Navigating Visual Information: Understanding Audience Perception and Evaluation in Different Data Visualizations

Joesuan, Ch'ng and *Sing Yan, Looi*[*]

Faculty of Science, Universiti Tunku Abdul Rahman, 31900 Kampar, Perak, Malaysia.

Abstract. In the age of Big Data where data literacy is vital across diverse domains, the prevalence of misleading visualization raises significant concerns. Examining the extent of such visualizations is crucial since viewers often lack the ability to choose the form of presentation. This study aims to investigate the impact of intentionally misleading data visualizations on cognitive biases by exploring factors that potentially influence perceptions and evaluations. A factorial experimental design with three factors involving methods of data visualization, audience academic background and the sequence of data presentation is employed. A total of 60 undergraduate students with two different major programmes from a local higher educational institution participated in this experiment. These students were tasked with responding to predesigned questions based on two different sets of infographics addressing the same issues. The findings indicate that both data presentation and analytical background significantly influence audience evaluation. Additionally, the order of data presentation reveals that audience evaluation is influenced by their initial negative impression. These results underscore the critical role of data literacy in enhancing the understanding of visual information, particularly in the context of public issues.

1 Background

In the era of Big Data, the volume of data presented in visual form is experiencing an unprecedented surge. Data visualization is defined as the utilization of images to represent information derived from data [1] that is displayed in tabular forms or various graphical representations. Contemporary organizations are intended to employ a significantly larger volume of data visualizations than before [2]. Generally, reliance on data visualization is crucial for extracting meaningful insights from extensive datasets [3]. Despite its prevalence, misleading data visualization can lead to misperception and misinterpretation, consequently resulting in faulty judgments and poor decisions. Misleading visualizations are mainly defined as charts that interfere with the viewer's ability to accurately read and compare values [4].

[*]Corresponding author: looisy@utar.edu.my

The severity of the issue amplifies when the data fails to support the intended claims. It is noticed that some visualizations are crafted to guide the public toward certain agendas that are commonly observed in political promotions and commercial advertisements. In such cases, there is a tendency to manipulate visualizations to create an appearance of support, causing misleading presentations [5]. Charts are relatively persuasive as they relate to logical and scientific thinking, especially when supporting an argument [6]. Furthermore, some visualizations require a higher level of data literacy and attention to be aware that deceptive visualization is presented. These deceptive graphs are intentionally produced through the tricks of plotting charts by displaying insufficient and dubious data, charts that suggest misleading patterns, and confusing the uncertainties presented in the chart.

Audiences generally rely on chart creators to convey messages graphically from the data. However, audiences remain uncertain about whether the visualizations are fully representing the whole picture of the issue. Graph creators frequently offer inadequate information to substantiate their arguments [4, 7-8]. In the pursuit of a society with less misinformation, one of the efforts is for the public to be less influenced by misleading visualizations. To address this challenge, it becomes essential to examine the factors that affect public perception and evaluation of data visualization. The question that is raised, for example, is whether an increase in data literacy could lead to a public less swayed by deceptive visualizations. What types of visualization techniques tend to be more misleading? Will the audience's prior perception affect judgment when they receive new information?

This study aims to investigate the potential factors that affect the audience's perception and evaluation of data visualization. A group of undergraduate students from two different major programs participated in this experiment. These students were tasked with responding to predesigned questions based on two different sets of infographics addressing the same issues. The factors include the academic background of the audience which represents different levels of data literacy, the form of infographics and the order of data visualization which assesses whether prior perception will affect judgment. By understanding these aspects, authorities can gain better knowledge for implementing measures to detect and prevent problems in comprehending data visualizations.

