

# Strategies for evaluating visual analytics systems: A systematic review and new perspectives

Md Rafiqul Islam<sup>1</sup> , Shanjita Akter<sup>2</sup>, Linta Islam<sup>3</sup>,  
Imran Razzak<sup>4</sup>, Xianzhi Wang<sup>5</sup> and Guandong Xu<sup>1</sup>

Information Visualization  
2024, Vol. 23(1) 84–101  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/14738716231212568  
journals.sagepub.com/home/ivi



## Abstract

In recent times, visual analytics systems (VAS) have been used to solve various complex issues in diverse application domains. Nonetheless, an inherent drawback arises from the insufficient evaluation of VAS, resulting in occasional inaccuracies when it comes to analytical reasoning, information synthesis, and deriving insights from vast, ever-changing, ambiguous, and frequently contradictory data. Hence, the significance of implementing an appropriate evaluation methodology cannot be overstated, as it plays a pivotal role in enhancing the design and development of visualization systems. This paper assesses visualization systems by providing a systematic exploration of various evaluation strategies (ES). While several existing studies have examined some ES, the extent of comprehensive and systematic review for visualization research remains limited. In this work, we introduce seven state-of-the-art and widely recognized ES namely (1) dashboard comparison; (2) insight-based evaluation; (3) log data analysis; (4) Likert scales; (5) qualitative and quantitative analysis; (6) Nielsen's heuristics; and (7) eye trackers. Moreover, it delves into their historical context and explores numerous applications where these ES have been employed, shedding light on the associated evaluation practices. Through our comprehensive review, we overview and analyze the predominant evaluation goals within the visualization community, elucidating their evolution and the inherent contrasts. Additionally, we identify the open challenges that arise with the emergence of new ES, while also highlighting the key themes gleaned from the existing literature that hold potential for further exploration in future studies.

## Keywords

Evaluation strategies, visual analytics, visual analytics systems, visualization research, interactive system

## Introduction

Visual analytics systems (VAS) are a relatively new field that is effective and intuitive for data analysis or data mining. Nowadays, it is used to discover problems and obtain insights from various researchers. Existing studies have highlighted the significance of justifying both its usefulness and usability aspects. To address this, we conduct a comprehensive review of over 100 papers published in top-tier journals and conferences over the past 6 years, focusing specifically on their utilization of evaluation methods. Our main objective is to gain a comprehensive understanding of the diverse evaluation techniques employed in visualization research as a whole.

<sup>1</sup>Data Science Institute, University of Technology Sydney, Broadway, NSW, Australia

<sup>2</sup>School of Computer Science, Taylors University, Subang Jaya, Selangor, Malaysia

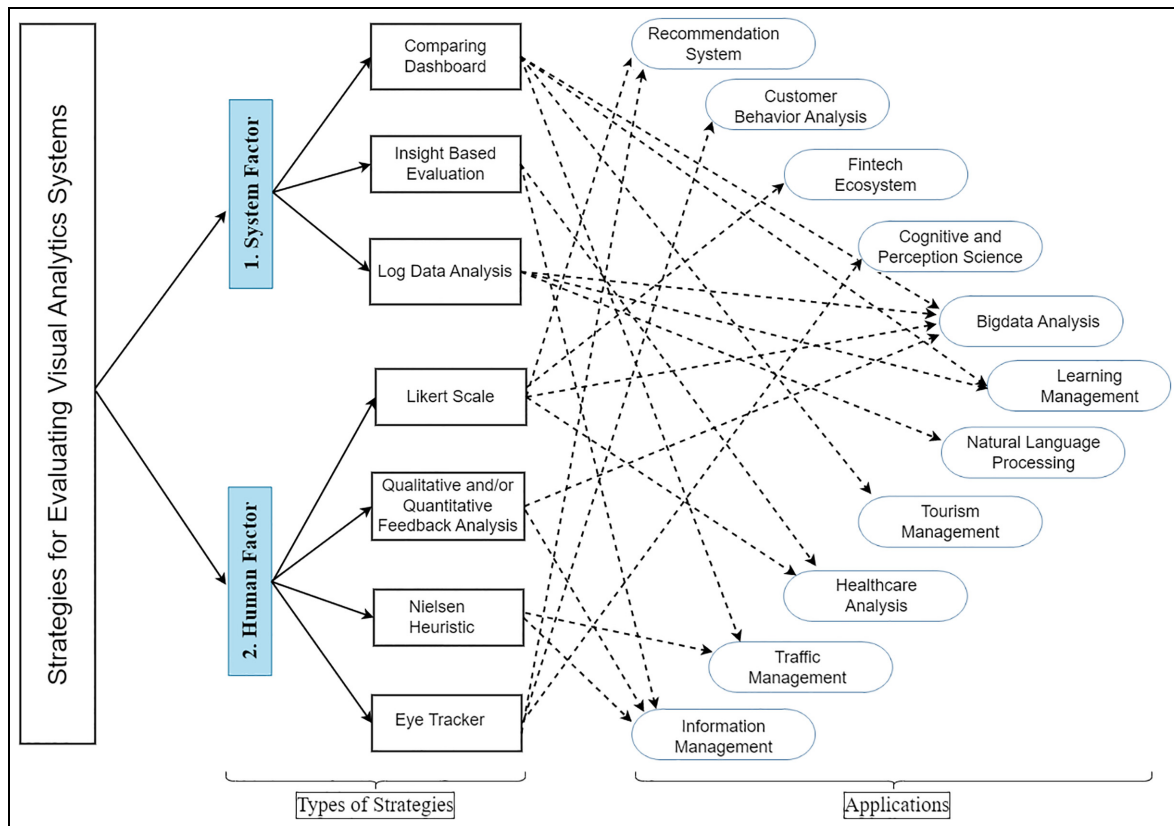
<sup>3</sup>Department of Biological Science, Louisiana State University, Baton Rouge, LA, USA

<sup>4</sup>School of Computer Science and Engineering, UNSW, Sydney, NSW, Australia

<sup>5</sup>School of Computer Science, University of Technology Sydney, Ultimo, NSW, Australia

### Corresponding author:

Md Rafiqul Islam, Data Science Institute, University of Technology Sydney, Ultimo, Broadway, NSW 2007, Australia.  
Email: mdrafiqul.islam@uts.edu.au



**Figure 1.** Evaluation strategies for various applications.

**Strategies for evaluating visual analytics systems:** Evaluation offers invaluable insights into a visualization system, greatly enhancing the effectiveness of visual interactive systems (VIS) in a more profound manner. The growing volume of literature on how to undertake evaluation and the increasing number of research publications that include a formal or informal assessment serve as evidence of the significance of evaluating visualization systems. For instance, Mandal et al.<sup>1</sup> proposed a novel VIS for discovering knowledge and hidden opportunities from massive and complex data. By incorporating more intelligence into the analysis process and dynamic volumes of information through visual representations and interaction approaches, the model automatically learns to close the information gap. This article contributes to the existing body of work by examining the comprehensive evaluation procedures employed in the peer-reviewed visualization papers that have not been previously subjected to systematic assessments. Figure 1 provides the primary evaluation methods applied to visual analytics (VA) research in various applications. For example, eye trackers are used in multiple areas, such as recommendation systems Saraiya et al.<sup>2</sup>

**What are the differences between this survey and former ones?** Numerous prior research works

have explored the differences between VA and information visualization (InfoVis) from various perspectives. These investigations encompass data analysis, perception, cognition, and human-computer interaction (HCI).<sup>3,4</sup> Additionally, the existing surveys primarily focus on (1) evaluating a visualization system for log data<sup>5</sup>; (2) assessing the impact of user characteristics and different layouts on interactive visualizations tailored for decision making,<sup>6-8</sup> (3) exploring visualizations aimed at enhancing the understanding of machine learning (ML) models<sup>9</sup>; (4) investigating the latest developments in predictive VA Lipton<sup>10</sup>; interactive ML Plaisant et al.<sup>11</sup>; interpretable ML Lipton<sup>10</sup>; and surveys of multidimensional visualization techniques.<sup>12,13</sup> However, only a few studies touch on our area of interest. Hence, this paper focuses on evaluation techniques categorized as human factors and system factors, which have been extensively utilized in diverse visualization applications.<sup>14-17</sup> Additionally, we specifically emphasize how these existing assessment approaches contribute to addressing the aforementioned issues, rather than providing a broader perspective on their application.

**Contribution of this survey:** In this paper, we present a state-of-the-art review that goes beyond more exploration of evaluation strategies. It delves into

the most significant evaluation techniques and uncovers novel findings by critically examining their application in various domains. In addition, this study highlights the key challenges faced in existing evaluation research, particularly in the field of VIS, and proposes multiple research directions to address these challenges. By assessing the advantages and disadvantages of different evaluation procedures, this paper provides valuable insights to enhance the evaluation process. In summary, our primary contributions are as follows:

- First, we present an in-depth overview of VIS, identify the novel evaluation methods, and explore how these methods play a significant role in visualization research.
- Second, we describe the popular state-of-the-art evaluation methods in the visualization community and offer a historical perspective by exploring numerous applications with these evaluation practices.
- Third, we assist researchers to identify, justify and refine their evaluation approaches. Additionally, we apply the lessons learnt from previous research and detail several opportunities, challenges, and future directions to help researchers recognize and avoid the known pitfalls.

