Critical Design Strategy: a Method for Heuristically Evaluating Visualisation Designs

Jonathan C. Roberts, Hanan Alnjar, Aron E. Owen and Panagiotis D. Ritsos

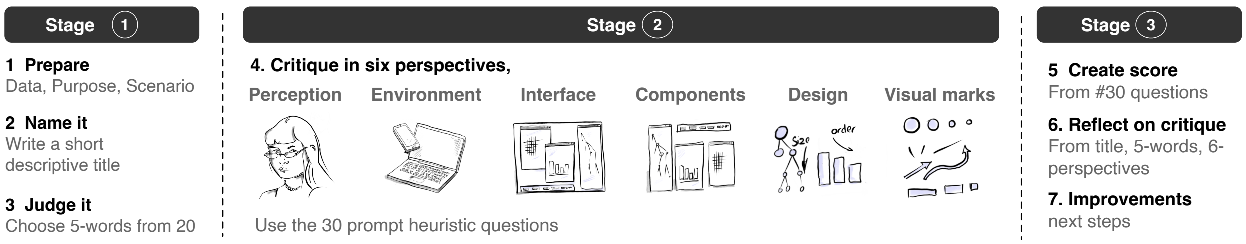


Fig. 1: The Critical Design Strategy consists of three stages: ❶ overview, ❷ detail, and ❸ review. The assessor begins by considering the design holistically—naming and summarising it while selecting five keywords from a set of twenty. Next, they conduct an in-depth evaluation across six perspectives, using 30 heuristic questions or directly engaging with semantic differential word pairs (opposite adjectives). Finally, the assessor reflects on their critique, assigns an indicative score, and determines areas for improvement.

Abstract— We present the Critical Design Strategy (CDS)–a structured method designed to facilitate the examination of visualisation designs through reflection and critical thought. The CDS helps designers think critically and make informed improvements using heuristic evaluation. When developing a visual tool or pioneering a novel visualisation approach, identifying areas for enhancement can be challenging. Critical thinking is particularly crucial for visualisation designers and tool developers, especially those new to the field, such as studying visualisation in higher education. The CDS consists of three stages across six perspectives: Stage 1 captures the essence of the idea by assigning an indicative title and selecting five adjectives (from twenty options) to form initial impressions of the design. Stage 2 involves an in-depth critique using 30 heuristic questions spanning six key perspectives—user, environment, interface, components, design, and visual marks. Stage 3 focuses on synthesising insights, reflecting on design decisions, and determining the next steps forward. We introduce the CDS and explore its use across three visualisation modules in both undergraduate and postgraduate courses. Our longstanding experience with the CDS has allowed us to refine and develop it over time: from its initial creation through workshops in 2017/18 to improvements in wording and the development of two applications by 2020, followed by the expansion of support notes and refinement of heuristics through 2023; while using it in our teaching each year. This sustained use allows us to reflect on its practical application and offer guidance on how others can incorporate it into their own work.

Index Terms—Visualisation design, Design critique, Pedagogy, Visualisation theory, Information visualisation, Teaching visualisation

1 Introduction

One of the most important skills that a visualisation designer needs to possess is that of critical thinking. Creating effective visualisations is a combination of technical skills, domain knowledge, and attention to detail, along with design and critical thinking. There are many ways to improve technical skills. Proficiency in data analysis and use of visualisation software such as Tableau, ggplot, D3.js or matplotlib can be gained through tutorials, workshops, classes, and books, and so forth. Domain knowledge can be learnt, or experts involved may provide it. Yet, there are few methods that help people think critically over their visualisation designs. Critical thinking methods outlined in the literature typically consist of broad catalogues of overarching concepts. Additionally, due to the ad hoc nature of design, it is often difficult for people to know how to structure their critical thought process. Indeed, people must consider different designs, layouts, and visualisation arrangements, while gauging the suitability of each option in the use-case they set to tackle.

The requirement for having such skills is becoming more crucial to a wider community of people, as data becomes more pervasive across various industries and sectors. The demand for individuals with strong critical thinking and visualisation design skills is growing rapidly. In particular, students of various levels and disciplines, seek to learn how to create data visualisations, as means of analysis and support for their work. While expert visualisation developers rely on their knowledge, and experience of what works (or does not), learners struggle to understand where to start, or how to organise their critical thinking. Heuristic guidelines [34, 50, 82], where people evaluate a tool against a set of recognised usability principles (the *heuristics*) can help. But even with these strategies, learners can find it difficult to critique designs, due to lack of experience, and often struggle to know how to go about critiquing designs in an systematic way.

To address these challenges, we present the Critical Design Strategy (CDS), extending our IEEE VIS 2023 poster presentation [56]. Consisting of three stages, and several thought-provoking segments, the CDS aids individuals in critically reflecting on visualisation designs. It serves as a valuable tool to structure critical thinking for design visualisation. It is especially useful for educators in the classroom, facilitating the generation of insights that can be leveraged to craft critique reports (documents that provide feedback on a specific design, which offer constructive criticism and recommendations for improvement). Stage ❶ of the strategy gets individuals to consider the design holistically. In Stage ❷ they dive into detail, and systematically consider 30 aspects of the design (from six perspectives). In Stage ❸ individuals calculate an average score, reflect on their critique and decide what to do next.

- *Jonathan C. Roberts, Aron E. Owen and Panagiotis D. Ritsos are with Bangor University. E-mail: {j.c.roberts aron.e.owen p.ritsos}@bangor.ac.uk.*
- *Hanan Alnjar is with University of Basrah, and previously Bangor University, h.r.alnjar@uobasrah.edu.iq*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

We contribute: (i) a broad discussion on the rhetoric of critique with a particular focus on visual critique (Sec. 3), (ii) a detailed breakdown of the Critical Design Strategy (Sec. 4), including an in-depth exploration of each of the three stages, six perspectives and 30 heuristic questions. In addition, we provide supplementary material with comprehensive notes that can be used by teachers. (iii) The design and evolution of the CDS, from its inception in 2017/18 through multiple workshops to its adaptations over the years (Sec. 5, and Fig. 5). This section includes several evaluations, including an assessment of the use of the 20-words, which led to refinements, an initial evaluation with 30 students, and a review of the latest version. (iv) Finally, we reflect on its use in our teaching across eight cohorts (see Fig. 5) and highlight key focus areas to help others adopt and apply it in their own work (Sec. 6).

