your data right away and it will not be used.
The data that we have collected will be destroyed in the future.
No data will be used without your explicit consent.
Your participation in this experiment is confidential.
Now please read this consent form carefully, which explains your rights as a participant and the conditions of the study, and sign it if you agree with these terms.
You might have a copy of the consent form for your own records
Before we go on with the instructions of the experiment, I would like to let you know that we appreciate your helping us in this study.

<Experimenter asks the participant to sit well>

<Experimenter hands out the consent form and Demographic form to participants>

<Participant filled out two forms and hands them back to Experimenter>
+

<Experimenter explains the table software and different table conditions to the participant>
The software was designed for you to see the tasks and data table pictures. You could change the task screen to the data table screen once you know exactly what the task is asking for by pressing the space bar. Also, the software could help the experimenter to record your answer, and the time spent and save those data. There will be four different letters shown in each corner of the screen, it is designed for better eye tracker precision, you don't need to care about it. There will be four different conditions you will see.
(In the order the participant sees, different pages)
Plain

Film	Rotten Tomatoes	MrQE	Metacritic
Avengers: Age of Ultron	74	86	66
Cinderella	85	80	67
Ant-Man	80	90	64
Do You Believe?	18	84	22
Hot Tub Time Machine 2	14	28	29
The Water Diviner	63	62	50

Zebra Striping table

Film	Rotten Tomatoes	MrQE	Metacritic
Avengers: Age of Ultron	74	86	66
Cinderella	85	80	67
Ant-Man	80	90	64
Do You Believe?	18	84	22
Hot Tub Time Machine 2	14	28	29
The Water Diviner	63	62	50

Color Shading

Film	Rotten Tomatoes	MrQE	Metacritic	Netflix
Avengers: Age of Ultron	74	86	66	7.1
Cinderella	85	80	67	7.5
Ant-Man	80	90	64	8.1
Do You Believe?	18	84	22	4.7
Hot Tub Time Machine 2	14	28	29	3.4

Bar Chart

Film	Rotten Tomatoes	MrQE	Metacritic
Avengers: Age of Ultron	74	86	66
Cinderella	85	80	67
Ant-Man	80	90	64
Do You Believe?	18	84	22
Hot Tub Time Machine 2	14	28	29
The Water Diviner	63	62	50

<Experimenter opens the python file called “training”>
Now, you will use the software to finish the training, the questions will be similar to the ones you will do in a real experiment, during the training, we will simulate the real experiment from the training.

Teach how to use the software
Press space to move on to the next page. Remember to prevent jumping to a different page by pressing the space bar too hard and quick. When you see the red or blue font, it means you have to wait for my response to move on, the instruction is in black font, you need to read that carefully, and once you feel confident with what it is asking, press space to see the table and find the answer.

When found the answer
When you find the answer, don’t tell me the answer right away, instead, press the space bar to move to the next page and then tell me the answer, it could help me to record the time you have spent on that task.

Ok, you may press the space now once you are ready for the training

Real Experiment

<Experimenter asks the participant to wear eye tracker glass and ask the participant to do calibration>
Please sit well, if you are not feeling uncomfortable from your glasses, please keep your eyes focused on the card that is in my hand. <Wait until the screen shows “Calibration success”, if it shows ‘failed’, try all over again>.

<Experimenter opens a python file called ‘experiment –(the number)’ and speaks to the participant>
Now we are about to start the real experiment, the software you use will be the same as the one you did in the training, the difference is there will be more tasks and in between each task, I will hand out a questionnaire that needs you to fill out.

Note

Again, you can use the mouse to find the answer, but not your finger, please sit well during the entire experiment, no questions while you are looking at the data table, you can ask me questions when you are on the task instruction page or the page after the data table. Again, don't give me the answer right after you found it except for the describe task, always press the space bar then tell me. Try to get it right, but do it as fast as possible you can, it is ok to make mistakes.

When you see the red or blue font, it means you must wait for my response to move on, the instruction is in black font, you need to read that carefully, and once you feel confident with what it is asking, press space to see the table and find the answer.

