Exploring Pie Charts and Part-To-Whole Alternatives: Eye-tracking Approach

D. Bačić, A Krbanjevic and N. Jukić
Loyola University Chicago, Chicago, USA
dbacic@luc.edu, akrbanjevic@luc.edu, njukic@luc.edu

***Abstract* - In the field of business information visualization (BIV), there are numerous options for displaying data. One popular yet debated choice is the use of pie charts, as their effectiveness and efficiency have been called into question. Despite their widespread use, there is limited empirical research on the effectiveness and efficiency of pie charts compared to other display types. Additionally, existing research primarily focuses on measuring user impact through accuracy and time, neglecting the impact on users' physiological attention and cognitive resources. This study aims to fill this gap by comparing the use of pie charts to seven alternative data representation types in a part-to-whole task. The study employed a randomized, within-subject experiment with 21 participants, utilizing eye-tracking technology to evaluate user performance (accuracy and time) and cognitive effort (eye fixation). The results showed that pie charts were more accurate than bar charts but required more cognitive effort than stacked bars and treemaps. Donut charts required the most cognitive effort among all tested data representations. The study highlighted that time to complete the task may not be the best indicator of user experience and accuracy while elevating the importance of accounting for users' cognitive resources.**

Keywords - pie chart; cognitive effort; eye tracking; data visualization; pie chart alternatives; part-to-whole task

I. INTRODUCTION

Given the relative ease of chart creation using commercial business intelligence tools, visual representations such as pie charts, bar charts, treemaps, and bubble charts are gaining popularity amongst data analysts and other business practitioners. Continued practitioner (especially their executive ranks) and academic fascination with pie chart data representation when displaying solutions to part-to-whole problems is a driving motivation for the present research. There is a lack of conclusive empirical data focused on understanding overall pie chart effectiveness, efficiency, and comparison to alternative display types. Similarly, the available research is almost exclusively focused on measuring user impact through accuracy and time and is not addressing users' physiological attentional resources and cognition. We address this gap by conducting an experiment comparing pie charts to seven alternatives and evaluating user impact relative to performance (accuracy, time) and cognitive effort using eye-tracking technology.

Despite their popularity, the use of pie charts has become heavily criticized [1], [2], in several foundational [3] and recent studies favoring the alternative forms of displays [4]–[7]. For proponents of pie charts, there is a relative consensus that pie charts are best when used for part-to-whole relationships, with the slice size relating to other parts or the whole [7]. However, the proliferation of pie chart misuse (3D display option, large number of pie slices, poor labeling, and color choices) and general lack of understanding of how users 'read' pie charts [8] has resulted in two vocal camps and a vigorous debate centered on the pie chart's appropriateness and user impact.

This current study aims to impact the ongoing discussion and debate about pie charts. More specifically, this study seeks to explore a more thorough evaluation of pie charts by comparing them to alternative options. This evaluation will focus on the impact of pie charts and other part-to-whole alternatives on users and will consider various aspects, including visual attention, cognitive effort, judgment accuracy, and speed of judgment. In the process, the study will utilize the latest eye-tracking technology.

The study will begin by briefly addressing the relevant literature related to pie charts and alternative options, as well as cognitive effort measured through eye-tracking. This will provide the necessary context and background information to understand the purpose and significance of the experiment. Following the literature review, the study will provide an overview of the experiment procedures and setup. The next stage of the study will analyze and present the collected experimental data. The results will be discussed and analyzed, taking into account the findings' implications and limitations and providing suggestions for future research.

II. LITERATURE REVIEW

The first instance of a pie chart was in William Playfair's 1801 Statistical Breviary [9], which defined the pie chart as a simple information graphic whose principal purpose is to show the relationship of a part to the whole [10]. The academic critiques of pie charts eventually led to their growing popularity within business dashboards, flashy presentations, and popular media [2], [11]. A study pitted pie charts against bar charts and ultimately found that pie charts outperformed stacked bar charts but not all types of bar charts [3]; another set of different studies found that pie charts were useful and effective in part-to-whole comparisons [12] for quick comparison calculations. In the context of alternative charts, a study found that treemaps performed worse than pie charts in accuracy, while a pie chart took longer to read and that slice size itself may impact user performance irrespective of display type [4].