2 Background

We are strong advocates for using (and developing) *explanatory frameworks*. In the context of science education, explanatory frameworks are particularly relevant for helping learners develop a deep understanding of scientific concepts [76]. They can be found across domains, spanning cell theory, evolutionary theory, energy transfer and plate tectonics. The visualisation community also emphasises their research, e.g., Bach et al [8] call for theoretical frameworks tailored to visualisation education and those offering practical guidelines. Explanatory frameworks serve as guiding structures for individuals to navigate and comprehend diverse subjects effectively, and as roadmaps that help learners organise thoughts, explore information systematically, and understand complex concepts more easily. In visualisation there are several explanatory frameworks, such as Bertin's method of mapping data onto visual properties [10], the data-flow paradigm, Shneiderman's visual information-seeking mantra [67], Amar and Stasko's knowledge and task-based framework [3], and Munzner's nested model [48].

This work builds on three of our previous *explanatory frameworks*; our Explanatory Visualisation Framework (EVF) [61], a design and build strategy for courses where people create explanatory visualisations, the Five Design-Sheets (**FdS**) [58, 59], a method to consider alternative designs through sketching, and the Critical Thinking Sheet (CTS) [60], a strategy to encourage learners to critically think, and sketch their algorithm, before coding.

We started to develop the Critical Design Strategy (CDS) around 2015, while reflecting on critical thinking, design and pedagogical approaches in general. Our pedagogic strategy is twofold. First, learners choose a dataset, analyse it, and perform a low-fidelity design study – we use the FdS, which they write as a design report. Second, learners develop a prototype visualisation based on their design and submit both their implementation and a reflective report, and in most cases we request students to use Processing.org to create their visualisations. We chose Processing to encourage the development of unique and innovative visual solutions rather than relying on predefined visualisation types such as bar charts or line graphs. This gives two graded submission points (Assessment 1 and 2), each with a report: (a **design report**, and an **implementation report with critical reflection**).

We realised, however, that learners struggled to critique their visualisations, and did not know how to structure their reports. They expressed attitudes, such as "*I like it*" or it "*it looks good*", rather than trying to objectively judge the designs/visualisations against knowledge of design thinking [80], tasks [67], retinal variables [10], Gestalt theory [49, 59], good design principles [55], and so forth — all of which had been taught and explained in previous lectures. When grading, we consistently provided feedback, urging learners to *"justify your decisions"* and "*structure your critical thinking report*" as we often found their reports lacked clear organisation. With critical thinking, learners need to (1) ask questions, (2) collect information, (3) contemplate alternative potential solutions, (4) understand and empathise different viewpoints, and finally (5) communicate in a clear way [33]. Learners were not doing any of these tasks effectively. What we felt was required is a formal structure and a set of steps to help people critique their ideas in a systematic way.

The CDS provides an 'explanatory framework' [76], with three core parts (overview, detail, review), see Fig. 1, and a structure for results reporting and discussion. We introduced the CDS in our teaching around 2017 [2], refined the questions and structure until 2020, and published a IEEE VIS 2023 poster presentation [56]. We give two lectures: one on critical thinking, and another specifically on the CDS. We have used this strategy with an Information Visualisation (InfoVis) module; a 20 credit compulsory module for an MSc in Data Science, and optional for MSc in Advanced Computer Science (CS) and BSc Computer Science (MSc version is 180 credits long). We estimate over 300 students have used the CDS. In our classes students use the CDS (at least) twice. First, after ideating on their design, using the FdS, they carry out a CDS critique on their realisation design (sheet 5 of the FdS). Second, they critique their implementation using the CDS, and write a reflective report which follows the CDS's structure (see Sec. 4, Fig. 5).

3 Related Work

Critical analysis is at the heart of academic tradition of reason and argument and is taught across the curriculum [11, 52]. We first consider general rhetoric and critique, and then visual reflection and visualisation specifically.

3.1 Rhetoric and critique

Especially for writing critical essays, the art of contrastive rhetoric writing, which inspires this work, encourages the reviewer to break the work into individual parts, and identify strategies that the author has used to persuade the audience. Likewise, in a design context, we are breaking the visual depiction into different categories, considered individually, to make an overarching judgement on the design. Influenced by Aristotle, Cicero described five canons of rhetoric': *inventio* (invention/creation), *dispostio* (arrangement), *elocutio* (style), *memoria* (memory/recollection) and *pronuntiatio* (delivery) [66], which, although intended for public speaking, describe the writing process as well. These terms can also be mapped to the Aristotle's means of persuasion, of *ethos*, *pathos* and *logos*: detailing the trustworthiness and credibility of the orator, how they use emotion to tender support and their logical argument. These subsections influence the six perspectives of the CDS. Consequently, there are several skills that learners aspiring to be critical thinkers require [17, 30, 31, 36, 51], such as being informed, honest and open, orderly, and not shying from hard work. Facione writes: "*The ideal critical thinker is habitually inquisitive, well-informed, .. honest in facing personal biases, prudent in making judgements, willing to reconsider, clear about issues, orderly in complex matters, .. focused in inquiry, and persistent..* " [33].

While we now have a written, rather than oral, rhetoric tradition, these ideas have been refined and taught across levels in our education system, for all types of learning, and different critical thinking tasks. Acronyms such as SOAPSTone are useful, for writers to consider the Speaker, Occasion, Audience, Purpose, Subject and Tone of a written document. There are several well-known educational models that include critical reflection, such as: a) Borton's [14] 'what', 'so what', and 'now what', b) the five W's *who*, *what*, *why*, *when*, *where* [32], and c) Roberts' et al. should-you, could-you, what-if-you [59]. Many of these models stem from problem solving (e.g., Polya [54] and Duncker [27]), whereas educational concepts are often influenced by Dewey [25], David Kolb's reflective model of experiential learning [41], and Ennis' [30] on dimensions of critical thinking (*logical*, *criteria* and *pragmatic*). More importantly, critical thinking should not only happen at the end of the design process. As Borton [14] suggests, it needs to take place when considering past activities, while something is taking shape and for anticipating what could happen in the future. While useful, however, these structures are not design or visualisation specific.

3.2 Visual critique and visualisation

Design reflection is an important aspect of general interface design and therefore visualisation tool design. There are many ways to evaluate interfaces, including *automatically* (using algorithms and metrics), *empirically* (assessed by real users [19]), *formally* (using formulas and usability measures), and *informally* (based on the skill and experience of evaluators [50]). Our method fits within the informal classification, which are classified as heuristic methods, where experts judge compliance of an interface against predefined heuristics [50]. Our idea is that the CDS process is used formatively alongside the design process (e.g., ideate with the FdS [58], perform a CDS on the proposed design, adjust design, implement, re-evaluate with CDS).