The paper is organized as follows: first, a thorough study of VAS is provided in the background section, followed by a detailed discussion of the article collection and analysis in the methodology section. Then the next section overviews the evaluation methods and their terminology. The opportunities, difficulties, and future directions are discussed in the discussion section. Finally, the paper is concluded in the last section.

## Background study

### *What is a visual analytics system?*

The study of analytical reasoning made possible by interactive visual interfaces is known as VA Cook and Thomas.<sup>18</sup> It is an analytical technique that enables the combination of knowledge and facilitates understanding from vast, dynamic, confusing, and often contradictory datasets, allowing for the identification of both anticipated and unexpected information. When employing human judgment to draw conclusions from data, this approach relies on a comprehensive understanding of the underlying cognitive and perceptual concepts and the reasoning process Stadler et al.<sup>19</sup>

VAS assists users by presenting data effectively. It involves taking raw data as input, modeling it, and delivering the output in a comprehensible format Kosara.<sup>20</sup> Interactive visualization plays a vital role in this process as it allows users to easily identify patterns and trends when presented with a visual overview of the data, as opposed to analyzing hundreds of rows in a spreadsheet. This approach aligns with how the human brain processes information. Visualizing data enhances its value as data analysis aims to extract insights. Without the utilization of VA, data analysts may be able to reach conclusions based on their analysis, but effectively communicating those findings can prove challenging Liu et al.<sup>21</sup>

### *VAS in various application domains*

As shown in Figure 1, several interactive VAS have been applied in various application domains globally, where researchers have proposed and analyzed new techniques and discussed in detail how specific strategies might be applied to evaluate the system.<sup>22–30</sup> These researchers widened the scope of various applications by developing a framework that encourages the adoption of modernized design techniques in the visualization sector. Additionally, they also focused on establishing methodological insights while providing insight into specific areas of VAS evaluation. According to the existing studies, the key application domains where various interactive VAS are applied are (i) big data analysis, (ii) cognitive and perception science, (iii) recommendation systems, (iv) healthcare analysis, (v) customer behavior analysis, (iv) natural language processing, (vii) tourism management, and (viii) fintech ecosystem. For example, Liu et al.<sup>31</sup> designed a new visualization tool using a virtual reality platform. Green et al.<sup>32</sup> introduced perception and cognition-based visualization systems. They designed their system to maximize the cognitive strength of both humans and computers. A hybrid interactive visualized recommender system is presented by Bostandjiev et al.<sup>33</sup> to recommend items from social networking sites like Facebook, Twitter, etc. The system accepts user preferences as input and recommends an item on the interactive interface. Most of these studies applied various evaluation methods to justify their system. Therefore, we aim to provide a systematic review of the novel methods which have been used to evaluate interactive VAS.

### *Role of evaluation strategies in visualization research*

In recent times, interactive VAS have gained widespread usage as integral components of the creative process,

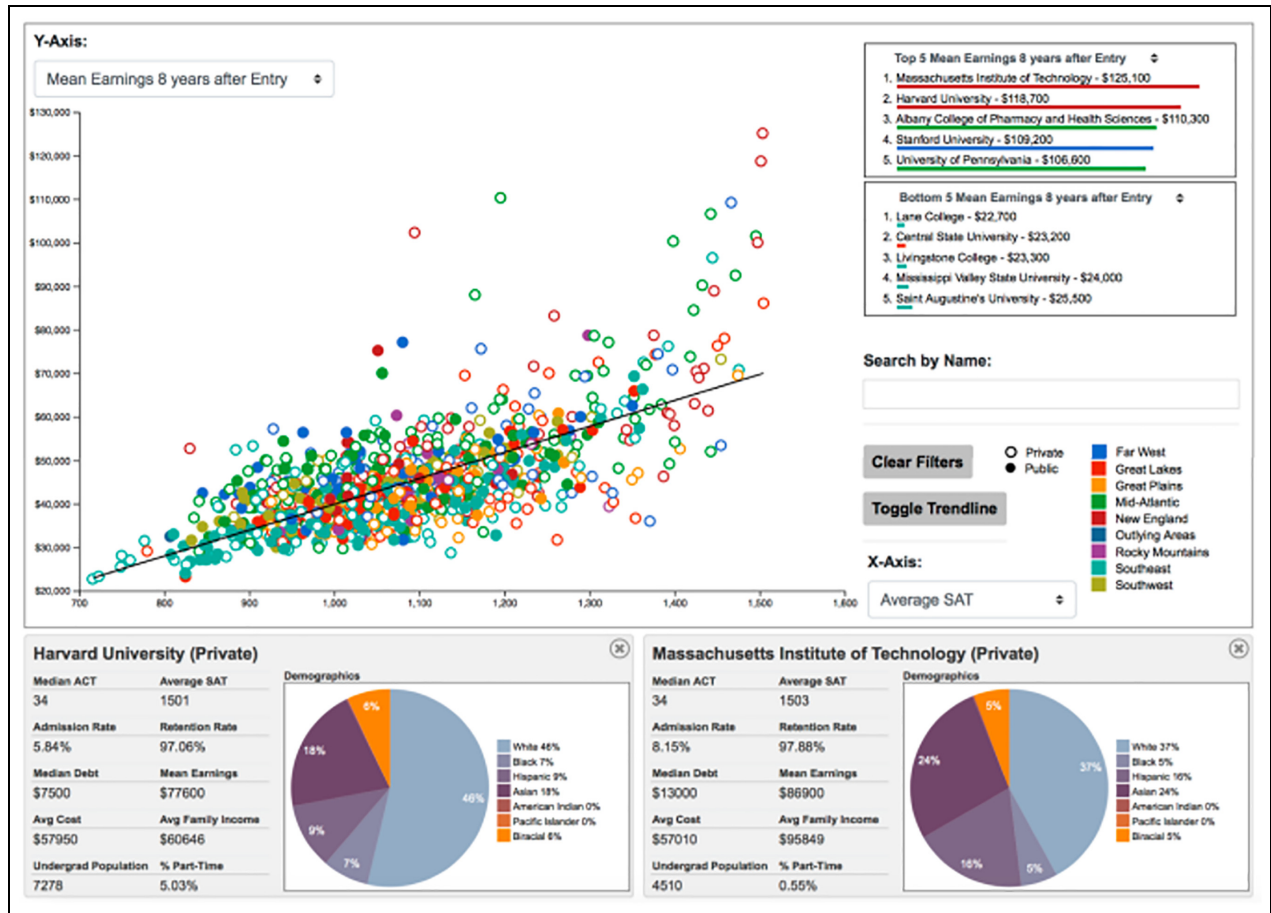


Figure 2. An interactive deep visual analytics system (Figure courtesy of Wang et al.<sup>36</sup>).

empowering users to formulate hypotheses, uncover patterns and anomalies, and refine theories. In the realm of visualization research, evaluation strategies assume a crucial role, acting as foundational elements that guarantee the effectiveness, usability, and practicality of visualizations. These strategies are designed to evaluate and validate the advantages and disadvantages of visual representations and interaction techniques. For example, an interactive VAS is developed to identify, isolate, and present the information to analyze the immense volumes of complex data as shown in Figure 2. Using this system, individuals gain the ability to interact with data, enabling the visualization of potential insights in a manner that was previously unattainable with static graphs. However, as the data is large and complex, it isn't always easy to examine in a prompt and efficient manner as it consumes significant time and money. Although several existing evaluation studies and experiments are helpful, there is an increasing demand for alternative evaluation methods that provide measurable advantages that promote the comprehensive

adoption of interactive VAS Christmann et al.<sup>24</sup> Therefore, evaluation strategies' play an increasingly important role in many domains because users may obtain unexpected findings that challenge their preconceived opinions, prompt new ideas, and lead to significant breakthroughs.<sup>34,35</sup>

Interactive VAS is used to gain insight into vast volumes of abstract data, such as tables, hierarchies, and networks. These visualizations extend beyond conventional reports and records and prove valuable in various domains, including big data analysis Keim et al.,<sup>37</sup> health care analysis Stadler et al.,<sup>19</sup> customer behavior analysis Abdulla et al.,<sup>38</sup> natural language processing Tupikovskaja-Omovic and Tyler,<sup>39</sup> recommendation systems Varu et al.,<sup>14</sup> fintech ecosystems Basole and Patel,<sup>40</sup> and tourism management systems Almaimoni et al.<sup>41</sup> The goal is to recognize patterns, structures, and anomalies, allowing specialists to evaluate large amounts of data. Geographical information systems can also benefit from a visualization system Folorunso and Ogunseye.<sup>42</sup> Regional planning, transportation planning

and management, weather forecasting, and mapping rely on VAS such as MapQuest and SmartForest Geisler.<sup>43</sup> Furthermore, visualization plays a crucial role in contemporary architectural and medical applications. Notably, the National Library of Medicine's Visible Human Project Ackerman<sup>44</sup> provides an extensive digital library of anatomical images capturing the human body with exceptional precision, taken at 1 mm intervals from both male and female cadavers. These images are invaluable resources for health professionals, enabling the identification of diverse diseases and ailments, ranging from damaged ligaments to severe conditions leading to fatalities.