Although academic criticism of pie charts continued [4], [7], [8], it is clear that an ongoing debate has prompted a necessity to investigate further user impact when assessing pie charts and their alternatives. While there has been some progress made in our understanding of pie chart [13] and its alternatives usage through observing user performance [4]–[7], [14] little progress has been made in measuring users' visual attention and cognitive effort, a primary contribution of the present study.

As part-to-whole data visualization literature evaluated user performance through accuracy and time, suggestions have been made that cognitive processes and cognitive effort may be promising avenues to clarify existing findings and inconclusive results [2], [4]. Cognitive effort can be defined as cognitive resources (perceptions, memory, and judgment) needed to complete a task [15]. There are several ways to measure cognitive effort, physiological [16] and perceptual [17]. Technological advances have been made in more recent times, and relevant research found numerous ways to measure cognitive effort physiologically through fMRI, EEG, and eye tracking [15], [16], [18]–[21].

Eye-tracking has been established as the primary physiological sensing technology used to analyze visual stimuli, as it provides valuable insight into a user's attention and effort while they are reading and scanning displayed information [20], [21]. This is achieved by capturing metrics based on eye fixation, saccades, blinking, and pupil dilation [18]. Beyond psychology, eye tracking has been widely adopted in various fields to measure visual attention distribution and assess the impact of user interactions. Salient to this research, studies have found that visualizations with improper use of colors can lead to increased cognitive effort and longer decision-making times [22]. Additionally, eye tracking has been utilized to examine conventional dashboard design guidelines [23], cognitive effort as the cognitive fit mechanism using tables and graphs [21], and to evaluate the impact of pie chart visual properties (arc, angle, area) [13]. Fixation-based eye tracking measures (fixation is the stabilization of the eye on an object), such as fixation count and duration, continue to be the preferred method for evaluating cognitive processes and effort [16], [21], [13].

III. Experiment

We conducted an experiment to evaluate users' performance and cognitive effort when evaluating pie charts and alternative charts. In our experiment, we measured user performance through accuracy, speed, and effort. Demographic data (education level, age, gender identity, and major) and task answers were collected through a survey module integrated into the iMotions biometric platform. Eye tracking data was collected using the SmartEye AI-X EyeTracker (0.5-degree accuracy, 60hz sample rate, 0.01-degree precision). All images were presented on a 22", 1920 x 1080 resolution monitor, and collected data were stored and analyzed using iMotions software (version 9.1). The default iMotions I-VT filter setting was used to process gaze information, including minimum fixation duration of 60ms.

A. Experimental Design

We conducted a within-subject experiment where chart type, chart order, and part size (2%, 3%, 5%, 9%, 16%, 19%, 62%, 67%) were fully randomized to avoid bias due to the learning effect. Subjects were asked to estimate the size of the part relative to the whole (100%) for a category.

A randomization algorithm was created to present each subject with all eight chart types and all part sizes (in total). For example, subject 1 was presented with eight experimental conditions (part size in parenthesis): pie chart (participant is shown 1 of the 8 pie charts, for example, the one with 16%-part size), horizontal bar (5%), vertical bar (2%), lollipop chart (3%), bubble chart (19%), stacked bar (9%), treemap (67%), and donut chart (62%). The combinations for subject 2 will then be randomized so that they were also given one instance of each part size combined with different charts yet still exposed to each chart type (also randomized) (Figure 1).

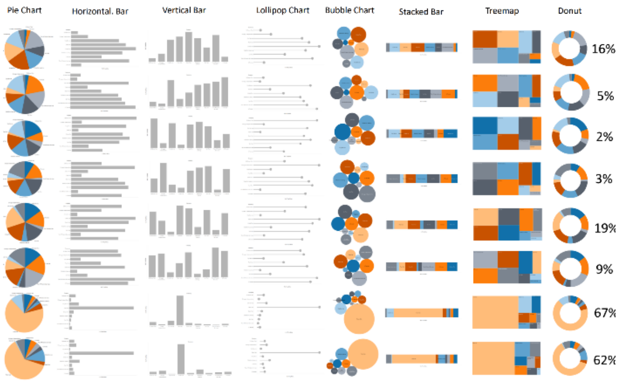

Figure 1: Fully randomized within subject design

B. Participants

Students from a large private midwestern university participated and were recruited from two business undergraduate classes (Data Visualization, Introduction to Information Systems). Students were rewarded with a small number of class bonus points for participating. There were 26 participants, 50% males, all undergraduates, mostly business majors, 77% with junior or senior class standing, and an average age of 20.53 years (Table 1).