Kosara [42] explains that critical inspection is a useful form of evaluating visualisations and writes "*visualisation criticism could be a tool for further developing and increasing the usefulness of visualisation theory*" [43] but does not offer a formal structure. We already have strategies that help us consider alternative visualisation designs, such as the FdS method [58], sketching user interfaces (e.g., Buxton [16]), and Munzner's Nested Model [48], which provide guidance how to develop and consider visualisations. To facilitate *creative thinking* for visualisation design, we need to employ *critical thinking*. Critical thinkers seek to find alternative perspectives that they analyse in an open-minded way, avoiding hasty decisions [52]. If we are unable to assess the products we create and produce, and ascertain what to change, then we will not be able to improve our solutions. Yet, getting towards this goal is not only about considering alternatives (divergent thought) but also employing both convergent and critical thinking approaches to reach a suitable solution [54]. Inevitably, critical thinking must take place throughout the whole design process, from deciding what data to select, how to process and enhance the data, to what visual design to choose, etc. Thus, developers need to simultaneously employ bottom-up and top-down thinking. They need to be open to alternative potential solutions (however unusual they are), while concurrently envisioning the end output, how to create it and be able to judge whether the problem at hand is solved [27].

Bottom-up methods [16, 58]) may help people ideate alternatives, but the process is still challenging. It is difficult to be open-minded, unbiased, put personal opinions aside, and it is too easy fixate on one solution (c.f., Facione [33]). It can be problematic to know how to organise the information appropriately to make the right decisions. Working together in a group can help to moderate views. Jackson et al. [39] employ a group critiquing process that is integrated with sketching and design. We, likewise, integrate critical analysis with the design, but in our case the individual critic (rather than a group) makes judgements on the visualisation, using our formal structure for the critical process. Saraiya et al. [63] also use insight-based methods, and provide a structure of eight characteristics: observation, time, domain value, hypothesis, directed vs. unexpected insights, correctness, breadth vs. depth and categorisation. This latter work is the closest to ours, yet we develop a more systematic structure, define a set of question that can be readily taught, learnt, and used in various situations including critique assignments. The CDS also delivers a result (score), which can help people to compare compare previous critical analysis of visualisations, and indicate improvement.

Top-down critical thinking also requires the user to imagine how a potential solution would appear, how it would be utilised in the intended environment, and how it could work using available technologies. Experts, when characterising the problem [48], build up a broad picture of solutions in their mind. They can draw on past experiences, and are able to abstract the ideas and create many visions of potential futures, in order to decide on the right result [71]. But for learners this is much more challenging [36]. They do not have the structures, schema and cannot rely on years of experience. Learners need *explanatory frameworks* [76] to facilitate the systematic exploration and interpretation of the idea. Thus, as researchers, we need to create structures and methods to support people in their critique.

Informal styles of evaluation are useful, especially to help focus and organise personal thoughts during the design process and to aid critical thinking [15]. Amar and Stasko [3] walk the user through thinking about rationales and tasks, and Munzner [48] implicitly advocates critical thinking throughout the whole visualisation creation process. In fact, ideas of visual inspection have a long history. Bertin [10] expounds three levels to read a graphic, from the elementary (looking at visual variables and marks), to look at patterns within the presentation, to observing the whole. Cleveland [21] also describes three perception operations: detection, assemble and estimation. Lohse [47] uses similar ideas to help classify visual representations, and develops a model to understand graphical depictions [46], while inspection techniques were used by Conversy et al. [22] to observe visualisations. Such work provides guidelines on how to design visualisations, which are detailed in many books (e.g., [49, 59, 81]). Yet, while these books provide a rich tapestry of information, it is difficult for learners to aggregate this into a single critical review structure.

There is a growing interest in evaluating visualisations [5, 18, 38, 44, 53, 65, 79]. However, most of the current techniques require that the tool has been fully built, and evaluate usability or user experience of the tool [44]. Consequently, several researchers have produced classifications on the different types of evaluation strategies. For instance, Plaisant [53] focused on four areas of: (i) controlled experiments, (ii) usability evaluation of the tool, (iii) controlled experiments comparing tools and (iv) case studies of tools in settings. In another classification, Isenberg et al. [38] extended the seven scenarios of Lam et al. [44] and provide a comprehensive review of evaluation techniques in visualisation. Using their categorisation we can place our CDS in the category of Understanding Environments and Work Practices (UWP), as a user would need to make judgements on how they imagine the tool fitting into the environment where it would be used. If the CDS is used as part of the refinement process then it becomes an informal evaluation, and could be classified within their Evaluating Visual Data Analysis and Reasoning (VDAR) category.

There are several works on heuristics in visualisation, highlighting benefits in designer/expert collaboration [68, 75]. Heuristics can be general, e.g., "*Consider people with colour blindness*" or specific "*ensure visual variable has sufficient length*", or "*provide multiple levels of detail*" [82]. Many researchers focus on ten (like Nielsen [50]) while others aggregate several lists. E.g., Forsell and Johnson present ten [34, 35], Zuk and Carpendale present twelve [82] drawn from Tufte [77], Ware [81] and Bertin [10]. Engelbrecht et al. [29] aggregates a list. Scholtz [64] explains that evaluators of the VAST challenge drew heuristics from Zuk and Carpendale [82], Shneiderman's visual information seeking mantra [67] and Amar and Stasko's knowledge and task-based framework [3]. Other heuristic lists include [24, 26, 73, 78]). Heuristics help people think, but they are typically not encapsulated in a formal structure. One exception is Wall et al. [79] who structure them around *insight*, *time*, *essence* and *confidence* [70]. Additionally, Eppler and Burkhard [32] organises them by *what*, *why*, *who*, and *when*. Our approach is also structured, but we follow a double-diamond approach [23] of starting broad (summary), delving deep (30 questions in six perspectives), and a final general step (reflection).

Finally, Wall et al. [79] present a heuristic-based evaluation methodology, used by experts, for assessing finalised interactive visualisations, in terms of perceived value. In comparison, the CDS is directed to non-experts and is designed to provoke critical thought during the design process and help the user contemplate their design decisions, and it can be used to consider how the implementation could be improved.

4 The Critical Design Strategy (CDS)

The final incarnation of the CDS comprises of three stages, carried out sequentially: **overview**, **detail** and **review** Fig. 1. In this section we describe the purpose of each stage as well as the process the *appraiser* needs to follow. We provide vignettes to summarise each stages and to emphasise the important points. The subject of appraisal is a visualisation artefact that can be a sketch, an interactive visualization tool, a physicalisation or even a poster. This artefact, displays *data*, was crafted by a *designer*, coded by a *developer* and is eventually going to be used by a *user*. These individual roles could be achieved by different people, or the same person; e.g., a learner designs a data visualisation and then develops the code to display it.