To ensure VAS quality and to compare the results, we need to evaluate the systems. As shown in Table 1, several evaluation methods are used to assess VAS, including Likert scales, log data analysis, eye trackers, dashboard comparison, etc. Many of these techniques have been employed to acquire research findings in the social and behavioral sciences. In the following section, we provide a brief overview and discussion of the benefits, limitations, and applications of evaluation methods predominantly employed in current visualization systems. The overall impact of a visualization system hinges on its advanced visualization and interaction capabilities. However, several research considerations arise, such as determining the evaluation scope, selecting the appropriate number and type of users, defining relevant tasks, choosing datasets, collecting relevant data, and conducting appropriate data analysis. Consequently, the research indicates that VAS are not

only effective and impactful they also attract significant attention to verify the performance of the models.

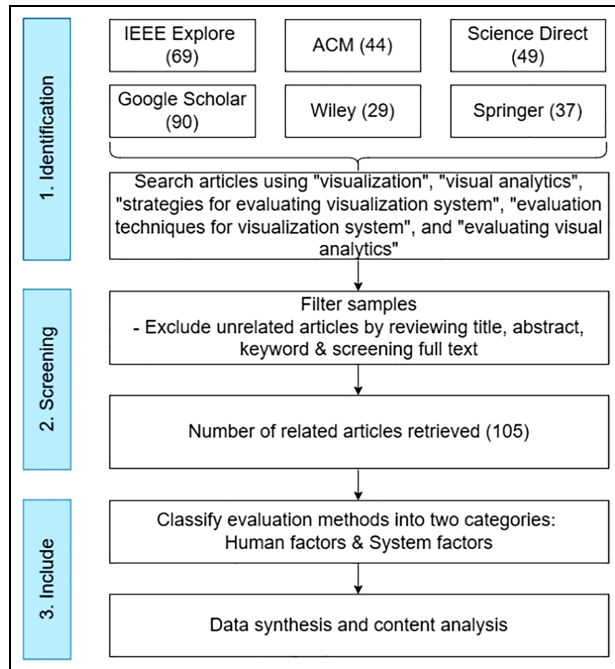
## Study methodology

A systematic approach to performing a literature review can reveal themes and generalizable insights that aid in the better construction of an overview of the body of knowledge on a particular topic. For instance, valuable methods proposed for conducting a systematic literature review (SLR) can propel advancements in the field of information science (IS) through both research and practical applications Yasmin et al.<sup>51</sup> SLRs can motivate the development of evidence-based guidelines for practitioners and related investigations. The process involves three key rounds: formulating the research topic, locating and evaluating relevant studies, and finally, conducting a comprehensive document review where the data are consolidated. By applying the SLR technique, a deeper understanding of the current strategies employed in reviewing VAS research can be achieved. Figure 3 (study methodology) illustrates the sample collection and analysis process.

**Research questions:** As detailed in many existing studies,<sup>51,52</sup> an SLR should follow clearly defined research questions where the criteria are clearly stated before the review is conducted. These research questions are designed to assist in the selection of relevant, thorough, and comprehensive studies. After identifying the research questions an SLR should assist the researcher to identify, select and pursue appropriate

**Table 1.** Existing evaluation methods.

| Paper                             | Evaluation methods       | Purpose of the analysis   | Applications               |
|-----------------------------------|--------------------------|---|----------------------------|
| Saraiya et al. <sup>45</sup>      | Insight-based Evaluation | Conduct a longitudinal study to examine how data visualizations are used to provide insight.  | Visual Interactive System  |
| Dowding and Merrill <sup>46</sup> | Nielsen's Heuristics     | Create a heuristic evaluation checklist that may be used to assess InfoVis production.  | Information Visualization  |
| Gena <sup>47</sup>                | Qualitative Analysis     | Comprehensive overview of methods and techniques used for the evaluation.   | Human-Computer Interaction |
| Kurzahls et al. <sup>48</sup>     | Eye Tracking             | Analyzed existing research works and present an overview of how eye tracking is currently utilized to evaluate visualization techniques, as well as how the gaze data is evaluated. | Visual Interactive System  |
| Onoue et al. <sup>49</sup>        | Qualitative Analysis     | Proposed a visual analytics systems for evaluation structures, to enable such effective analysis.   | Visual Analytic System     |
| Deriu et al. <sup>50</sup>        | Dashboard comparison     | Conducted a survey on human factor and automated evaluation methods for dialog systems.   | Dialog System              |



**Figure 3.** Review methodology for article collection and analysis.

research directions. Therefore, our current study aims to answer the following research questions, and to address each one with appropriate solutions.

*RQ1: What are the existing evaluation methods used in visualization research?*

VA is an emerging research topic that deals with a wide range of research globally. Although the primary goal of this study is to describe several strategies for evaluating VAS, it is difficult to cover all the evaluation strategies to monitor, interpret, and extract useful information. Therefore, based on the peer-reviewed articles, we determine the relevance of seven evaluation methods employed in visualization research namely: (1) dashboard comparison; (2) insight-based evaluation; (3) log data analysis; (4) Likert scales; (5) qualitative and quantitative analysis; (6) Nielsen's heuristics; (7) eye trackers.

*RQ2: Are the existing evaluation strategies valid in theory, understanding, and knowledge for evaluating visual analytics systems?*

It is crucial to investigate the justifications and the specifics of the methods used when evaluating past studies to assess VAS to determine whether it can produce impartial and fair judgments when using human-

centered methods.<sup>53,54</sup> While reviewing descriptions of different aspects of these strategies, it is crucial to identify the types of strategies used, namely (1) system factors and (2) human factors, and to comprehend the underlying motivations behind their selection. It is important to analyze how these evaluation strategies were implemented, the input characteristics involved, the training parameters, potential biases, and any notable features Song et al.<sup>16</sup> More importantly, depending on the task which is being undertaken by the VAS especially risk detection. For instance, theoretical frameworks generated from and validated by the health sciences are used to support the human-factor-based evaluation technique that was first introduced by Wang et al.<sup>36</sup>

*RQ3: How effective are the strategies in terms of both computational performance and meeting the user's needs?*

As shown in Table 4, several strategies have been proposed to evaluate the performance of interactive VAS, such as Likert scales, eye tracker, log data analysis, model comparison, and so on. According to the existing studies, these strategies provide valuable insights for testing VAS in many application domains from the users' perspective as well as from a computational perspective.

*RQ4: What system artifacts have been created, and how did they perform when used in practical situations?*

Developing an interactive VAS with unique evaluation artifacts is considered a primary research focus within the visualization community.<sup>55,56</sup> Using these design artifacts, we can truly determine the effectiveness of VAS. Thus, it is crucial that researchers not only create and assess VAS but also to integrate them into applications so that humans can evaluate them in real-world settings Kamaleswaran et al.<sup>57</sup> For example, in risk detection, it is essential to know how the model is integrated into a system.<sup>9</sup>

**Literature resources:** To answer the aforementioned research questions, we initially collected 318 sample articles. Through a meticulous examination of their titles, keywords, and adherence to inclusion and exclusion criteria, we narrowed the selection to 105 articles on which to perform the SLR. The SLR encompassed an in-depth analysis of 105 peer-reviewed papers published between 2016 and 2022, focusing on the evaluation of VAS. Figure 3 illustrates the entire process of the peer-reviewed article collection and review, where the proposed approach



for collecting sample articles takes a multi-disciplinary perspective.

**Searching keywords:** As indicated in Figure 3, the search was performed across six electronic databases and identified relevant articles between 2016 and 2022. We searched these databases using the keywords “visualization,” “visual analytics,” “strategies for evaluating visualization system,” “evaluation techniques for visualization system,” and “evaluating visual analytics” to capture the most relevant articles. However, we didn’t include non-English language publications, book chapters, newspaper items, unpublished articles, or articles that needed to be more scientific.

**Study selection.** First, we evaluated the titles of each article and then applied the previously established inclusion and exclusion criteria, obtaining a selection of possibly relevant articles. Following this, the team members manually reviewed each article and obtained the whole text before critically evaluating its contents. Duplicated research articles and those that may not address all of the research questions were omitted. After filtering, we retrieved 318 freely available full articles from six databases as shown in Figure 3. From this, 60 duplicates were eliminated. Additionally, 40 articles were discarded as newspaper or unreviewed internet content, while another 69 articles were eliminated for focusing on visuals rather than VA and lacking target information. Finally, 44 articles were removed after human inspection of the title, keyword, abstract, and full text revealed unrelated information. After classifying the articles into four groups, we verified that 105 addressed our study question and underwent the review analysis.

**Information synthesis and review analysis.** Content analysis is the practice of directly examining the content of any human interaction process, including verbal, visual, and written materials while analyzing the data using qualitative and quantitative research methods. Additionally, qualitative content analysis condenses the original data. Although both deductive and inductive approaches are widely employed, inductive analysis is the most useful when there have been no previous studies on the phenomenon or when the phenomenon needs to be more cohesive. This work uses an inductive approach to classify and categorize attributes from a few solution application design papers. We searched for each item thoroughly, describing procedures, evaluation methodologies, and design processes.

## Overview of evaluation methods and terminology: State-of-the-art

As indicated in Figure 3, a keyword search was conducted across six electronic databases to locate relevant papers from diverse publishers between 2016 and 2021. We identified the related studies by utilizing specific keywords to obtain the most relevant articles and to satisfy the PRISMA criteria Page et al.<sup>52</sup> In this section, we examine seven evaluation methods specifically within the context of VA. Our analysis involves reviewing over a 100 articles, with a primary focus on their application in VAS. These evaluation methods have been widely employed in assessing various aspects of VAS, as summarized in Table 2. Notably, recent research trends indicate an increased adoption of these evaluation methods. Consequently, we categorize these methods into two groups: system factors (SF) and human factors (HF), as illustrated in Figure 1. Table 3 lists the selected research studies and their advantages, and disadvantages.