2 Related Work

2.1 Misleading visualization

The study of misleading visualizations started long before the internet. "How to Lie with Statistics," written by Darrell Huff in 1954, stands as a pioneering work in the realm of misleading visualization and misapplication of statistics. From forming biases in sampling selection to choosing a suitable average that best supports one's argument, and even portraying graphs to deceive the audience, Huff laid the foundation for critical thinking about statistics. Recent research indicated that various visual tricks have been identified and classified in constructing misleading charts. Cowles et al. [9] identified four categories of visual health misinformation including misleading charts and visuals taken out of context. Doan [8] highlighted the techniques used in creating misleading visualizations about COVID-19. Lisnic et al. [4] pointed out the common reasoning errors of misleading visualizations in a large-scale data visualization about COVID-19 in Twitter posts.

To cultivate a better validity of graph visualization, books and publications about the common pitfalls of visualization and guidelines to construct high-integrity graphs have emerged. For instance, Kelleher et al. [10] proposed ten guidelines, including emphasizing selecting meaningful axis ranges when plotting effective visualizations. Nguyen et al. [5]

examined common visualization pitfalls in scientific publications. The frequently discussed misleading visualization techniques included manipulating scales, displaying insufficient data and unrelated causal inference [4, 6, 8].

Several studies have investigated the effects of visualization on audience interpretation, judgment, and decision-making. Woller-Carter et al. [11] proposed that misleading graphs could lead to judgment errors and hence affect decision-making. Axford et al. [12] suggested that employing more detailed data visualization is essential due to the risk of misinterpretation. Larkin [13] conducted an experiment designed to test whether the decisions of Air Force Institute of Technology students are affected by misleading charts. The author designed six charts, each with a high-integrity chart (control group) and a misleading chart (experiment group). Respondents were instructed to indicate whether they agreed with a proposed statement based on the given chart. The results showed all the answers to the six charts were significantly different, leading to the conclusion that misleading visualization techniques in charts could influence respondents' judgment.

2.2 Potential factors

Visual presentation significantly influences judgment and interpretation across various dimensions. Some fundamental components in misleading visualization include truncated axes, area encoding and 3D pie charts. Witt and Dhami [14] discovered that organizing icons proximally improved statistical reasoning by 35%–44%. Alhadad [15] advocated for the application of analytical visualizations in education to improve students' interpretation of relevant topics. However, some studies showed that the presentation of data does not impact the audience's perception. The audience might not be misled by simple charts that deliver direct messages, such as truncating the y-axis to exaggerate the trend [16].

The extent to which users are misled by the way data is portrayed could be caused by several influencing factors. The sequence of information acquisition plays an important role in shaping audience perception. Denrell [17] argued that negative initial impressions exhibit greater stability than positive impressions. Moreover, the reader's interpretation of graphs is subject to the influence of prior knowledge about a given subject [18]. However, Maltese et al. [19] and Driessen et al. [16] claimed that individuals who are more competent in the context of data literacy are less likely to be misled by deceptive data visualization. Maltese et al. [19] revealed that there is a notable difference in data literacy between experts and novice viewers in the science and mathematics fields. Driessen et al. [16] claimed that the contextual information in charts has a more substantial impact compared to an exaggerated axis and viewers' graph literacy. Lo et al. [20] discovered that the audiences can demonstrate enhanced skills in detecting misleading charts and displayed increased openness to suggest modifications in the visual design if they are allowed to reflect in the chart settings.

To the best of our knowledge, the primary focus of previous researchers on misleading visualizations has been individual charts rather than a collection of charts on a specific subject [8, 13]. Even when studies address charts on a general topic, the emphasis has predominantly been on categorizing the visual tricks employed in graph plotting [4, 8]. In this study, we direct our attention toward evaluating the conveyed message from a collection of charts presented in the form of infographics on a specific topic. Past experimental studies were limited to examining distinct data visualizations on separate individuals. In contrast, we are investigating the impact of various data visualizations on the same individuals using inferential tests, specifically pairwise *t*-test.

3 Method

3.1 Data and infographics

Two infographics depicting road accidents in Malaysia are created with open data sourced from the Polis Diraja Malaysia official website as illustrated in Figure 1. Infographic A is intentionally designed to portray a less severe road condition in Malaysia while Infographic B is crafted to highlight a more serious traffic situation. The methodologies employed in constructing these two infographics are derived from recommendations provided by prior researchers and authors of scholarly works on misleading visualization in the previous section.