<Participant answers 'yes, I'm ready, then experimenter asks participant to press the space bar.>

Describe Task
<Participant sees the first task instruction, experimenter says>
This is the describe task, you can say whatever you saw from the table, when you feel you have said everything you want to say, feel free to press the space bar. There is no correct answer or wrong answer.

The first task is to ask you to describe what you have seen in the picture, and I will briefly explain what speak aloud is

Speak aloud protocol explanation

Speak aloud protocol, is usually called Think aloud protocol. **Speak aloud protocol** is a method used to gather data in usability testing in product design and development.

Essentially, this involves asking real users to think aloud as they are performing a set of specified tasks using the product being tested. Users are asked to say whatever they are looking at, thinking, doing, and feeling, as they go about their tasks. This enables observers to see how users react to the product and make changes accordingly. Observers at such a test are asked to objectively take notes of everything that users say, without attempting to interpret their actions and words. This method is especially helpful for determining users' expectations and identifying what aspects of the interface are confusing.

In the describe task, you can talk and press the space button whenever you feel you have described enough things, remember, there is no correct answer and no wrong answer.

<Participant starts to do the describe task>

<Participant keeps doing the task until she/he sees blue font text, then experimenter speaks to them>
Now, you have finished the describe task, from now, you will see different types of tasks, Again, don't give me the answer right after you found it, always press the space bar then tell me. Try to get it right, but do it as fast as possible you can, it is ok to make mistakes.

When finished describe task
<Experimenter open software training>
Before we move onto the next task, I want you to get used to the software. Now, you need to complete several very easy tasks via the software. When you see red and blue font, it means you have to wait for my response to move on. When it is black font, it means you should read the instruction carefully and remember what it is asking for, once you are ready, you can press space button to find the answer. Important thing is, you must say your answer after you press the space button, when you find the answer, don't tell me directly, instead, press the space button then tell me.

When participant finish training
<Experiment reopen real experiment software>
If you have no question with the software, you may restart the experiment to see different tasks.

Non-describe Task
Finding maximum value vertically
<Experimenter gives the task instruction sheet and explains to the participant about the second task: Find the row which has the largest value in a given column>
This is another task you need to finish; it asks you to find the column and then find which row has the largest value in that column. Please read the instruction carefully and say the answer.

When participants feel confident of the task.
<Participant presses the space button and starts to work on the second task.>

<Participant finishes one trial, experiment start to collect the given answer by clicking it on the monitor screen, then speaks to the participant that ‘You can press the space button to move forward’>

The procedure will be repeated until

<When the participant is on the page that indicates she/he has finished this type of task, the experimenter hand out a questionnaire to the participant and says>

You have finished this task, please fill out the questionnaire, and when you finish, give it back to me.

<Participant finished the questionnaire and gives back to the experimenter.>

<Experimenter allows the participant to press the space bar to continue the experiment.>

Finding maximum value horizontally
<Experimenter gives the task instruction sheet and explains to the participant about the third task: Find the row which has the largest value in a given column>
This is another task you need to finish; it asks you to find the row and then find which column has the largest value in that row. Please read the instructions carefully and say the answer.

When participants feel confident of the task.
<Participant presses the space button and starts to work on the second task.>

<Participant finishes one trial, experiment start to collect the given answer by clicking it on the monitor screen, then speaks to the participant that ‘You can press the space button to move forward’>

The procedure will be repeated until

<When the participant is on the page that indicates she/he has finished this type of task, the experimenter hand out a questionnaire to the participant and says>

You have finished this task, please fill out the questionnaire, and when you finish, give it back to me.

<Participant finished the questionnaire and gives back to the experimenter.>

Find cell value

<Experimenter gives the task instruction sheet and explains to the participant about the fourth task: Find the row which has the largest value in a given column>

This is another task you need to finish; It asks you to find the value according to the column name and row name that has been given in the task instruction. Please read the instructions carefully and say the answer.

When participants feel confident of the task.

<Participant presses the space button and starts to work on the second task.>

<Participant finishes one trial, experiment start to collect the given answer by clicking it on the monitor screen, then speaks to the participant that 'You can press the space button to move forward'>

The procedure will be repeated until

<When the participant is on the page that indicates she/he has finished this type of task, the experimenter hand out a questionnaire to the participant and says>

You have finished this task, please fill out the questionnaire, and when you finish, give it back to me.