### System factors

System factors play a vital role in every stage of the visualization process, from designing to evaluating visualization tools. These factors focus on the technical aspects and characteristics of the system itself and encompass various technical aspects that directly impact the effectiveness and usability of visualization systems. By considering system factors, designers and evaluators can ensure the development of robust and efficient visualization tools.

**Dashboard comparison.** Conducting a dashboard comparison is a valuable approach to evaluating visualization systems. This evaluation method has been proved to be productive, efficient, and provides substantial insights into the effectiveness of visualizations. By comparing different dashboards, researchers can assess internal structures, functionalities, system behavior, accuracy, efficiency, and interactivity. The dashboard comparison evaluation technique is particularly effective when the analyzed parameters are closely aligned.

Several researchers have embraced evaluation techniques to assess their visualization systems. One noteworthy example is the study conducted by Keim and Kriegel<sup>93</sup> in the early stages of visualization development. In their research, they compared their visualization system with alternative techniques to validate its effectiveness in analyzing multidimensional datasets. To gauge the model’s efficiency, they meticulously





**Table 3.** Overview of key representative works involving human and system factors.

| Types of methods | References                    | Evaluation Technique     | Advantages   | Disadvantages   |
|------------------|-------------------------------|--------------------------|--|---|
| Human Factors    | Qian et al. <sup>64</sup>     | Likert Scale             | 7 point Likert scale accurately identified the highest scoring ML-based visualization recommender system according to the human expert's ratings | It failed to measure the actual attitudes of respondents as only a few are rated efficiently rather than all users. |
|                  | Shao et al. <sup>85</sup>     | Eye Tracker              | Eye trackers have been used to collect actual visual data on the entire screen space and detect off-screen times.                                | Failed to identify a few users who wear contact lenses or have long eyelashes.                                      |
|                  | Ali et al. <sup>89</sup>      | Qualitative Analysis     | Conducted qualitative analysis to evaluate various learning analytic tools and identified the best one   | Generated some redundant data which mislead the final output.   |
|                  | Zöller et al. <sup>90</sup>   | Nielsen's Heuristics     | To ensure the visual variable is of sufficient length  | Considered several limitations such as redundancy, heuristics conflict, generalizable problems, etc.                |
| System Factors   | Chang et al. <sup>91</sup>    | Insight Based Evaluation | Used to explore the user's insight and capture actual data   | Expert's review can be repeatable while evaluating insights.  |
|                  | Lerche and Kiel <sup>92</sup> | Log Data Analysis        | Analyzing log data allows for a more accurate and effective evaluation outcome   | Difficulty in handling massive amounts of data.   |
|                  | Zuk et al. <sup>60</sup>      | Dashboard Comparison     | Easier to visualize the comparisons so it can quickly and effectively identify which one achieves the better performance                         | Lack of similar parameters.   |

evaluated multiple test cases, revealing significant disparities and providing valuable insights. Moreover, Bourqui et al.<sup>71</sup> explored the interrelation between human and computer vision based on deep learning techniques and evaluated the graph visualization model. They compared their performance with user evaluations and other existing methods to enhance their model's accuracy and execution. Stehle and Kitchin<sup>65</sup> evaluated a real-time archived data visualization technique, and proposed a city dashboard visualization system, and effective strategies for improving the dashboard design. They asserted that their visualization technique is significantly more effective than the existing ones. designed a visualization system ExtraVis, to evaluate and overcome road traffic incidents and assist in traffic management system control. They compared their approach to three incident dashboards and explored the practical benefits and techniques that are not available in the existing systems. As a result, the technique of dashboard comparison, particularly in visualization, has a wide range of applications since it is thorough and makes it easier to discover the scope of existing systems and overcome their limitations.

*Insight-based evaluation:* Insight has been commonly stated as the broader purpose of InfoVis, and it seems to capture the intuitive notion of visualization's purpose North.<sup>94</sup> Insight-based evaluation is a technique aimed at acquiring valuable insights through visualization that differentiates a system from others and underscore its unique qualities when compared to alternative models. To evaluate this capability, visualizations can be measured in terms of insights. Typically, visualizations are evaluated using heuristics and expert reviews or controlled studies that measure user performance on specified tasks. Measuring insight enables a direct comparison to be made of visualization design alternatives and directly leads to visualization refinement and improvement.

Researchers evaluated several visualization systems using insight-based evaluation approaches and identified many issues, challenges, and advantages. Visualizations are frequently used to help the user gain insight into a data set. For example, North<sup>94</sup> introduced insight-based evaluation as a methodology to assess the effectiveness of visualization in facilitating individuals' acquisition of meaningful insights from the presented information. In the proposed qualitative

insight analysis, the users verbalize their findings using a think-aloud protocol so that evaluators can capture the users' insights. A more modern approach is to conduct insight-based evaluations, in which participants are given open-ended, complex tasks and asked to report on the insights gained. Another research study promoted the insight-based evaluation of visualizations which required teams to report on insights gained while exploring data North.<sup>94</sup> Saraiya et al.<sup>45</sup> developed and implemented an insight-based approach to visualization evaluation that we believe may be applied to a variety of data domains. Their evaluation technique focuses on identifying and quantifying insights gained through exploratory visualization. Furthermore, they expanded their explanation and discussion of this insight-based evaluation technique and applied the method to bioinformatics visualizations Saraiya et al.<sup>95</sup> Their insight definition allowed them to quantify insight generation using a range of insight characteristics, allowing them to assess bioinformatics visualization technologies' open-ended insight capability. However, this process requires time to capture insights. The researchers overcome these difficulties and limitations (user motivation) in future work. Therefore, it is clear from the aforementioned research that insight analysis is far more effective and is widely used as an evaluation technique.

*Log data analysis:* The evaluation, analysis, and comprehension of computer-generated records known as log analysis. Many programmable technologies produce logs, including networking devices, operating systems, apps, and more. A log is a collection of messages in a sequence that describes what is going on in a system. Log files can be broadcast to a log collector over an active network or saved in files to be reviewed later. Log analysis is the delicate process of evaluating and interpreting these messages to obtain insight into the system's inner workings.

Researchers have explored various fields using log data analysis. For example, Vartak et al.<sup>96</sup> explored FlowSense as a new natural language interface for visual data exploration in a dataflow visualization system which enables the user to grasp the underlying parsing status using real-time feedback on certain labeled utterances. Lerche and Kiel<sup>92</sup> provided a linear model of student achievement that combines prior knowledge and log file-extracted online behavior as predictors. The model displayed a good fit with data obtained in three separate scenarios. He et al.<sup>84</sup> targeted automated log parsing for the large-scale log analysis of modern systems. They analyzed various state-of-the-art log parsing approaches in depth, evaluating their accuracy, efficiency, and efficacy in relation to future log mining jobs. They conducted an extensive evaluation on synthetic and real-world data sets,

and their findings suggest that their parallel log parser, POP can reliably and effectively handle large-scale log data. Therefore, the abovementioned research shows that log data analysis is useful and is widely used as a VA evaluation technique.

### *Human factors*

Human factors (HF) evaluation and design techniques are well-established across various HCI domains, however HF evaluation in VR is complex and encompasses multiple aspects such as human performance, cognition, and sensory capabilities. Consequently, HF plays a pivotal role and makes substantial contributions to the visualization process, tool design, and evaluation. When conducting HF analysis, it is advisable to employ objective metrics that measure performance based on quantifiable characteristics, such as the number of errors made or the time taken to complete a task.

*Likert scale/validation points analysis:* As indicated in numerous surveys, the most widely utilized evaluation tool is the Likert scale. A Likert scale is a psychological tool for assessing attitudes, values, and opinions. It involves individuals completing a questionnaire where they express their level of agreement or disagreement with a series of statements. The Likert scale assumes a linear relationship between the strength and intensity of an experience. The respondents' agreement with different statements is typically rated on a five, seven, or nine-point scale. Each item is assigned a numerical score, enabling quantitative data analysis and visualization through graphs or charts.

Various researchers have implemented the Likert scale to evaluate data in a visualization system Islam et al.<sup>97</sup> The Likert scale is popular in survey research because it allows personality traits or perceptions to be operationalized quickly. For example, in 2020, Qian et al.<sup>64</sup> proposed the first end-to-end graphical recommendation system based on ML. They formalized and described a generic learning framework to solve the problem of ML-based visualization recommendations and used trained models to automatically generate, and evaluate a list of recommended views for new data sets which are unknown to arbitrary users. Indeterminate Likert scaling based on TRINS was introduced by Kandasamy et al.<sup>28</sup> To address inconsistent, uncertain, vague, and indeterminate records. Yu and Silva<sup>77</sup> designed and evaluated a combination of the LoRaWAN network with AR visualization. The instrument is a practical application addressing the sensor infrastructure maintenance use case to study the utility of such a combination in a close-to-life scenario. The designed application helps to locate faulty sensors and keep track of data accuracy. Ferrer et al.<sup>83</sup>

compared user perception of two approaches to temperature data visualization in tangible augmented reality on mobile phones: (i) the current particle-based visualization and (ii) novel virtual human-based visualizations. Visualizing the Likert scale using horizontal diverging stacked bar charts is an excellent method to see how the participants respond to questions or statements on a survey or questionnaire. However, only a limited number of visual analytic frameworks have been developed, and the Likert scale procedure proves to be comprehensive and effective in evaluating those data visualization systems compared to others. As a result, most Likert-type scales require a diverging stacked bar chart to effectively convey their findings and insights.