1. **Overview**. After suitable preparation, assign a name, summarise its essence, and holistically critique by selecting five words.
2. **Detail**. Critique artefact. The appraiser conducts a thorough critique by responding to 30 heuristic questions across six perspectives.
3. **Review**. Finally, the appraiser reflects on both the overall critique and the detailed analysis to identify the next steps.

Stage 1 – Overview

Assign a name to the design:
Summarise essence:

Circle 5 (first impression) words:

clear confusing sensible indifferent clever reliable pointless indistinctive complex organised moderate spectacular useless average bad fulfilling useful fair vague beautiful

Fig. 2: Following adequate preparation, assign name, summarise essence, conduct a holistic critique by selecting five descriptive words.

4.1 Stage 1 – Overview

The primary objective of the first stage is to ensure a thorough understanding of the topic and to make holistic assessments of the artefact. Critical thinking necessitates individuals to be "*well-informed*" [33]. Individuals should adequately **prepare** and ensure a thorough understanding of both the challenge and associated data. Data visualisation cannot be pursued without access to data. It crucial to consider the composition of the data and its organisational aspects, such as sparsity and structure. This involves identifying variables, understanding their nature (categorical, ordinal, quantitative, etc.), and recognising the purpose for which the data was collected. Additionally, comprehending the main objective of the visualisation and the intended user tasks is essential. Contextual information, including the creator's intent and the environment in which the visualisation will be utilised, should be understood to ensure effective use.

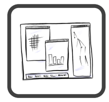
To confirm understanding, the appraiser **name** the artefact/design, and summarise its **essence**. The act of naming the design commences the critical thinking process. Crafting a brief, concise title (of two or three words) compels consideration of what is crucial [58]. The holistic critique continues, by **circling five** of the twenty words (Fig. 2). This task acts as a preliminary, intuitive assessment. While such initial assessments may be wrong, they are reviewed in the Review stage 3.

4.2 Stage 2 – Detail

During the second stage, the aim is for the appraiser to conduct a comprehensive critique (Fig. 3), by considering 30 heuristic questions in six perspectives (**User, Environment, Interface, Components, Design, Visual Marks**), each with its own Likert scale. These six perspectives encourage a top-down approach in the critique, with the questions themselves encouraging reflection while providing a structure for appraising different design viewpoints. The critique of this stage can be written as a report (as when we use the CDS in assessments). An overall score can be calculated at this stage. We detail some of the possible questions to reflect upon at this stage in our Supplementary material.

User: The first step is to consider the user's perspective and empathise with their point of view, and expectations in terms of their skills and experience. Topics to reflect upon include task suitability, how understandable the visualisation is, its trustworthiness and usefulness, echoing more traditional sub-dimensions of User Experience.

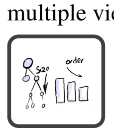
Environment: The next step is to consider the suitability of the proposed environment for end usage. This needs to consider the overall scenario of usage, where will this take place physically, access to different technologies and their potential interoperability. The appraiser should consider whether the design of the artefact is appropriate and well-suited for the intended purpose, environment, need, and with appropriate ergonomics? E.g., a static display for an e-book, while a 3D environment is suitable for an immersive head-mounted display. Likewise, considerations should assess what interaction affordances are available for sensemaking in terms of the environment. The term *interaction* should be understood in the appropriate context. Rather than simply evaluating whether it is an interactive Human-Computer Interaction (HCI) tool (e.g., using a mouse), assess whether the level and mode of interaction align with the specific situation and usage context. For example, if the artefact is a visualisation poster, interaction occurs when people physically move closer or further away from it. Ergonomics questions should be considered, such as whether physical interaction can take place, say by moving closer/away from a powerwall [4], but environmental obstacles may impede such interaction.



User Interface: Following the critique on the usage environment, the assessor should reflect on the user interface available. This is more often an interactive, programmed interface encompassing different interactive visualisation techniques, often employing specific hardware (e.g., HMDs). However, the interface can also be physicalisations, printed outputs and even specialised representations such as haptic visualisations. Depending on the interface flavour, the appraiser must consider the specific features of the user interface, such as menus, buttons, drag-and-drop functionality, visual programming, etc.

Much like in the case of the environment, ergonomics and usability are important elements in terms of the suitability of the interface. For instance, a drag-and-drop command may produce an ergonomic visualisation interface — it is easy to control, and quick to operate — yet the number of available options may hinder the user's experience. Likewise, certain interactions in an immersive environment may be difficult to manage over time and the user get tired, or motion sickness affects them. Visual displays are often made from many different facets and coordinate views: are the sizes of these views right for the task? E.g., in a comparison task comparing data A to data B, we may assume that the size given over to display both datasets is about the same (so not to bias one to the other). Or, in a webviewer, we may assume a central view will sit alongside adverts (are the size of the adverts suitable; perhaps too big or small in comparison to the visual display). Is spacing utilised efficiently, as it can enhance clarity and emphasise certain elements yet it can be detrimental as it can lead to confusion by disrupting desired relationships between elements.

Components: Components are specific visual elements or depictions that can be identified and isolated for individual consideration. Identifying and understanding these components is essential for analysing the structure and effectiveness of the design/visualisation as a whole. Each component serves a distinct purpose within the design, artefact or visualisation, contributing to the overall presentation and conveying specific information to the viewer. Components can include various graphical elements such as charts, graphs, tables, icons, or other visual representations of data; an identified depiction [20], along with menus, labels, help-information and so forth. They can be displayed in different ways, e.g., grid or in a tabbed window [1]. The appraiser needs to consider if important components are missing, such as interactive elements, interaction handlers and buttons, or even appropriate data types. Likewise, the type of output should facilitate interpretation and task execution, and align with the purpose of the presentation and the required task e.g., representing continuous data with a line graph and categories with a bar chart. Considerations should also include whether the relationships between different parts of the display are evident, such as in the case of multiple view visualisation.



Design: Design encompasses organising any part of the system, which involves considerations like colour balance, item alignment, and styling. This category focuses on more visually aesthetic aspects of the visualization or the supporting interface. While attractiveness is subjective, humans tend to prefer balanced designs [81], those that are well-proportioned. Colour choices fitting with the task at hand are important [13, 37], whereas UIs following good design principles [55, 69] are always preferred. In terms of how the design facilitates sensemaking, there are questions regarding the representation, aggregation and appropriate (statistically) depiction of the data. Finally, basic yet important elements such as legends, labels, titles, etc. [28] can be of paramount significance, and only ommited if that makes sense.