<Participant finished the questionnaire and gives back to the experimenter.>

<Experimenter allows the participant to press the space button to continue the experiment.>

Proportionally different

<Experimenter gives the task instruction sheet and explains to the participant about the last task: Find the row which has the largest value in a given column>

This is the last task you need to finish; It asks you to find the most proportional difference row between two rows and provides the corresponding column.

Proportional difference explanation

Proportion is a mathematical concept, which states the equality of two ratios or fractions. It refers to some a category over the total. When two sets of numbers, increase or decrease in the same ratio, they are said to be directly proportional to each other.

Imagine you are viewing two food review websites, the first one is in the range of 0-10, and the other one is in the range of 0-100. Food A is rated 4 on the first website and 70 on the second website, Food B is rated 8 on the first website and 60 on the second website. Regardless of the authority on each website, can we say Food A is much better than Food B because Food A is 10 points ahead of Food B on the second website, but only have 4 points difference on the first website? No, we cannot say that because of two websites are taking different measurements.

When participants feel confident of the task.
<Participant presses the space button and starts to work on the second task.>

<Participant finishes one trial, experiment start to collect the given answer by clicking it on the monitor screen, then speaks to the participant that ‘You can press the space button to move forward’>

The procedure will be repeated until

<When the participant is on the page that indicates she/he has finished this type of task, the experimenter hand out a questionnaire to the participant and says>

You have finished this task, please fill out the questionnaire, and when you finish, give it back to me.

<Participant finished the questionnaire and gives back to the experimenter.>

<Experimenter allows the participant to press the space bar to continue the experiment.>

The procedure will be repeated until
<Participant is on the page that says ‘You have finished all the tasks’, experimenter says>

Congratulation, you have finished all the tasks, please help me to fill the rest of the post-study questionnaire.

<Participant finishes the questionnaire and gives it back to the experimenter.>

End of Experiment
<Experimenters give money to participant>

<Experimenters ask the participant to sign a receipt>

Thank you very much for your participation. Do you have any questions?

End of Study Script

Start of Demographic Survey

Draft of Demographic Survey

Note that specific formulations might vary slightly

Pre-study questionnaire

[to fill before performing the tasks in the study]

Demographic Information

This information is collected for demographic purposes only. All questions are optional.

Age: _____

Sex: ☐ Male ☐ Female ☐ Other _________________

Are you currently a student? ☐ Yes ☐ No

If you are a student, please indicate your **current level of ongoing education**. If you are not a student, please indicate **the highest level of education** you have **completed**: (lower level to higher level)

☐ High school or equivalent
☐ Vocational/technical school (2 year)
☐ University degree (4 year), e.g. BSc, BA
☐ Master’s degree
☐ Doctoral degree (PhD)
☐ Professional degree (MD, JD, etc)
☐ Other _____________________

What is your field of study or expertise? ____

How often do you encounter data tables in your professional life?
Note that we define tables in a very generic way, as any two-dimensional display of information that uses rows and columns to show data (e.g., numbers, categories, notes) of multiple items (usually as rows) in which each item has multiple attributes (usually in the different columns). This includes:

- Excel files
- Excel files printed on paper
- Itemized budgets in table forms

- List of files that contain multiple columns (e.g., name, size, type) in your operating system
- Tables in scientific or economic documents, including articles.

Tick the appropriate box.

□ I come across data tables multiple times per day
□ I come across data tables most days
□ I come across data tables most weeks
□ I come across data tables most months
□ I rarely come across data tables

Please add here any other information that you think we should know:

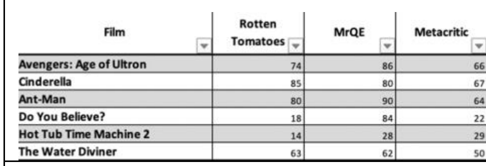
____

End of Demographic Survey

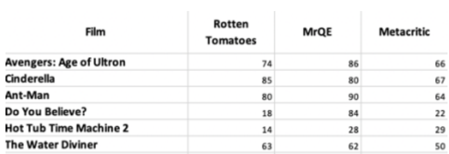
Start of Post-Study Questionnaire

Questionnaire

Rank the different conditions that you have worked with from 1 to 4, where 1 means most helpful or best, and 4 means least helpful based on the aspect described in the columns.