*Qualitative and quantitative feedback analysis:* This is a comprehensive approach to assessment that combines both quantifiable results and metrics with individualized conversations. Quantitative feedback involves collecting factual data on employee and company performance, which can be used for routine evaluations or to establish new goals. On the other hand, qualitative feedback involves more personalized discussions about performance and quality. It can be gathered through various methods, including focus groups, document/material reviews, ethnographic involvement, and interviews. Both qualitative and quantitative feedback formats have their own advantages in different contexts and both may be necessary for a thorough review of a system, website, or mobile app.

Various researchers used qualitative evaluation and quantitative feedback analysis for data visualization. For example, Haleem et al.<sup>7</sup> presented a CNN-based model to evaluate the readability based on the coordinates of nodes and edges in the graph layout. They used previous representative algorithms to create the network design dataset and traditional methods to mark these design images with readability values as a fundamental truth. The suggested CNN model is then trained using this graphic layout image and the readability metric values provided. Steyn et al.<sup>25</sup> identified how feedback might be strengthened in the context of competency-based education. They concluded after reviewing the research that an evaluation that should precede feedback should be planned to maximize its didactic effect. Ali et al.<sup>89</sup> conducted two qualitative studies to evaluate two versions of LOCO-Analyst, a learning analytics tool. They update the system by updating the graphical user interface and applying data visualization techniques to show its generated inputs.

Therefore, from the aforementioned research, it is evident that qualitative evaluation and quantitative feedback analysis have emerged as highly productive and widely employed techniques in the field of VA.

These techniques offer valuable insights and play a crucial role in enhancing the effectiveness of VA methods.

*Nielsen's heuristics:* Heuristic evaluation is a streamlined and easily integrable process that can be incorporated within development iterations. It is as an informal method of usability analysis, involving a group of evaluators who are given an interface design and asked to provide comments and feedback. This approach was initially proposed by Nielsen and Molich.<sup>98</sup> It was originally developed as a usability engineering method in which the evaluators were asked to consider the system's fundamental technical limitations. It is also a well-known bargain evaluation technique in HCI and visualization. The primary aim of this evaluation method is to reveal usability flaws in an existing design. Without communicating with others, heuristic evaluations allow us to identify and focus on specific concerns. Furthermore, heuristic evaluation is used to discover usability issues with particular elements and how they affect the overall user experience.

Several researchers used Nielsen's heuristics to analyze various visualization systems and compiled a list of their flaws and limitations. For example, Strobel et al.<sup>99</sup> deployed heuristics evaluation techniques on visualization and usability guidelines. They summarized expert reviews and stated that heuristic evaluation should be used to analyze visualization systems. Dal et al.<sup>100</sup> discovered the benefits and drawbacks of a hierarchy visualization tool using heuristic evaluation. Shneiderman<sup>13</sup> proposed the well-known visual information-seeking mantra which is the heuristic evaluation of fact visualization primarily based on task and usefulness. Using heuristic evaluation, the mantra represents the summary data retrieved via experience, occasional empirical evidence, and practice designing visualizations. Christmann et al.<sup>24</sup> extracted the usability problems through heuristic evaluation. They identified the issues with the usability approach. They concluded the contribution of a heuristic evaluation to total usability efforts reduced the potential for adverse consequences. Zoller et al.,<sup>90</sup> performed a meta-evaluation to look at the challenges of heuristic assessment for InfoVis, and additionally mentioned the generalizability and categorization of these heuristics. They evaluated the usefulness of heuristic utilization and identified implications for further research into the heuristic evaluation process in InfoVis. Still, researchers much prefer heuristic evaluation rather than others. As a result, the heuristic evaluation technique has a wide range of applications, particularly in visualization, since it is more thorough and makes it easier to discover the scope of existing systems and overcome their limitations.

*Eye trackers for evaluation systems:* Eye tracking is a well-known evaluation technique which was first proposed to evaluate maps but it has been increasingly used in the last 10–15 years Du et al.<sup>88</sup> Eye tracking has become a popular method for analyzing user behavior, becoming a new human-computer interaction and visualization research approach. Eye tracking is a unique method to study usability issues. Using eye tracking to evaluate visualization systems is advantageous and useful. It is also utilized for the analysis of human-computer interactions and to augment user interaction.

Several researchers proposed and implemented eye tracking support for VAS, which can be extended to be more supportive and adaptive by exploring eye tracking evaluation systems. For example, Silva et al.<sup>101</sup> used eye tracking to control a degree-of-interest display when analyzing hierarchically organized data, as shown in Figure 4. They proposed a framework to explore the eye tracker's raw data and demonstrate the value of incorporating eye tracking into the VA system. Blascheck,

John and Kurzhals<sup>58</sup> provided a novel method for visualizing eye tracking data, which is significant for user behavior analysis and overall evaluation. They analyze various VA systems using eye tracking techniques whenever they analyzed participants' information. Another example is an exploration conducted on automatically guided data by identifying user interests using an eye tracker Shao et al.<sup>85</sup> Popelka et al.<sup>102</sup> introduced an eye tracking system named the EyeTribe tracker, which is designed for qualitative data recording. They used an eye tracking system to study and examine the quality of the experiment design and analyze the result to explore the strategies for design improvement. They also proposed EyeTribe accuracy evaluation method. IrinaFabrikant et al.<sup>103</sup> utilized eye tracking to assess a map series illustrating the progression of a phenomenon over time, as well as user comprehension of weather maps.

Based on the aforementioned research, it is evident that eye tracking technology is a highly productive and extensively utilized evaluation technique in the field of VA. Compared to data accuracy, the use of eye tracking technology has gained prominence due to its comprehensiveness and practicality in evaluating VA systems. Although there are already a few VA systems in existence, the application of eye tracking provides a comprehensive and practical means to assess and compare these systems against others.

## Discussion

As the field of data visualization continues to evolve, researchers have introduced new methodologies and

have engaged in extensive discussions on the utilization of specific evaluation methods. These advancements aim to make the evaluation of data visualization more accessible and user-friendly. Table 4 overviews these methodologies and their contributions to the field. Following a thorough evaluation of the literature, seven common situations of evaluation research were identified by Lam et al.,<sup>5</sup> which provide a helpful overview. In addition, although evaluation might be a long and laborious process, there is a reasonable probability that essential solutions will be accepted and implemented in real-world scenarios.

In our study, we analyzed the existing surveys in the field. Over the years, numerous evaluation systems have been proposed, as highlighted in Table 1. However, it is worth noting that several researchers have focused on applying single or multiple evaluation methods based on specific fields, rather than providing a systematic review of all the evaluation methods commonly used in the visualization field. Therefore, there is a need for a comprehensive survey that encompasses all the evaluation methods extensively applied in the visualization domain. This survey paper aims to provide a systematic review of evaluation procedures that are effectively applied in the field of VA. Despite being a topic of great interest among researchers, a comprehensive survey that demonstrates the strategies of evaluation methods remains somewhat elusive. While most researchers have utilized various evaluation methods for analyzing VAS, our survey paper takes a strategic approach by providing an overview of all significant evaluation methods. Our main objective is to offer a comprehensive and strategic understanding of the most important evaluation methods in the field of VAS, allowing readers to gain valuable insights at a glance.

## Challenges

In the following, we outline vital insights into the challenges facing future research using the aforementioned evaluation techniques.

- **Selection of relevant evaluation strategies:** A significant challenge lies in identifying the most suitable evaluation strategies to include in any given application. Given the multitude of available strategies, it is imperative to meticulously select those that are highly relevant to effectively evaluate visual analytics systems.
- **Validation of evaluation metrics:** It is crucial to validate the effectiveness of the evaluation metrics used in the identified ES. Researchers should examine the reliability, validity, and sensitivity of these metrics in capturing the desired

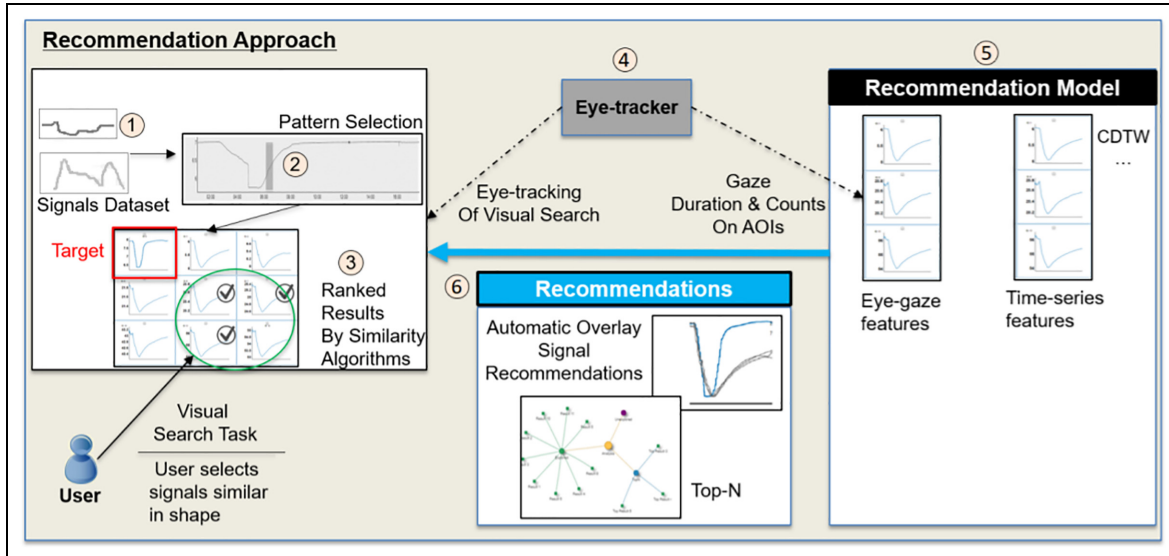


Figure 4. Eye tracking Evaluation System (Figure courtesy of Shao et al.<sup>85</sup>).