Visual marks: The last step requires the appraiser to consider the graphical marks utilised in the visualisation and their arrangement, ensuring the appropriate marks are used in the correct locations with the correct attributes. Evalu-

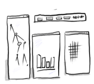
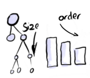
Stage 2 – Detail (comprehensive critical evaluation)						
Perspective		**Heuristic question**	-2 -1 0 1 2	**Range of answers, from poor to good**		
User 	#1	Is suitable for the user and task	○○○○○	Unsuitable	↔	Suitable
	#2	Is understandable for user and task to hand	○○○○○	Incomprehensible	↔	Understandable
	#3	It doesn't require guesswork	○○○○○	Requires guesswork	↔	Clear assumptions
	#4	Is trustworthy	○○○○○	Distrustful	↔	Trustful
	#5	Would be useful	○○○○○	Useless	↔	Useful
Environment 	#6	It would fit in with other technologies	○○○○○	Wrong setting	↔	Right setting
	#7	Uses suitable technology	○○○○○	Unsuitable technology	↔	Right technology
	#8	Has appropriate interaction	○○○○○	Unsuitable interaction	↔	Appropriate interaction
	#9	Its sizing is correct	○○○○○	Unsuitable size	↔	Suitable physical size
	#10	Gives a positive ambience	○○○○○	Poor vibe/ambience	↔	Positive ambience
Interface 	#11	Suitable user interface	○○○○○	Unsuitable GUI	↔	Suitable GUI
	#12	Ergonomic interface	○○○○○	Uncomfortable	↔	Ergonomic
	#13	Facets are sized suitably	○○○○○	Poorly proportioned	↔	Suitable sized facets
	#14	Interface suitably spaced	○○○○○	Poor facet spacing	↔	Relevant spacing
	#15	Suitable quantity of interface parts	○○○○○	Unsuitable facet quantity	↔	Suitable facet quantity
Components 	#16	Has all necessary components	○○○○○	Missing components	↔	All necessary components
	#17	Has all suitable output/view types	○○○○○	Unsuitable types	↔	Suitable view types
	#18	Clear relationships between parts	○○○○○	Unclear correspondences	↔	Clear view relationships
	#19	Task can be easily performed	○○○○○	Task unfulfilled	↔	Task easily performed
	#20	Suitable organisation of components	○○○○○	Poor component layout	↔	Good component layout
Design 	#21	Inspiring design	○○○○○	Uninspiring	↔	Inspiring
	#22	Aesthetic and visually attractive	○○○○○	Unattractive	↔	Visually attractive (aesthetic)
	#23	Good composition and space utilisation	○○○○○	Poor layout	↔	Good composition
	#24	Suitable coverage of data/underpinning facets/concepts	○○○○○	Unsuitable coverage	↔	Suitable coverage
	#25	Clear instructions, labels, legends to give context	○○○○○	Poor labels/legends	↔	Suitable legends/labels
Visual marks 	#26	Right choice of channels to communicate things clearly	○○○○○	Poor choice of channels	↔	Good channel choices
	#27	Communicates appropriate relationships/morphisms	○○○○○	Inappropriate mappings	↔	Appropriate mappings
	#28	The types of marks used, communicate things well	○○○○○	Inappropriate mark types	↔	Suitable mark types
	#29	Components are shown at the right level of abstraction/detail	○○○○○	Poor scale/zoom	↔	Good scale/zoom
	#30	Nothing is hidden that shouldn't be hidden	○○○○○	Overplotting	↔	Clear display, easy read

Fig. 3: Conduct a comprehensive critical evaluation of the artefact/design. Follow the questions (in the six perspectives: User, Environment, interface, components, design, marks), recording the answers in the Likert scale. Make notes that justify your decisions.

ating for redundant ink or chartjunk [77], aspects of design [81], and how it remains memorable [12] are important considerations. Issues relating to scale, mark legibility, mapping validity and clarity for the display medium available are also important. For multisensory visualisation systems, considerations beyond visual depictions, such tangible, vibrotactile and auditory representations require special considerations, including in terms of mapping efficiency, data transformation as well as accessibility [40]. Similarly, for immersive visualisations, issues such as occlusion should be also considered.

4.3 Stage 3 – Review

The goal of this final stage (Fig. 4) is to distil the key findings and observations, and translate these insights into actionable steps that contribute to an improved design/artefact. Initially a score is calculated by summing the Likert scale values and reflect on each part (the name, essence, six perspectives and so forth). Although the average score derived from the Likert scale calculation provides utility, it may be misinterpreted, as it conceals numerous facets of the actual critique. Therefore, it should be interpreted in conjunction with other insights. Review the six perspectives and the 30 questions. Identify any standout perspectives among the six. Assess the strengths and weaknesses, highlighting areas for improvement and identifying aspects to be strengthened.

After assessing the critique, it is important to consider the next steps. Perhaps a redesign to address identified issues and improve the overall design, which may involve brainstorming potential improvements, such as refining layout, focusing the adjusting visual elements, enhancing usability, or incorporating user feedback. The next steps would then include outlining a plan for redesign, which may include tasks such as conducting further research, gathering additional user input, creating prototypes, and implementing changes iteratively. It is essential to establish clear goals and objectives for the redesign and to regularly evaluate progress to ensure that the new design effectively addresses identified needs and enhances overall usability and user experience.

Stage 3 – Review	
Create score.	**Reflect on parts:**
Improvements and next steps:	

Fig. 4: The final stage involves synthesising the various perspectives and insights, gathered throughout the critique.

5 Design and Evolution of the CDS

In this section, we explore the design and evolution of the CDS over the past eight years (illustrated in Fig. 5), providing a critique of each version and explaining the rationale behind the development of its various forms. Throughout the design process, several methods were used to evaluate it, with the most significant being the application of the CDS in our teaching courses (see Sec. 2), where it was tested in real-world conditions.

5.1 CDS V.1 and V.2 2017-2019

The first version of the CDS was designed as a structured questionnaire. We started with a 2-day workshop investigating critiquing vocabulary (see Fig. 6), then designed the preliminary CDS structure and performed a talk-aloud evaluation [2]. We recruited 10 participants (six identified as male and four as female) with age between 25-40. Experiences ranged from ICT consultant, marketing expert, three doctoral students (visualisation, NLP and engineering), a postdoc in English, a mature student, with the remaining people were computing undergraduates. We formed two teams of five, and provided light lunch. We asked participants to consider eight tasks: (**T1**) individually write a definition of critical analysis and underline key words in their definition; (**T2**) In groups, brainstorm over 15 words associated with critical analysis and produce a word-cloud. (**T3**) Critique the given six images (Fig. 6); (**T4**) Reflect on the critiques and consider if all critiques were the same.