TABLE 1: PARTICIPANTS

Gender				**Age**		
Male	13	50.0%		18	1	3.8%
Female	12	46.2%		19	1	3.8%
Non-Binary	1	3.8%		20	10	38.5%
				21	11	42.3%
Class				22	3	11.5%
Freshman	2					
Sophomore	14			**Major**		
Junior	6			Business	25	96.2%
Senior	4			Non-business	1	3.8%

C. Experimental Conditions

As mentioned, subjects were given a baseline pie chart and seven alternative charts (Figure 2) while controlling for the number of categories (each chart displayed eight).

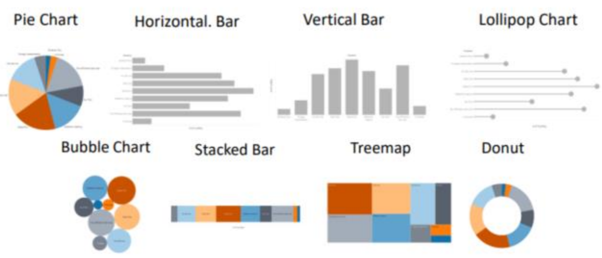

Figure 2: Experimental conditions

We implemented standard best practices found in chart design, such as using a color-blind palette. For each chart, except the donut chart, the categories were directly labeled on the chart itself. The color was used in addition to text to identify categories in all cases except in horizontal bar, vertical bar, and lollipop charts to preserve graphical best practices and mimic what practitioners experience when designing these types of charts.

D. Variables

Cognitive effort and visual attention were measured through fixation duration (average length of all fixations) and count (the number of eye fixations). Accuracy was measured as the subjects' answer nominal % difference (error) relative to the correct answer (part size). Time was measured in milliseconds.

IV. Data Analysis

Data from five participants were excluded from the analysis due to poor eye-tracking calibration and misunderstanding of the task. After removing invalid data points from five respondents, the remaining data from the 21 participants (43% males), all undergraduates, average age= 20.6 yrs (SD=0.18)) was used for analysis.

Before evaluating performance across experimental conditions (chart types), participant performance is compared across all possible task solutions for all chart types combined (Table 2).

TABLE 2: ACCURACY ACROSS PART SIZES

Task Solution (a)	Avg. Solution (b)	Error (Nom) (b-a)	Error Size % (b-a)/a
2%	6.6%	4.6%	232%
3%	9.3%	6.3%	211%
5%	14.2%	9.2%	185%
9%	18.9%	9.9%	110%
16%	27.6%	11.6%	72%
19%	28.8%	9.8%	52%
62%	68.8%	6.8%	11%
67%	71.6%	4.6%	7%

Recall that task solutions (for example, in the case of a pie chart, the solution is slice size % relative to total pie) were varied and randomized to minimize the potential bias if using constant/single task solution. This approach eliminates the learning bias and aids in the generalizability of the results. The average error size % ranges from 7% (for tasks with 67% as the correct solution) to near or over 200% (for tasks with 2%, 3%, and 5% as correct solutions, which was to be expected).

In addition to the error size, users' performance data included time, fixation count, and fixation duration. Given the within-subject design, and since we wanted to specifically compare the user performance using a typical 2D pie chart to each alternative, a pairwise t-test comparing the respective means was conducted (t-test, paired two sample for means). Its findings are summarized in Table 3 (Accuracy and Time) and Table 4 (Fixation count and duration).

In case of accuracy, a significant difference was found in four pairs; between pie (M=4.643, SD=4.090) and horizontal bar (M=16.238, SD= 18.3356); t(20)=-3.1195, p=0.0054, pie and vertical bar (M=15.119, SD= 14.734); t(20)=-3.2431, p=0.0041, pie and lollipop (M=13.667, SD=14.2945); t(20)=-3.1051, p=0.0056, and pie and bubble (M=8.00 , SD=5.992); t(20)=-2.1871, p=0.0408.