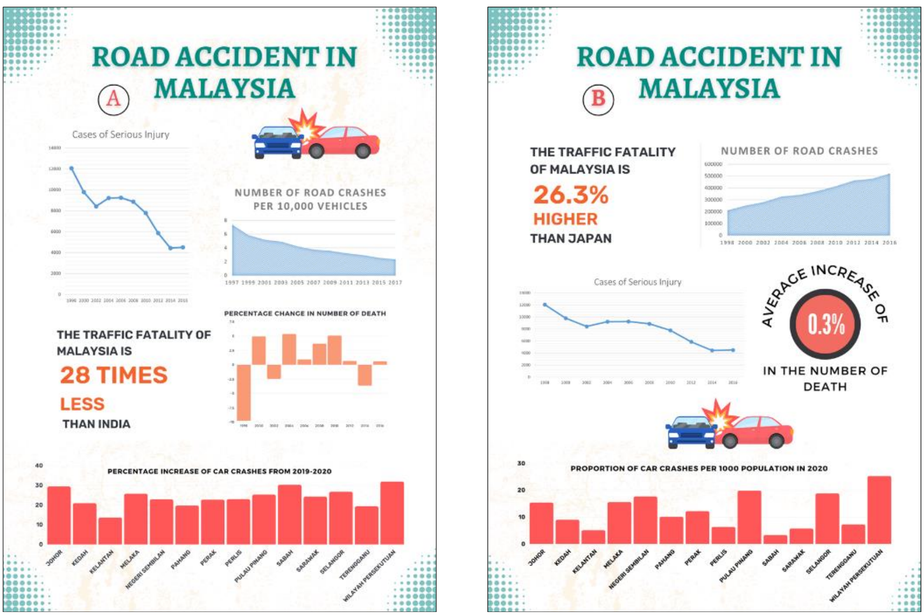

Fig. 1. Infographic A (left) and Infographic B (right)

Firstly, manipulation is made to the y-axis scale in the charts [4, 6-8]. Infographic A showcased an enlarged y-axis concerning cases of serious injuries, enlarging the decrease in such cases. Conversely, Infographic B compressed the y-axis, downplaying the trend and diminishing the perceived seriousness of serious injury cases. Secondly, the infographic is designed with intentional omission of information [4, 7-8]. The information about the population change is not disclosed in Infographic B. Respondents rely solely on the number of cases to judge the seriousness of road accidents in Malaysia.

Furthermore, both infographics introduced an arbitrary threshold. A bar chart of the percentage change in the number of deaths over time is revealed in Infographic A. Nonetheless, Infographic B, which attempted to deliver a more serious message, solely displayed the statement ‘Average Increase of 0.3% in the Number of Deaths’. This selective presentation could potentially mislead individuals into believing a continuous increase in the number of deaths. Both visualizations showed changes in death cases in the same time frame, spanning from the year of 1998 to 2016.

Moreover, the infographic consisted of irrelevant information [7]. The traffic fatality of Malaysia in 2020 is compared with the traffic fatality of other countries, although road conditions of other countries are irrelevant to Malaysia's traffic condition. In Infographic A, since Indonesia had a high traffic fatality rate, partly due to its large population, comparing Malaysia's traffic fatality to the country shows a less severe traffic fatality in Malaysia.

3.2 Experimental design

The respondents are selected from the undergraduates enrolled in two distinct academic programs, one enrolled in statistical programme and the other in non-statistical programme. The selection process involved employing cluster sampling for students in the non-statistical program. This group serves as a control to examine the impact of academic background. A total of 60 undergraduates participated in this experiment. Twenty students enrolled in statistical programme while 40 students are from the academic backgrounds unrelated to statistics. They are tasked to rate the seriousness and trend of road accidents in Malaysia after viewing the two infographics in different sequences. Thirty students are randomly assigned to answer questions based on Infographic A, followed by Infographic B (Sequence 1: A → B). Conversely, the other thirty students are tasked with responding to questions based on Infographic B first followed by Infographic A (Sequence 2: B → A). Each group is allocated a total of 10 minutes to complete the respective questionnaire.