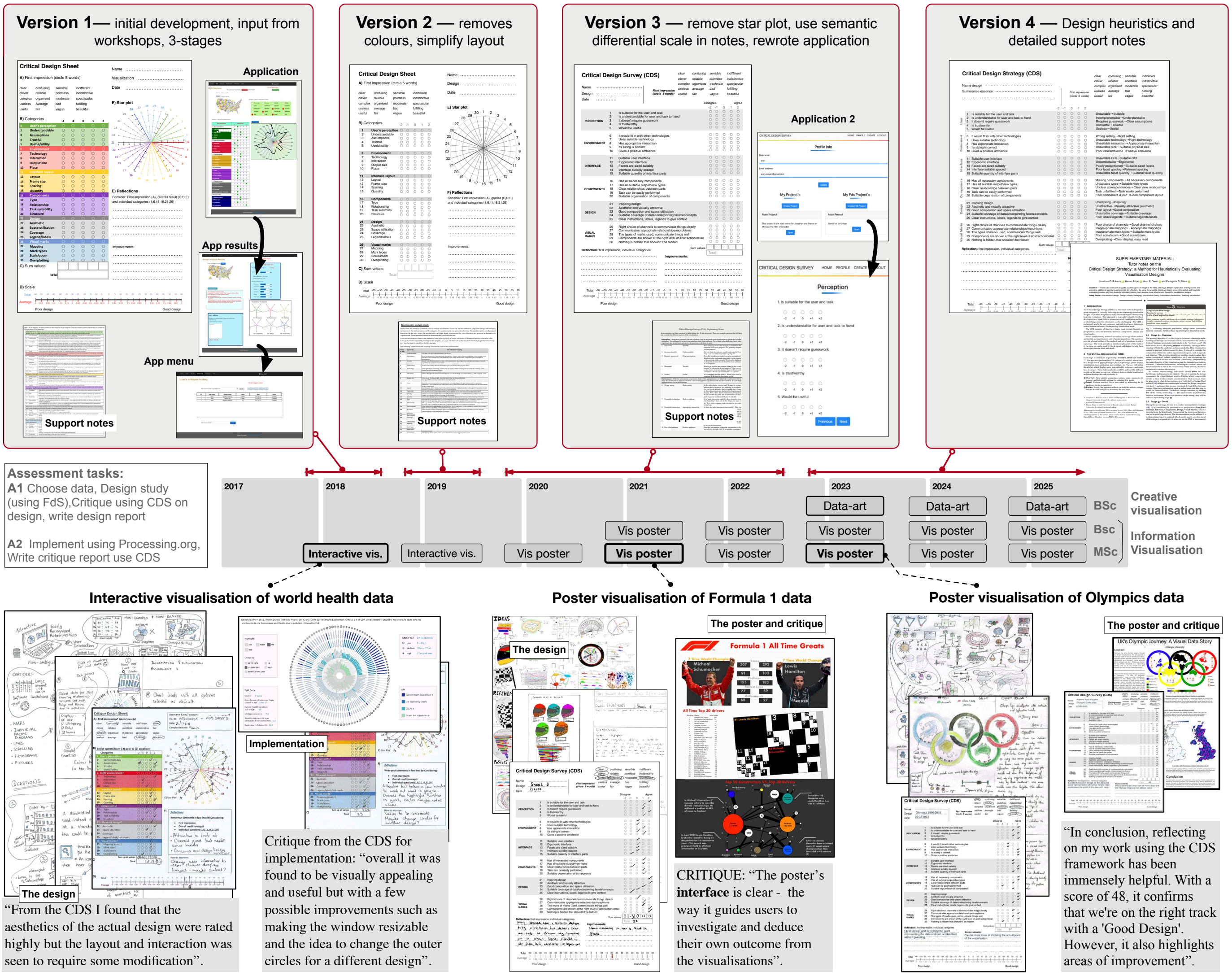


Fig. 5: Timeline of CDS Versions, aligned with the courses where they were used (MSc or BSc in information Visualisation and Creative Visualisation) and the tasks (either build an interactive visualisation, or a static visualisation poster, or data-art piece). Three students' work are included, where they perform a design study then the CDS; then implement their work and perform another critique with the CDS. Quotes of their use of the CDS are included. Version 1 was developed from workshops, introducing the initial sheet, support notes, and web application while establishing the three primary stages, and six design perspectives. Version 2 refined the layout by removing colours and simplifying the sheet. Version 3 brought substantial changes, including the removal of the star plot and a complete rewrite of the application. Version 4 introduced semantic differential words (pairs of opposite adjectives to clarify design extremes) along with expanded and revised support notes.

(**T5**) Write notes explaining similarities and/or differences. (**T6**) Think about and discuss the process of critiquing. (**T7**) Sketch a diagram to represent the main stages of the critiquing process (e.g., draw a flow diagram). (**T8**) Using the same process critique in turn the parallel coordinate plot and radar plot visualisation. Two observers took notes, and we recorded conversations, which we later transcribed.

Analysis of the transcript and notes were informative, confirming ideas of critical thinking from Facione [33]. For instance, participants focused on the idea of serious examination and judgement; one wrote "*critical analysis is to perform in-depth thinking, for making fair and balanced judgements*". Tasks T2,3,4 required teamwork, and to create a word cloud. When comparing Van Gogh's sunflowers to the kid's artwork (**T3**), *"I guess this is basic and this advanced"* (pointing in turn to the pictures) and another said *"one influenced the other"*. At that stage a participant added the word *"influence"* to the word cloud, and promptly said: *"we had some bias to some objects. We got influenced by our prior knowledge. We got influenced by thinking about the objects when we considered them"*; demonstrating awareness of bias [33]. To analyse the vocabulary, we looked at the words that the candidates underlined (**T1**), word clouds (**T2**) and their discussions (**T3** and **T4**). Underlined words were predominantly verbs, e.g., *think*, *formulate*, *imagine*, *improve*, and *analyse*. Participants used adjectives in written definitions, particularly with **T3**, and the word clouds. They include *good*, *bad*, *useful*, *clear*, and *fair*. We summarise their words: appealing, basic, busy, classic, colourful, confusing, dull, fit for purpose, fun, functional, influential, lack of context, minimalist, modern, old fashioned, plain, pretty, realistic perspective, reusable, simple, sophisticated, realistic scale, surreal, too detailed, unreadable.