TABLE 3: ACCURACY AND TIME

Chart Type	Accuracy (Nom. Error)		Time (Milliseconds)	
	Mean	SD	Mean	SD
Pie Chart	4.643	4.090	14192.3	8094.3
Horizontal Bar	16.238	18.3356	17801.1	6142.21
	t(20)=-3.1195, **p=0.0054**		t(20)=-1.9218, **p=0.069**	
Vertical Bar	15.119	14.734	12715.4	5967.0
	t(20)=-3.2341, **p=0.0041**		t(20)=0.8635, p=0.3981	
Lollipop	13.667	14.294	1773.1	11315.4
	t(20)=-3.1051, **p=0.0056**		t(20)=-1.3099, p=0.2051	
Bubble Chart	8.00	5.992	13761.2	8122.8
	t(20)=-2.1871 **p=0.0408**		t(20)=0.2192 p=0.8287	
Stacked Bar	4.667	4.3050	13117	5865.6
	t(20)=-0.0198, p=0.9844		t(20)=0.6330, p=0.5339	
Treemap	5.857	6.7474	14569.3	5551.97
	t(20)=-0.8172, p=0.4234		t(20)=-0.1998, p=0.8436	
Donut Chart	5.5238	4.366	13361.7	7252.17
	t(20)=-0.7941, p=0.4365		t(20)=0.5352, p=0.5979	

In the case of time, a significant difference was found in one pair; between pie (M=14192.3, SD=8094.33) and horizontal bar (M=17801.1, SD=6142.21); t(20)=(-1.9218), p=0.069[1].

In the case of fixation count, a significant difference was found in three pairs; between pie (M=37.6, SD=18); and horizontal bar (M=51.81, SD=17.64); t(20)=-2.7510, p=0.0123; pie and lollipop (M=51.8, SD=31.4); t(20)=-1.8604, p=0.0776; and pie and donut (M=31.381, SD=14.2658); t(20)=1.7324, p=0.0986. For the fixation duration, a significant difference was found in five pairs; between pie (M=286.921, SD=67.48) and horizontal bar (M=243.707, SD=41.907); t(20)=3.841, p=0.0010; pie and vertical bar (M=237.404, SD=49.686); t(20)=4.31, p=0.0003, pie and lollipop (M=234.853, SD=39.30); t(20)=4.293, p=0.0004, pie and treemap (M=246.744, SD=56.93); t(20)=2.894, p=0.01, and pie and donut chart (M=326.574 , SD=90.588); t(20)=-3.206, p=0.004.[2]

[1] To minimize the potential for Type 1 error results could be evaluated α level of 0.014 (=0.1/7 – seven tests). At this level, the difference in accuracy between a pie chart and bubble chart, and in time between a pie chart and vertical bar becomes insignificant.

[2] When evaluating the results at the α level of 0.014 (to control for Type 1 error), the findings remain unchanged for fixation duration, while for fixation count the only the difference between pie and horizontal chart is significant.

TABLE 4: FIXATION COUNT AND DURATION

Chart Type	Fixation Count		Fixation Duration	
	Mean	SD	Mean	SD
Pie Chart	37.6	18	286.921	67.48
Horizontal Bar	51.81	17.64	243.707	41.907
	t(20)=-2.7510, **p=0.0123**		t(20)=3.841, **p=0.0010**	
Vertical Bar	38.143	16.593	237.404	49.686
	t(20)=-0.1042, p=0.9181		t(20)=4.31, **p=0.0003**	
Lollipop	51.8	31.4	234.853	39.30
	t(20)=-1.8604, **p=0.0776**		t(20)=4.293, **p=0.0004**	
Bubble Chart	37.238	19.071	275.665	48.073
	t(20)=0.0738, p=0.942		t(20)=0.908, p=0.375	
Stacked Bar	36.905	16.577	266.184	60.208
	t(20)=0.1471, p=0.885		t(20)=1.484, p=0.1534	
Treemap	42.5	15.2	246.744	56.93
	t(20)=-0.9708, p=0.3432		t(20)=2.894, **p=0.01**	
Donut Chart	31.381	13.793	326.574	90.588
	t(20)=1.7324, **p=0.0986**		t(20)=-3.206, **p=0.004**	

V. DISCUSSION

Our goal was to understand the relative performance of selected alternatives compared to commonly used pie charts. Unlike existing research that primarily focused on the evaluation of accuracy and speed, we also focused on capturing users' visual attention and cognitive effort by measuring their eye fixation count and duration. In the process, we discovered several important findings.

First, regarding accuracy (estimation error), pie charts outperformed alternatives focusing on data encoding through length, including a horizontal bar, vertical bar, and lollipop. This confirms that data encoding through length is ineffective for part-to-whole tasks. A likely explanation is that bar charts provide a challenging way to estimate the total compared to area-based charts. Interestingly, the pie chart also outperformed the bubble chart (by almost 100%), a data representation that encodes data through the size of the area of the circle. Participants did not perform better when using other data representations that rely on data encoding through the area (stacked bar, donut chart, and treemap) than when using a traditional 2D pie chart. When it comes to time, the only significant difference measured when comparing the pie chart to a horizontal bar. Therefore, time may not be the best way to understand users' efforts or cognitive processes that lead to differences in accuracy.