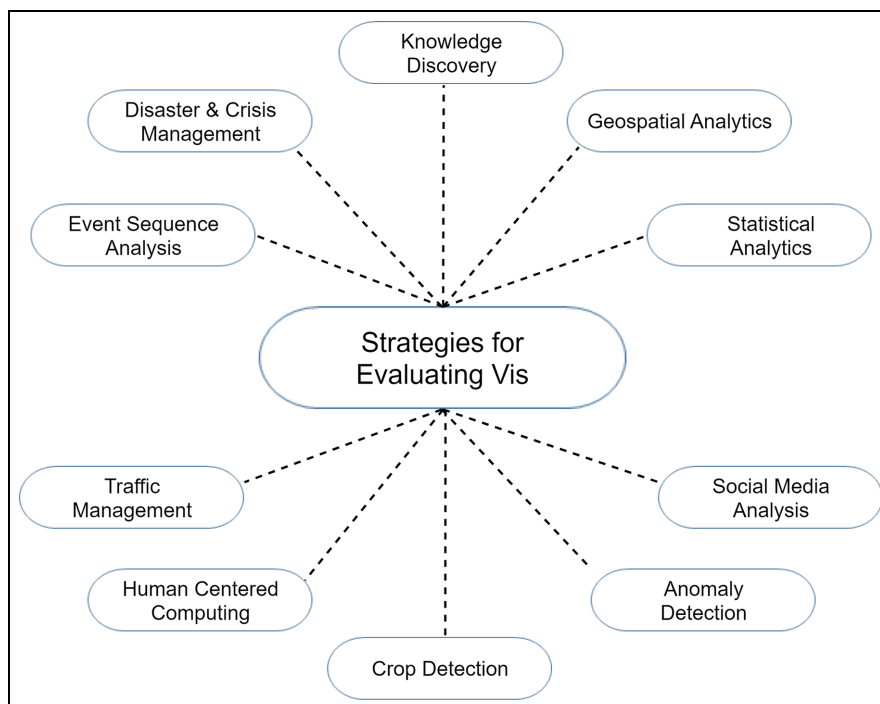


Figure 5. Miscellaneous applications of evaluating visualization system.

outcomes and assessing the performance of visual analytics systems accurately.

- **Addressing the limitations of existing evaluation strategies:** Recognizing and acknowledging the drawbacks of the chosen evaluation methodologies is a significant problem. This involves identifying any potential gaps in their

ability to fully capture the complexity and subtleties of visual analytics systems and devising ways to fill such gaps.

- **Interpretation and synthesis of findings:** Analyzing and synthesizing the findings from a wide range of evaluation strategies can be complex. Ensuring a clear and coherent presentation

**Table 4.** Overview of various visualization framework and their evaluation techniques.

| Framework          | Likert Scale | Eye trackers | Log data analysis | Dashboard comparison | Insight based evaluation | Qualitative Analysis | Quantitative Analysis | Nielsen's heuristics | References                       |
|--------------------|--------------|--------------|-------------------|----------------------|--------------------------|----------------------|-----------------------|----------------------|----------------------------------|
| ExtraVis           |              |              |                   | ✓                    |                          | ✓                    | ✓                     |                      | Zuk et al. <sup>60</sup>         |
| LSTMVis            |              |              |                   |                      |                          |                      |                       |                      | Strobel et al. <sup>99</sup>     |
| VizML              |              |              |                   | ✓                    |                          |                      |                       |                      | Hu et al. <sup>75</sup>          |
| Data2Vis           |              |              |                   |                      | ✓                        |                      |                       |                      | Dibia and Demiralp <sup>74</sup> |
| Generic Learning   | ✓            |              |                   |                      |                          |                      |                       |                      | Qian et al. <sup>64</sup>        |
| DGViz              |              |              |                   |                      |                          | ✓                    |                       |                      | Li et al. <sup>67</sup>          |
| ActiVis            |              |              |                   |                      |                          | ✓                    | ✓                     |                      | Kahng et al. <sup>15</sup>       |
| Recommender System |              | ✓            |                   |                      |                          |                      |                       |                      | Shao et al. <sup>85</sup>        |
| FlowSense          |              |              | ✓                 |                      |                          |                      |                       |                      | Zerafa et al. <sup>62</sup>      |
| InfoVis            |              |              |                   |                      |                          |                      |                       |                      | Saraiya et al. <sup>45</sup>     |
| MicroarrayVis      |              |              |                   |                      |                          |                      |                       |                      | Amar and Stasko <sup>17</sup>    |
| DAV                |              |              |                   |                      |                          |                      |                       |                      | Swaid et al. <sup>87</sup>       |
| Dqnviz             |              |              |                   |                      | ✓                        |                      | ✓                     |                      | Webster and Watson <sup>68</sup> |
| AirVis             |              |              |                   |                      |                          | ✓                    |                       | ✓                    | Deng et al. <sup>76</sup>        |

of the key themes, trends, and insights emerging from the literature can be challenging but essential for providing meaningful contributions to the field.

- **Identification of open challenges:** Identifying and addressing the open challenges in evaluating visual analytics systems is crucial. This involves critically analyzing the current state of evaluation strategies and highlighting areas that require further exploration and innovation.
- **Integration of historical context:** It can be challenging to provide the historical background of evaluation procedures in the visualization industry. This entails following the development of these tactics over time and comprehending their historical importance in light of the status of visual analytics evaluation at the moment.

### Future directions

In this article, various existing efforts related to the application of several evaluation techniques have been detailed from different perspectives. However, there are still gaps in the research that need to be addressed. Therefore, the following potential future research directions are proposed:

- The protocols related to heuristic evaluation approaches demonstrate a concept for deploying a heuristic method that uncovers new insights in the field of heuristic evaluation. However, a heuristic evaluation questionnaire can be used to rate the assessment, using a five-point Likert scale.
- The eye tracking data indicated that the users had difficulty extracting information from specific areas and frequently hunted for precise details in non-relevant regions of the screen. According to eye tracking evaluation data, a critical aspect is that the screen was too small. This limitation can be mitigated by exploring new strategies to extract user data.
- Insight-based evaluation techniques are being increasingly used. It is challenging to acquire accurate insights and analyze them to evaluate a visualization system. Although humans rely on explainable evaluation methods, they interpret them to make decisions. Therefore, adding explainability to insights will make them more valuable. Researchers can explore new strategies and apply advanced technology to gain precise insights.
- Heuristics might be redundant, inconsistent, or even context-specific, hampering the heuristic evaluation process. There could be flaws in its

validation and there is a need for more rigor, robustness, and standardization in its analysis.

- As new technologies continue to evolve in the field of visual analytics, it is crucial to evaluate their impact and effectiveness.<sup>105</sup> Further research can explore the utilization of established ES to assess the effectiveness of advanced technologies such as augmented reality, virtual reality, or natural language processing in visual analytics systems. Additionally, future studies can also investigate the possibility of combining and integrating diverse evaluation methodologies to enhance our understanding of these systems as shown in Figure 5.
- Shneiderman and Plaisant<sup>104</sup> combine observation in the typical user environment, automated activity monitoring, and long-term involvement with researchers. Longitudinal research was not used as an evaluation approach in any of the papers in our research. So adapting new strategies to longitudinal studies will be a direction for future research.
- Very few studies employed observational techniques to study how users interact with visualization systems where system interactions are logged and analyzed as part of the evaluation. Thus, several new directions for security visualization system evaluation and design are explored by logging and analyzing interactions.

## Conclusion

It is necessary to use novel evaluation methods to improve a visualization system's development. Existing studies reveal that the general level of rigor of reporting evaluations needs to be improved. There needs to be a detailed discussion on how evaluation methods work in visualization research to significantly improve the impact of research results. Therefore, in this paper, we provide a comprehensive review with an emphasis on seven evaluation methods, namely (1) dashboard comparison; (2) insight-based evaluation; (3) log data analysis; (4) Likert scales; (5) qualitative and quantitative analysis; (6) Nielsen's heuristics; (7) eye tracker. We investigated the limitations of these works and identified the open research challenges facing all seven methods. We also provided several future research directions.