When discussing the similarities and differences (**T5**) team 1 distilled their ideas into four categories, expressing that these cross-cut all the object and image types: (1) the *style* (and layout) of an image, (2) its *appearance* and what it looks like, (3) *comparison* with other ideas and to an idea of what it could look like, and (4) whether it was suitable for the *context* (or environment) where it would be used. Participants acknowledged that they needed to put their own emotions aside, become more 'academic' in their approach, and that some objects/pictures engendered more emotional responses than others. If they understood the image and its purpose, or if it was something they had seen, liked

Fig. 6: Photograph of participants at the critical-analysis workshop (left). Thumbnails of critiqued images (right). Minard's map of Napoleon's Russian campaign (© public domain); modern version of Minard's map (DkEgy CC BY-SA 4.0), radar plot, D3.js parallel coordinate plot. Reprint of Vincent Van Gogh's sunflowers, child's painting of same.

or used before, then their emotional response was stronger. They also emphasised that to critique the objects each participant needed to understand the *context* where it would be used. These responses match similar judgements by Ennis [30], where users need to employ *logic* to understand the judgement *criteria* and be *pragmatic* in their approach. For **T5**, team 2 likewise included *appearance* and *comparison*, but also emphasised the importance of the *initial thoughts* to the design, and in comparison to deep thought. This result is significant, and implement it by the action to 'circle 5 words' of the list of 20.

Finally, addressing the last three tasks (**T6**,**T7** and **T8**) we focused on team 2, as they completed the tasks in more detail and their answers are a superset of what group 1 produced. They produced a useful four-stage process, where the observer (1) thinks about the problem and defines the terms of critical analysis, (2) brainstorms ideas and organise their thoughts, (3) produces a balanced review by putting emotions aside, and finally (4) reflects on the analysis to reject or accept the critique. This is another important result because it confirms the process described by Polya [54] on problem-solving, where users should understand the problem, devise a plan, carry it out, and confirm the answer.

5.2 Developing the CDS strategy (V1)

To develop a list of suitable 'first-impression' words, we started with the list of adjectives. We used card-sorting, to decide on a list of 20 words. We made a compromise between having too many words and confusing the user, to not having enough words that could inhibit expression. We wanted to balance positive, neutral and negative sentiment words, and for that we used scores from the SENTIWORDNET 3.0 lexicon [7]. Sentiment analysis depends on the position of the word and how it is used, and each of our words have multiple sentiment scores in the lexicon [72]. We placed all words in a table and recorded an average of all scores in the lexicon for each word. After discussion we simplified the list to seven positive (**average, beautiful, clear, clever, reliable, sensible, spectacular**), seven negative (**bad, complex, indistinctive, pointless, confusing, useless, vague**) and six neutral (**fair, fulfilling, indifferent, moderate, organised, useful**) words.

For our first design, we tried to follow the System Usability Scale (SUS) [9], and alternated between positive and negative wording of questions. The advantage of this strategy is that it combats response acquiescence, minimising the results from participants who (say) make quick, injudicious, selections on one side of the Likert scale. We developed a list of questions, including "*I found the colour map suitable"*. *"I would imagine that most non-expert users would not easily understand data that is presented by this visualisation without any prior knowledge of the visualisation domain"*. *"My first impression is that I have a high confidence in this visualisation"*. *"I felt the visualisation has not achieved the design goals."* However, we faced several challenges. First, it was difficult to distil the subject to ten questions, as we did not believe these were enough for us to perform an in-depth critique of a visualisation design. Second, the questions seemed too convergent; we wanted the user to really think deeply about their work, not merely judge it whether it was right or wrong. Third, when we tested these questions, we found it difficult to complete the responses of alternating positive to negative question wording. We discussed this problem with colleagues, who had not been involved in our design or workshop, who likewise found it difficult to use.

Reflecting on this issue we realised that the goal of the SUS is different, as it is designed for summative assessment from end-users, with the results analysed once a prototype has been finalized. However, we want the designer/developer to use our method for their own work in a formative manner and make decisions on their own creations. For our situation, there is no need to alternate positive and negative to combat response acquiescence – why would a user not be honest with themselves – and the questions were too convergent to achieve our goals of prompting the user to think critically about the problem. We therefore looked to other questionnaires.

The UEQ [62] measures user-experience through a set of six categories: *Attractiveness*, *Efficiency*, *Perspicuity*, *Dependability*, *Stimulation* and *Novelty*, gauged using a 5-part Likert scale of (alternating positive and negative) 26 questions. The UEQ takes longer to complete than the SUS, and the creators used statistical analysis to develop a set of weights, to balance the results to an 'average user' [45]. However this further complicates how the results are calculated. We did not want users to have to input their values into a spreadsheet, so decided to implement a quick way to calculate the scores. Weighing up the different strategies, we decided to follow the UEQ design to develop a list of 30 questions.

To develop the six perspectives of the CDS, we were heavily influenced by the words of one workshop participant (drawing on Shneiderman's mantra [67]) who said *"when we are critiquing we need to be clear what we are looking at, we first need to understand what it is, then put aside bias and emotion. There are different levels of detail. We should look to the big picture before looking at details"*. We can consider design critique in a similar way. Start by thinking holistically and drill down into the finer detail. People can critique a design by thinking about the *user*, then the usage *environment*, the *interface* that would be used in said environment, and so on, until we get to individual graphical marks that are used to create the visualisation.

The next task was to develop the individual questions of the framework. Starting with key terms from the workshop we used a card-sorting process to categorise them into the six perspectives. We show these terms in Tab. 1. Placing the cards on a table, as a team, we moved cards with similar meaning words on top of each other, with the goal to reduce the overall quantity of indicative words to five per category. (Five cards, in six perspectives, gives 30 questions, which is close to 26 topics in the UEQ). We deliberated over individual parts, for instance whether we should include *'design'* or more specific terms, such as *'colour'*. From these prompt words we added words with the opposite meaning, and placed them on a Likert scale; 0 for a poor value (left), and 5 a good value (on the right). This meant that we could add up the values and calculate an overall score, without requiring to compensate for alternating scales (as in the SUS).

We deliberated much about creating an overall score of the user's critique. On one hand, having a single value that can be awarded to a visualisation is useful. For example, if a user uses the CDS for a second time, the overall score difference would allow them to quickly decide whether the visualisation has improved. On the other hand, a single average score, could potentially mask bad perspectives. This is similar to Anscombe's quartet [6] of statistical graphs, that look completely different but have similar statistical properties. We deliberated over using six scores (one for each category), and even considered not including a score at all. We decided to keep the score as we believe the benefits outweigh the negatives.