Second, our findings on physiological indicators of attention and effort provide a more nuanced understanding of user impact. More specifically, although there was no difference in accuracy and time, our results indicate that the pie chart requires a longer fixation duration than the stacked bar and treemap. Hence, this is the first study to provide biometric evidence that pie charts require more effort to achieve the same level of accuracy and speed. Alternatively, we found that the donut chart requires more effort than all tested alternatives, including the pie chart. Our findings suggest that the recent increase in popularity of donut charts may come at the expense of users' cognitive effort and its use should be carefully considered in practice.

Third, our experiment suggests that any experiments using part-to-whole analysis must control the potential impact of part size. We found that small data parts (slices/areas/segments) generate more significant accuracy errors when compared to larger data parts. Although not surprising, we provide empirical evidence of the bias and caution others in generalizing any findings if not controlling for a potential effect of a data part. We successfully controlled for part size by randomizing the data part in addition to randomizing chart type and chart type order.

VI. LIMITATIONS AND FUTURE RESEARCH

Future research should consider other ways in which eye-tracking data can provide insights through gaze heatmaps (Figure 3), fixation sequences for viewing patterns, areas of interest for understanding revisits, and potential confusion and uncertainty. Similarly, saccadic movement-based metrics such as peak amplitude count, peak velocity, and pupillometry can provide a more nuanced understanding of user impact. One way to deploy this technology in future research is to focus on part-size user implications. In our exploratory post-hoc analysis of gaze data (heat map replays), we found initial indications of users potentially deploying different visual strategies when evaluating pie slices due to their size (small vs. large) or position (relative to easy to estimate 0, 45, 90, 135, and 180 angles). This potential research stream further amplifies the need for care in experiment design (controls) regarding data/part size.

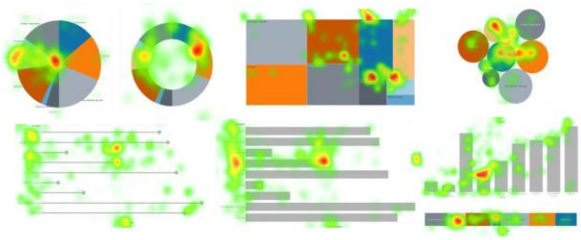

Figure 3: Gaze heat map

While our findings offer practical implications, they should be generalized only to the part-to-whole task type used in the present experiment. Future research should consider the use of pie charts in other task types. For example, tasks to assess the overall distribution of categories and understand the dominant/minority groups. Another research avenue may be analyzing business insights from pie charts and their alternatives.

Although a fully randomized and within-subject design allows for detecting significant relationships, future research should focus on replicating our findings with more participants. The present study, with only 21 participants, allows for Type II error (false negatives), therefore, the results should be interpreted cautiously. The participants were all undergraduate students from a single university, which may limit the generalizability of the results to other populations. Future research should consider expanding the population's size, age, occupation, and geographic distribution.

The experiment was conducted in a controlled lab setting, which may not accurately reflect how individuals use charts in real-world scenarios. While fixation duration and fixation count can provide insights into visual attention, they may not fully capture cognitive effort. Other measures such as subjective ratings or physiological (pupillometry, EEG, fMRI) measures could be used to supplement the fixation data.

Lastly, we used charts with eight categories. Future research should focus on understanding the potential impact of the number of categories (e.g., the number of pie slices) on user performance.

VII. CONCLUSION

In conclusion, the present study provides evidence that eye-tracking, particularly fixation duration, can add valuable insights regarding user impact when using pie charts and other data representations in solving part-to-whole tasks. First, although pie charts outperform length-based charts (vertical and horizontal bar charts and lollipop chart) when it comes to accuracy, this comes at the expense of more significant effort (fixation duration). On the other hand, while yielding the same accuracy, pie charts require more effort than alternative area-based charts (treemap and stacked bar). Second, in part-to-whole tasks, time is a poor predictor of accuracy and effort. Third, donut charts required the most cognitive effort among all tested data representations. Lastly, users tend to be biased toward overestimating the part size. However, the overestimation is not constant, suggesting a need for care when designing experiments focused on part-of-whole tasks.