The first question inquired about the participants' assessment of the severity of road accidents in Malaysia. The seriousness scale ranged from 1 (not serious) to 5 (very serious). The second question addressed the perceived trend of road accidents in Malaysia, with participants providing a rating from 1 (a better-off trend) to 5 (a worsening trend). Here, a score of 1 indicated an improvement in traffic conditions, while a score of 5 indicated a worsening trend of road accidents. Four questions are designed to assess the participant's comprehension of the graphs and to filter out respondents who might not take the task seriously. All the respondents met the visualization literacy benchmarks. Hence no one is excluded in the study.

4 Findings

The overall results indicate a significant difference in the measures of seriousness and trend for both infographics, as presented in Table 1. This suggests that the audience responses align with the infographic design intent. Specifically, Infographic A is intentionally crafted to portray a less severe road accident scenario. Conversely, Infographic B is designed to depict a more severe road accident situation and the audience responses are consistent with this portrayal. It is concluded that varying data visualizations significantly impact the perception and evaluation of the seriousness of road accidents, even when the same dataset is used in both infographics.

Table 1. Mean Score of Seriousness and Trend

		Infographic A	Infographic B
Seriousness**	Mean	2.800	3.867
	(Std. Error)	(0.112)	(0.108)
Trend **	Mean	1.983	3.917
	(Std. Error)	(0.100)	(0.115)

Note: ** denotes significance at 5%

4.1 Academic background

In Table 2, when examining the mean difference between both infographics by academic background, the score difference between statistical students and non-statistical students is 1.42 and 0.97 respectively. A higher difference in score indicated that the statistical students could better receive the designated message as infographic A is intentionally crafted to portray a less severe road accident scenario.

When comparing by infographic, it is noted that there is nearly no score difference in Infographic B for both groups of students. For infographic A, it is observed that the mean score of statistical students is lower in infographic A, aligning with the intended message. Consequently, non-statistics students exhibit a lesser ability to discern the conveyed message by Infographic A, which is meant to portray a not serious traffic condition. We posit academic background served as a proxy for proficiency in data literacy. It is evident that the level of data literacy plays a crucial role in decoding the message conveyed through data visualization.

Table 2. Mean Score of Seriousness by Academic Background

			Statistical Programme	Non-statistical Programme
Seriousness	Infographic A **	Mean	2.474	3.000
		(Std. Error)	(0.173)	(0.131)
	Infographic B	Mean	3.895	3.974
		(Std. Error)	(0.165)	(0.112)

Note: ** denotes significance at 5%

4.2 Sequence

The interpretation of content by a reader can be shaped by their prior perception of a specific subject. The results in Table 3 clearly show that the mean scores of both infographics differ across different sequences. The students who were initially exposed to a less severe visualization of road accidents tended to amplify their negative perception compared to those who directly viewed Infographic B which demonstrates a more severe condition. Conversely, students who initially viewed a negative visualization are more likely to rate the road accidents in Infographic B as less severe. This highlights a disparity between positive and negative initial impressions with negative impressions proving to be more stable than positive impressions [17]. Consequently, the impact of viewing a more severe road accident infographic (Infographic B) first is more pronounced than viewing Infographic A. This simply means if the audience had previously seen a data visualization with an opposing impression, they tended to exaggerate the impression of road accidents in the opposite way when another infographic was used.

Table 3. Mean Score of Seriousness by Sequence

			Sequence 1	Sequence 2
Seriousness	Infographic A*	Mean	3.000	2.621
		(Std. Error)	(0.169)	(0.142)
	Infographic B**	Mean	4.138	3.552
		(Std. Error)	(0.160)	(0.125)

Note: * and ** denote significance at 10% and 5% respectively

5 Conclusion

This study aims to examine the impact of presenting various data visualizations in the form of infographics with different sequences on the judgments and perceptions of audiences with diverse backgrounds regarding the seriousness and trend of road accidents. The results reveal distinct judgments and perceptions among audiences exposed to different infographics. A significant difference emerges in the scores related to the seriousness of Infographic A which is designed to convey a less severe traffic condition in different academic backgrounds. Proficiency in data literacy is indeed crucial for decoding the message within data visualization. The viewing sequence of the infographic impacts the seriousness scores which amplifies the latter's interpretation, particularly when a more severe infographic is viewed first.