<table>
<tr><th></th><th>Speed: rank conditions in terms of how they allowed you to be fast (1 is fastest, 4 is slowest)</th><th>Accuracy: rank conditions in terms of how they allowed you to be accurate (1 is most accurate, 4 is least accurate)</th><th>Rank the conditions in terms of your preference (1 is most preferred, 4 is least preferred)</th></tr>
<tr><td>Zebra Striping

<table>
<tr><th>Film</th><th>Rotten Tomatoes</th><th>MrQE</th><th>Metacritic</th></tr>
<tr><td>Avengers: Age of Ultron</td><td>74</td><td>86</td><td>66</td></tr>
<tr><td>Cinderella</td><td>85</td><td>80</td><td>67</td></tr>
<tr><td>Ant-Man</td><td>80</td><td>90</td><td>64</td></tr>
<tr><td>Do You Believe?</td><td>18</td><td>84</td><td>22</td></tr>
<tr><td>Hot Tub Time Machine 2</td><td>14</td><td>28</td><td>29</td></tr>
<tr><td>The Water Diviner</td><td>63</td><td>62</td><td>50</td></tr>
</table></td><td></td><td></td><td></td></tr>
<tr><td>Color Shading

<table>
<tr><th>Film</th><th>Rotten Tomatoes</th><th>MrQE</th><th>Metacritic</th><th>Netflix</th></tr>
<tr><td>Avengers: Age of Ultron</td><td>74</td><td>86</td><td>66</td><td>7.1</td></tr>
<tr><td>Cinderella</td><td>85</td><td>80</td><td>67</td><td>7.5</td></tr>
<tr><td>Ant-Man</td><td>80</td><td>90</td><td>64</td><td>8.1</td></tr>
<tr><td>Do You Believe?</td><td>18</td><td>84</td><td>22</td><td>4.7</td></tr>
<tr><td>Hot Tub Time Machine 2</td><td>14</td><td>28</td><td>29</td><td>3.4</td></tr>
</table></td><td></td><td></td><td></td></tr>
<tr><td>Bar Chart

<table>
<tr><th>Film</th><th>Rotten Tomatoes</th><th>MrQE</th><th>Metacritic</th></tr>
<tr><td>Avengers: Age of Ultron</td><td>74</td><td>86</td><td>66</td></tr>
<tr><td>Cinderella</td><td>85</td><td>80</td><td>67</td></tr>
<tr><td>Ant-Man</td><td>80</td><td>90</td><td>64</td></tr>
<tr><td>Do You Believe?</td><td>18</td><td>84</td><td>22</td></tr>
<tr><td>Hot Tub Time Machine 2</td><td>14</td><td>28</td><td>29</td></tr>
<tr><td>The Water Diviner</td><td>63</td><td>62</td><td>50</td></tr>
</table></td><td></td><td></td><td></td></tr>
<tr><td>Plain

<table>
<tr><th>Film</th><th>Rotten Tomatoes</th><th>MrQE</th><th>Metacritic</th></tr>
<tr><td>Avengers: Age of Ultron</td><td>74</td><td>86</td><td>66</td></tr>
<tr><td>Cinderella</td><td>85</td><td>80</td><td>67</td></tr>
<tr><td>Ant-Man</td><td>80</td><td>90</td><td>64</td></tr>
<tr><td>Do You Believe?</td><td>18</td><td>84</td><td>22</td></tr>
<tr><td>Hot Tub Time Machine 2</td><td>14</td><td>28</td><td>29</td></tr>
<tr><td>The Water Diviner</td><td>63</td><td>62</td><td>50</td></tr>
</table></td><td></td><td></td><td></td></tr>
</table>

Please explain your top choices: Why was the condition best for each aspect?

____

____

____

Do you have more comments?

____

End of Post-Study Questionnaire