In short, this study provides a comprehensive analysis of the various state-of-the-art evaluation methods and their implementation in different applications. This study will help enhance the current visualization system development landscape to solve multiple problems.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We would like to acknowledge the financial support from the Australian Research Council (ARC) under Grant No. LE220100078 and DP220103717.

## ORCID iD

Md Rafiqul Islam  <https://orcid.org/0000-0001-7209-3881>

## Data availability

No data was used for the research described in the article.

## References

1. Mandal A, Sinaeepourfard A and Naskar SK. VDA: Deep learning based visual data analysis in integrated edge to cloud computing environment. In: *Adjunct Proceedings of the 2021 International Conference on Distributed Computing and Networking*, 5 January 2021, pp.7–12. New York, NY, United States. Association for Computing Machinery.
2. Saraiya P, North C, Lam V, et al. An insight-based longitudinal study of visual analytics. *IEEE Trans Vis Comput Graph* 2006; 12: 1511–1522.
3. Keim D, Andrienko G, Fekete JD, et al. Visual analytics: definition, process, and challenges. In: *Information visualization*. Berlin, Heidelberg: Springer, 2008, pp.154–175.
4. Caban JJ and Gotz D. Visual analytics in healthcare – opportunities and research challenges. *J Am Med Inform Assoc* 2015; 22: 260–262.
5. Lam H, Bertini E, Isenberg P, et al. Empirical studies in information visualization: seven scenarios. *IEEE Trans Vis Comput Graph* 2012; 18: 1520–1536.
6. Sacha D, Sedlmair M, Zhang L, et al. What you see is what you can change: human-centered machine learning by interactive visualization. *Neurocomputing* 2017; 268: 164–175.
7. Haleem H, Wang Y, Puri A, et al. Evaluating the readability of force directed graph layouts: a deep learning approach. *IEEE Comput Graph Appl* 2019; 39: 40–53.
8. Thomas JJ and Cook KA. *A visual analytics agenda*. *IEEE Comput Graph Appl* 2006; 26: 10–13.
9. Li H, Xu M, Wang Y, et al. A visual analytics approach to facilitate the proctoring of online exams. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems 2021* May 6, 2022, pp.1–17.
10. Lipton ZC. The mythos of model interpretability: In ML, the concept of interpretability is both important and slippery. *Queue* 2018; 16: 31–57.
11. Plaisant C, Fekete JD and Grinstein G. *Promoting insight-based evaluation of visualizations: from contest to benchmark repository*. *IEEE Trans Vis Comput Graph* 2008; 14: 120–134.



12. Liu S, Maljovec D, Wang B, et al. *Visualizing high-dimensional data: advances in the past decade*. *IEEE Trans Vis Comput Graph* 2017; 23: 1249–1268.
13. Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In: *The craft of information visualization*. Elsevier, 2003, pp.364–371.
14. Varu R, Christino L and Paulovich FV. ARMatrix: an interactive item-to-rule matrix for association rules visual analytics. *Electronics* 2022; 11(9): 1344.
15. Kahng M, Andrews PY, Kalro A, et al. ActiVis: visual exploration of industry-scale deep neural network models. *IEEE Trans Vis Comput Graph* 2018; 24: 88–97.
16. Song H, Lee J, Kim TJ, et al. GazeDx: Interactive visual analytics framework for comparative gaze analysis with volumetric medical images. *IEEE Trans Vis Comput Graph* 2017; 23(1): 311–320.
17. Amar RA and Stasko JT. Knowledge precepts for design and evaluation of Information Visualizations. *IEEE Trans Vis Comput Graph* 2005; 11: 432–442.
18. Cook KA and Thomas JJ. *Illuminating the path: The research and development agenda for visual analytics*. Technical Report. Richland, WA: Pacific Northwest National Lab (PNNL).
19. Stadler JG, Donlon K, Siewert JD, et al. Improving the efficiency and ease of healthcare analysis through use of data visualization dashboards. *Big Data* 2016; 4: 129–135.
20. Kosara R. Visualization criticism-the missing link between information visualization and art, In: *2007 11th International Conference Information Visualization (IV'07)*, 2007, pp.631–636. New York: IEEE.
21. Liu S, Cui W, Wu Y, et al. A survey on information visualization: recent advances and challenges. *Vis Comput* 2014; 30: 1373–1393.
22. Kosara R, Healey CG, Interrante V, et al. User studies: Why, how, and when? *IEEE Comput Graph Appl* 2003; 23: 20–25.
23. Lu Y, Garcia R, Hansen B, et al. The State-of-the-Art in predictive visual analytics. *Comput Graph Forum* 2017; 36: 539–562.
24. Christmann O, Fleury S, Migaud J, et al. Visualizing the invisible: user-centered design of a system for the visualization of flows and concentrations of particles in the air. *Inf Vis* 2022; 21(3): 311–320.
25. Steyn C, Davies C and Sambo A. Eliciting student feedback for course development: the application of a qualitative course evaluation tool among business research students. *Assess Eval High Educ* 2019; 44: 11–24.
26. Isenberg T, Isenberg P, Chen J, et al. A systematic review on the practice of evaluating visualization. *IEEE Trans Vis Comput Graph* 2013; 19: 2818–2827.
27. Andrews K. Evaluating information visualization. In: *Proceedings of the 2006 AVI workshop on beyond time and errors: novel evaluation methods for information visualization*, 2006, pp.1–5.
28. Kandasamy I, Kandasamy WBV, Obbineni JM, et al. Indeterminate likert scale: feedback based on neutrosophy, its distance measures and clustering algorithm. *Soft Comput* 2020; 24: 7459–7468.
29. Chen C and Yu Y. Empirical studies of information visualization: a meta-analysis. *Int J Hum Comput Stud* 2000; 53: 851–866.
30. Forsell C. A guide to scientific evaluation in information visualization. In: *2010 14th International Conference Information Visualization*, 2010. pp.162–169. New York: IEEE.
31. Liu S, Wang X, Liu M, et al. Towards better analysis of machine learning models: A visual analytics perspective. *Vis Inform* 2017; 1: 48–56.
32. Green TM, Ribarsky W and Fisher B. Visual analytics for complex concepts using a human cognition model. In: *2008 IEEE Symposium on Visual Analytics Science and Technology*, 2008, pp.91–98. New York: IEEE.
33. Bostandjiev S, O'Donovan J and Höllerer T. Taste-weights: a visual interactive hybrid recommender system. In: *Proceedings of the sixth ACM conference on Recommender systems*, 2012, pp.35–42.
34. Islam MR, Razzak I, Wang X, et al. Ucb-Vis: understanding customer behavior sequences with visual interactive system. In: *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp.1–8. New York: IEEE.
35. Islam MT, Islam MR, Akter S, et al. Designing dashboard for exploring tourist hotspots in Bangladesh. In: *The 23rd International Conference on Computer and Information Technology (ICCI-T-2020)*, IEEE.
36. Wang J, Gou L, Shen HW, et al. Dqnviz: a visual analytics approach to understand deep q-networks. *IEEE Trans Vis Comput Graph* 2018; 25: 288–298.
37. Keim D, Qu H and Ma KL. Big-data visualization. *IEEE Comput Graph Appl* 2013; 33: 20–21.
38. Abdulla N, Ahmed Z, Gazzali M, et al. Measure customer behaviour using c4. 5 decision tree map reduce implementation in big data analytics and data visualization. *Int J Innov Res Sci Technol* 2015; 1: 228–235.
39. Tupikovskaja-Omovie Z and Tyler D. Eye tracking technology to audit google analytics: Analysing digital consumer shopping journey in fashion m-retail. *Int J Inf Manag* 2021; 59: 102294.
40. Basole RC and Patel SS. Transformation through unbundling: visualizing the global fintech ecosystem. *Serv Sci* 2018; 10: 379–396.
41. Almainoni H, Altuwajjri N, Asiry F, et al. Developing and implementing web-based online destination information management system for tourism. *Int J Appl Sci Eng Res* 2018; 13: 7541–7550.
42. Folorunso O and Ogunseye OS. *Challenges in the adoption of visualization system: a survey*. *Kybernetes*, 2008; 37(9/10): 1530–41.
43. Geisler G. Making information more accessible: A survey of information visualization applications and techniques, <http://www.ils.unc.edu/~geisg/info/infovis/paper.html> (1998, accessed 31 January).
44. Ackerman MJ. The visible human project®: From body to bits. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp.3338–3341. IEEE.