To ensure better clarity in delivering the message of charts, the creators of data visualization should follow the guidelines of designing data visualization and be aware of the design that would lead to misinformation. Individuals responsible for communicating messages through visualization, particularly visual journalists, or evidence-driven visual communicators, are urged to follow ethical principles in the creation of charts and infographics to maintain clarity [6]. On the other hand, the public is encouraged to cultivate a sufficient level of data visualization literacy to safeguard themselves from potential misinterpretations arising from data visualization. A limitation of this study is the restriction of respondents to higher tertiary education students. Future research efforts could include participants from the public with diverse backgrounds to gain deeper insights into the effects of various data visualizations.

References

1. S. Few, P. Edge, Data Visualization: Past, Present, Future (IBM Cognos Innovation Center, 2007)
2. M. Islam, S. Jin, *An overview of data visualization*, InProceedingsof the International Conference on Information Science and Communications Technologies, ICISCT, 4-6 Nov 2019, Tashkent, Uzbekistan (2019)
3. M. Sadiku, A.E. Shadare, S.M. Musa, C. M. Akujuobi, IJERAT, **2**, 11 (2016)
4. M. Lisnic, C. Polychronis, A. Lex, M. Kogan, *Misleading beyond visual tricks: How people actually lie with charts*, InProceedingsof the 2023 CHI Conference on Human Factors in Computing Systems, 20-23 April 2023, Hamburg, Germany (2023)
5. V.T. Nguyen, K. Jung, V. Gupta, Vis. Comput. Ind. Biomed. Art, **4**, 27 (2021)

6. A. Cairo, How charts lie: Getting smarter about visual information (WW Norton & Company, New York, 2019)
7. D. Huff, How to lie with statistics (Penguin UK, London, 1954)
8. S. Doan, J. Bus. Tech. Commun, **35***,* 73 (2021)
9. K. Cowles, R. Miller, R. Suppok. Med. Ref. Serv Q, **43**, 44 (2024)
10. C. Kelleher, T. Wagener. Environ Model Softw, **26**, 822 (2011)
11. M.M. Woller-Carter, Y. Okan, E.T. Cokely, R. Garcia-Retamero, *Communicating and distorting risks with graphs: An eye-tracking study.* InProceedingsof the human factors and ergonomics society annual meeting, 22-26 Oct 2012, Westin Boston Waterfront, USA (2012)
12. N. Axford, M. Beveridge, M. Ahern, F. Cortes, J. Martinez, A. Thompson. Res Sq, (2023)
13. A.A. Larkin, Misleading graphic: can decision makers be affected by their use. (Defense Technical Information Center, 1990)
14. J.K. Witt, M.K. Dhami, J. Appl. Res. Mem. Cogn, **12**, 241 (2023)
15. S.S. Alhadad, *Attentional and cognitive processing of analytics visualisations: Can design features affect interpretations and decisions about learning and teaching.* InProceedingsof the 33rd International Conference of Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education, ASCILITE, 28-30 Nov 2016, Adelaide, Australia (2016)
16. J.E. Driessen, D.A. Vos, I. Smeets, C.J. Albers, PLOS ONE, **17** (2022)
17. J. Denrell, Psychol Rev, **112**, 951 (2005)
18. P. Shah, Diagrammatic representation and reasoning (SpringerLink, London, 2022)
19. A.V. Maltese, J.A. Harsh, D. Svetina, J. Coll. Sci. Teach., **45**, 84 (2015)
20. L.Y-H. Lo, Y. Cao, L. Yang, H. Qu. IEEE Trans. Vis. Comput., **30**, 955 (2024)