REFERENCES

[1] E. R. Tufte, “The visual display of information.” Cheshire, Ct.: Graphics Press, 1983.

[2] C. N. Knaflic, *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2015.

[3] W. S. Cleveland and R. McGill, “Graphical perception: Theory, experimentation, and application to the development of graphical methods,” *J Am Stat Assoc*, vol. 79, no. 387, pp. 531–554, 1984.

[4] R. Kosara, “The Impact of Distribution and Chart Type on Part-to-Whole Comparisons.,” in *EuroVis (Short Papers)*, 2019, pp. 7–11.

[5] V. T. Nguyen, K. Jung, and V. Gupta, “Examining data visualization pitfalls in scientific publications,” *Vis Comput Ind Biomed Art*, vol. 4, no. 1, pp. 1–15, 2021.

[6] F. E. Sandnes, A. Flønes, W.-T. Kao, P. Harrington, and M. Issa, “Searching for extreme portions in distributions: A comparison of pie and bar charts,” in *Cooperative Design, Visualization, and Engineering: 17th International Conference, CDVE 2020, Bangkok, Thailand, October 25–28, 2020, Proceedings 17*, 2020, pp. 342–351.

[7] H. Siirtola, “The cost of pie charts,” in *2019 23rd International Conference Information Visualisation (IV)*, 2019, pp. 151–156.

[8] D. Skau and R. Kosara, “Arcs, angles, or areas: Individual data encodings in pie and donut charts,” in *Computer Graphics Forum*, 2016, vol. 35, no. 3, pp. 121–130.

[9] W. Playfair, *The statistical breviary*. Wallis, 1801.

[10] I. Spence and S. Lewandowsky, “Displaying proportions and percentages,” *Appl Cogn Psychol*, vol. 5, no. 1, pp. 61–77, 1991.

[11] S. Few, “Data visualization for human perception,” *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*, 2014.

[12] D. Simkin and R. Hastie, “An information-processing analysis of graph perception,” *J Am Stat Assoc*, vol. 82, no. 398, pp. 454–465, 1987.

[13] A. Krbanjevic and D. Bačić, “Exploring Pie Chart Visual Properties and Design Elements: Eye-Tracking Approach,” in *AMCIS 2022 Proceedings*, Aug. 2002.

[14] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman, “The science of visual data communication: What works,” *Psychological Science in the public interest*, vol. 22, no. 3, pp. 110–161, 2021.

[15] E. Cooper-Martin, “Measures of cognitive effort,” *Mark Lett*, vol. 5, pp. 43–56, 1994.

[16] M. Shojaeizadeh, S. Djamasbi, R. C. Paffenroth, and A. C. Trapp, “Detecting task demand via an eye tracking machine learning system,” *Decis Support Syst*, vol. 116, pp. 91–101, 2019.

[17] Dinko. Bačić and R. Henry, “Task-Representation fit’s impact on cognitive effort in the context of decision timeliness and accuracy: A cognitive fit perspective,” *AIS Transactions on Human-Computer Interaction*, vol. 10, no. 3, pp. 164–187, 2018.

[18] M. K. Eckstein, B. Guerra-Carrillo, A. T. M. Singley, and S. A. Bunge, “Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?,” *Dev Cogn Neurosci*, vol. 25, pp. 69–91, 2017.

[19] K. Kurzhals, B. Fisher, M. Burch, and D. Weiskopf, “Eye tracking evaluation of visual analytics,” *Inf Vis*, vol. 15, no. 4, pp. 340–358, 2016.

[20] D. Bačić, “Biometrics and business information visualization: Research review, agenda and opportunities,” in *HCI in Business, Government, and Organizations: 5th International Conference, HCIBGO 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings 5*, 2018, pp. 671–686.

[21] D. Bačić and R. Henry, “Advancing our understanding and assessment of cognitive effort in the cognitive fit theory and data visualization context: Eye tracking-based approach,” *Decis Support Syst*, vol. 163, p. 113862, 2022.

[22] P. Bera, “How colors in business dashboards affect users’ decision making,” *Commun ACM*, vol. 59, no. 4, pp. 50–57, 2016.

[23] D. Peebles and P. C.-H. Cheng, “Modeling the effect of task and graphical representation on response latency in a graph reading task,” *Hum Factors*, vol. 45, no. 1, pp. 28–46, 2003.