45. Saraiya P, North C and Duca K. An evaluation of microarray visualization tools for biological insight. In: *IEEE Symposium on Information Visualization*, 2004, pp.1–8. IEEE.
46. Dowding D and Merrill JA. The development of heuristics for evaluation of dashboard visualizations. *Appl Clin Inform* 2018; 9: 511–518.
47. Gena C. Methods and techniques for the evaluation of useradaptive systems. *Knowl Eng Rev* 2005; 20(1): 1–37.
48. Kurzhals K, Fisher B, Burch M, et al. Evaluating visual analytics with eye tracking. In: *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, 2014, pp.61–69.
49. Onoue Y, Kukimoto N, Sakamoto N, et al. E-grid: a visual analytics system for evaluation structures. *J Vis* 2016; 19: 753–768.
50. Deriu J, Rodrigo A, Otegi A, et al. Survey on evaluation methods for dialogue systems. *Artif Intell Rev* 2021; 54: 755–810.
51. Yasmin R, Salminen M, Gilman E, et al. Combining iot deployment and data visualization: experiences within campus maintenance use-case. In: *2018 9th International Conference on the Network of the Future (NOF)*, 2018, pp.101–105. IEEE.
52. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 2021; 88: 105906.
53. Valiati ER, Freitas CM and Pimenta MS. Using multidimensional in-depth long-term case studies for information visualization evaluation. In: *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization*, 2008, pp.1–7.
54. Wall E, Agnihotri M, Matzen L, et al. A heuristic approach to value-driven evaluation of visualizations. *IEEE Trans Vis Comput Graph* 2018; 25: 491–500.
55. Kurzhals K, Fisher B, Burch M, et al. Eye tracking evaluation of visual analytics. *Inf Vis* 2016; 15(4): 340–358.
56. Leung C, Kononov V, Pazdor A, et al. PyramidViz: visual analytics and big data visualization for frequent patterns. In: *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSci-Tech)*, 2016 Aug 8, pp.913916. IEEE.
57. Kamaleswaran R, James A, Collins C, et al. CoRAD: Visual Analytics for Cohort Analysis. In: *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, 2016 Oct 4, pp.517–526. IEEE.
58. Blascheck T, John M, Kurzhals K, et al. VA 2: A visual analytics approach for evaluating visual analytics applications. *IEEE Trans Vis Comput Graph* (2015); 22(1): 61–70.
59. Nazemi K, Burkhardt D and Kock A. Visual analytics for technology and innovation management. *Multimed Tools Appl* 2022; 81(11): 14803–14830.
60. Zuk T, Schlesier L, Neumann P, et al. Heuristics for information visualization evaluation. In: *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 2006, pp.1–6.
61. Qian X, Rossi RA, Du F, et al. Learning to recommend visualizations from data. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp.1359–1369.
62. Zerafa J, Islam MR, Kabir A, et al. ExTraVis: Exploration of traffic incidents using visual interactive system. In: *25th International Conference Information Visualization (IV 2021)*, 2021, IEEE, Institute of Electrical and Electronics Engineers.
63. Sahu K, Bai Y and Choi Y. Supervised sentiment analysis of twitter handle of president trump with data visualization technique. In: *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 2020, pp.0640–0646. IEEE.
64. Qian X, Rossi RA, Du F, et al. ML-based visualization recommendation: Learning to recommend visualizations from data. *arXiv preprint arXiv* 2020; 2009: 12316.
65. Stehle S and Kitchin R. Real-time and archival data visualisation techniques in city dashboards. *Int J Geogr Inf Sci* 2020; 34: 344–366.
66. Samuel J, Ali GGMN, Rahman MM, et al. COVID-19 public sentiment insights and machine learning for tweets classification. *Information* 2020; 11: 314.
67. Li R, Yin C, Yang S, et al. Marrying medical domain knowledge with deep learning on electronic health records: A deep visual analytics approach. *J Med Internet Res* 2020; 22(9): e20645.
68. Webster J and Watson R. Analyzing the past to prepare for the future: Writing a literature review. *MIS Q* 2002; 2002: xiii–xiii.
69. Beasley Z, Friedman A, Pieg L, et al. Leveraging peer feedback to improve visualization education. In: *2020 IEEE Pacific Visualization Symposium (PacificVis)*, 2020, pp.146–155. IEEE.
70. Lee C, Kim Y, Jin S, et al. A visual analytics system for exploring, monitoring, and forecasting road traffic congestion. *IEEE Trans Vis Comput Graph* 2020; 26: 3133–3146.
71. Bourqui R, Giot R, et al. Toward automatic comparison of visualization techniques: application to graph visualization. *ArXiv* 2019; arXiv–1910.
72. Ku CH, Chang YC, Wang Y, et al. Artificial intelligence and visual analytics: A deep-learning approach to analyze hotel reviews & responses. In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
73. Chen CY. Using an eye tracker to investigate the effect of sticker on line app for older adults. In: *International Conference on Human-Computer Interaction*, 2019, pp.225–234. Springer.
74. Dibia V and Demiralp C. Data2vis: automatic generation of data visualizations using sequence-to-sequence recurrent neural networks. *IEEE Comput Graph Appl* 2019; 39: 33–46.
75. Hu K, Bakker MA, Li S, et al. VizML: A ML approach to visualization recommendation. In: *Proceedings of the*

- 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp.1–12.
76. Deng Z, Weng D, Chen J, et al. Airvis: Visual analytics of air pollution propagation. *IEEE Trans Vis Comput Graph* 2020; 26: 800–810.
  77. Yu B and Silva CT. FlowSense: A natural language interface for visual data exploration within a dataflow system. *IEEE Trans Vis Comput Graph* 2020; 26: 1–11.
  78. GarciaCaballero HS, Westenberg MA, Gebre B, et al. V-Awake: A visual analytics approach for correcting sleep predictions from deep learning models. *Comput Graph Forum* 2019; 38: 1–12.
  79. Barcellos R, Viterbo J, Bernardini F, et al. An instrument for evaluating the quality of data visualizations. In: *2018 22nd International Conference Information Visualization (IV)*, 2018, pp.169–174. IEEE.
  80. Kwon BC, Choi MJ, Kim JT, et al. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Trans Vis Comput Graph* 2018; 25: 299–309.
  81. Chang C, Dwyer T, Marriott K, et al. An evaluation of perceptually complementary views for multivariate data. In: *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 2018, pp.195–204. IEEE.
  82. Du M, Chou JK, Ma C, et al. Exploring the role of sound in augmenting visualization to enhance user engagement. In: *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 2018, pp.225–229. IEEE.
  83. Ferrer V, Perdomo A, Ali HR, et al. Virtual humans for temperature visualization in a tangible augmented reality educational game. In: *2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)*, 2017, pp.1–6. IEEE.
  84. He P, Zhu J, He S, et al. Towards automated log parsing for large-scale log data analysis. *IEEE Trans Dependable Secure Comput* 2018; 15: 931–944.
  85. Shao L, Silva N, Eggeling E, et al. Visual exploration of large scatter plot matrices by pattern recommendation based on eye tracking. In: *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, 2017, pp.9–16.
  86. Ming Y, Cao S, Zhang R, et al. Understanding hidden memories of recurrent neural networks. In: *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2017, pp.13–24. IEEE.
  87. Swaid S, Maat M, Krishnan H, et al. Usability heuristic evaluation of scientific data analysis and visualization tools. In: *International Conference on Applied Human Factors and Ergonomics*, 2017, pp.471–482. Springer.
  88. Du F, Plaisant C, Spring N, et al. EventAction: Visual analytics for temporal event sequence recommendation. In: *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2016 Oct 23, pp.61–70. IEEE.
  89. Ali L, Hatala M, Gašević D, et al. A qualitative evaluation of evolution of a learning analytics tool. *Comput Educ* 2012; 58: 470–489.
  90. Zöllner MA, Titov W, Schlegel T, et al. XAutoML: A Visual Analytics Tool for Establishing Trust in Automated ML. arXiv preprint arXiv:2202.11954, 2022.
  91. Chang R, Ziemkiewicz C, Green TM, et al. Defining insight for visual analytics. *IEEE Comput Graph Appl* 2009; 29: 14–17.
  92. Lerche T and Kiel E. Predicting student achievement in learning management systems by log data analysis. *Comput Human Behav* 2018; 89: 367–372.
  93. Keim DA and Kriegel HP. Visualization techniques for mining large databases: a comparison. *IEEE Trans Knowl Data Eng* 1996; 8: 923–938.
  94. North C. Toward measuring visualization insight. *IEEE Comput Graph Appl* 2006; 26: 6–9.
  95. Saraiya P, North C and Duca K. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Trans Vis Comput Graph* 2005; 11: 443–456.
  96. Vartak M, Huang S, Siddiqui T, et al. Towards visualization recommendation systems. *ACM SIGMOD Rec* 2017; 45: 34–39.
  97. Islam MR, Liu S, Razzak I, et al. MhiVis: Visual analytics for exploring mental illness of policyholder’s in life insurance industry. In: *2020 7th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC)*, 2020, IEEE.
  98. Nielsen J and Molich R. Heuristic evaluation of user interfaces. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1990, pp.249–256.
  99. Strobelt H, Gehrmann S, Pfister H, et al. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Trans Vis Comput Graph* 2018; 24: 667–676.
  100. Dal CM, Freitas S, Luzzardi PR, et al. Evaluating usability of information visualization techniques. In: *Brazilian Symposium on Human Factors in Computing Systems*, 2002.
  101. Silva N, Shao L, Schreck T, et al. Sense. me-open source framework for the exploration and visualization of eye tracking data. In: *Proceedings of the 2016 IEEE Conference on Information Visualization*, 2016.
  102. Popelka S, Stachoň Z, Šašinka, et al. EyeTribe tracker data accuracy evaluation and its interconnection with hypothesis software for cartographic purposes. *Comput Intell Neurosci* 2016; 2016.
  103. IrinaFabrikant S, Rebich-Hespanha S, Andrienko N, et al. Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes. *Cartogr J* 2008; 45: 201–215.
  104. Shneiderman B and Plaisant C. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 2006, pp.1–7.
  105. Islam MR, Akter S, Ratan MR, et al. Deep visual analytics (DVA): applications, challenges and future directions. *Hum-Centric Intell Syst* 2021; 1: 3–17.