We evaluated this first version with three talk-aloud evaluation sessions (S1,S2,S3). In S1 we focused on comprehension of the initial thirty 'prompt' words (Tab. 1); S2 and S3 focusing on details of the questions and on the Likert scale and overall score. For each talk-aloud 90-minute session we used two researchers (one with expertise in information visualisation and the other in virtual reality). We explained the heuristics, they read the explanatory notes, and we video recorded the sessions and transcribed their comments. We used six different visualisations; two of each: line-graph, bar chart, parallel coordinate plot. Our goals were to ascertain if the strategy was easy, quick to

Table 1: The terms used in our card-sort exercise, for the 30 questions. Old words (~~striked~~), and new words after the talk-aloud exercise.

#1 ~~Perception~~Suitable	#11 ~~Interface~~GUI	#21 ~~Design~~Inspiring
#2 Understandable	#12 ~~Layout~~Ergonomic	#22 Aesthetic
#3 Assumptions	#13 Frame size	#23 ~~Space utilisation~~Composition
#4 Trustful	#14 Spacing	#24 Coverage
#5 Useful/utility	#15 Quantity	#25 Legend/labels
#6 ~~Environment~~Setting	#16 Components	#26 ~~Visual marks~~Channels
#7 Technology	#17 Type	#27 Mapping
#8 Interaction	#18 Relationship	#28 Mark types
#9 Output size	#19 Task suitability	#29 Scale/zoom
#10 ~~Place~~Ambience	#20 ~~Structure~~Layout	#30 Overplotting

complete and comprehensive. We made several changes from these talk-aloud evaluations. First, we adjusted some terms. E.g., (#10) *place* and *'ambience'*, and (#11) *interface* to *GUI*; Table 1 shows the ~~old~~ and new words. Second, participants said the 5-part Likert was useful, but took too long to calculate the result. They suggested to swap it from [1...5] to -2 to $+2$, with 0 being the mid-point. It is now quicker to sum the values, and the calculation gives an idea that the design was poor or good. Third, participants recommended adding a star plot to summarise the 30 scores. We implemented this in Version 1 and used it for the first two years, but removed it from Version 2 onwards due to the time it took to complete manually.

5.3 Analysis of Version 2, and developing V3 (2019-2023)

At this time the cohort size doubled, and the information visualisation module was incorporated into a wider range of programmes (data science and a general computing programme), attracting students with less advanced programming skills. As a result, we transitioned from having students develop an interactive visualisation of selected data (2018-2019) to designing a new visualisation for a poster display. This shift removed interaction from the coding process and placed a greater emphasis on design and storytelling (see Fig. 5). We continued to guide students in focusing their critical thinking on the three stages (❶ overview, ❷ detail, and ❸ review). Our lecture approach remained similar, covering topics such as visualisation history, design, and perception. In addition to a design lecture on the FDS method, we delivered an hour-long lecture on general critical thinking, followed by an explanation of the CDS structure. We maintained the two-assessment tasks: a design report and an implementation using Processing. Students were asked to complete a CDS sheet at least twice (once for each assessment), either by hand or digitally, and include a copy in their report. Most chose to complete it by hand and submitted a scan with their report.

By this point, we had been using the CDS in our teaching for two years, but we aimed to evaluate Version 2 more formally. To do so, we conducted two separate usability evaluation sessions with different student groups Group 1 (**G1**) had 10 PhD students from the Computer Science department; half had experience in creating visualisations with JavaScript and all had created charts in Excel. Group 2 (**G2**) were 20 BSc computer science major students, who were all taking the visualisation module and using processing.org to create visualisations. All participants were volunteers and none received compensation for participating. We gave a 15 minute introduction, consent form, printed CDS, explanation text of 30 categories, and six visualisation scenarios. They had an hour to reflect on each scenario and judge the visualisations. We chose the visualisations to include familiar and (potentially) unfamiliar forms. These were (i) a GapMinder bubble chart, (ii) an Excel pie chart, (iii) a treemap, (iv) a transport network map, (v) a time series graph, and (vi) an Excel bar chart. The written scenarios placed the visualisations in a particular context. E.g., the pie chart visualisation showed University annual research-grant income, and was to be included in the University's research report. Each participant was observed, and notes and timings recorded. Participants were asked to critique the visualisations in order. In **G1** eight participants completed six, wheras two completing five scenarios. In **G2**, only one participant did not finish within the hour. We note, that students in **G2** used the CDS three additional times, as they applied it in their teaching while submitting assessments 1 and 2.

To analyse the inter-item consistency (across PhD and undergraduate participants) we first considered all participants together (N=30). Cronbach's alpha values (Tab. 2) indicates that users did broadly answer in a similar, consistent way. Putting aside potential issues with Cronbach's α [74] where the calculation can be sensitive to the number of scales, we believe this is a positive result. To investigate the difference between **G1** (PhD) and **G2** (BSc) we used a t-test and made the dependent variable the total score of each visualisation. Tab. 3 shows the calculated significance ρ values for each visualisation. The ρ-value for the bubble, pie chart, treemap and bar chart visualisation are greater than the chosen significance level $a = 0.05$, so we don't reject H_0 the null hypothesis that there is no statistically significant difference (variance) of the means between users groups. However, the null hypothesis for equal variance can be rejected for the network map because ρ-value is less than 0.05. That means there is statistically significant difference for the mean values between groups. For timeline chart, ρ-value is equal to 0.05, yet the null hypothesis is rejected based on the variance between groups. In addition the associative SD for visualisations that has ρ-value < 0.05 have big mean difference between values (more than 4) from the average values to other visualisations.

Table 2: The internal-correlation of 30 users (N=30) users completing the categories on six different visualisations. Time taken for **G2** (N=20).

Visualisations	**Cronbach's** α	**Variance**	**Group2 time**, (sec)
Bubble chart	.89	0.080	15.6
Pie chart	.91	0.157	9.8
Treemap chart	.92	0.130	8.4
Network graph	.95	0.041	8.0
Time series graph	.92	0.068	7.0
Bar chart	.95	0.057	6.3

Table 3: Time taken for each visualisation. Independent samples t-test results show ρ-value at significance level a=0.05, associative mean and SD for each visualisation.

Visualisations	**Group**	**Mean**	**SD**	**Sig.** ρ **>0.05**
Bubble chart	1	102.6	18.4	.564
	2	108.0	13.2	
Pie chart	1	113.9	20.4	.623
	2	99.2	16.13	
Treemap chart	1	78.6	21.5	.728
	2	67.9	17.4	
Network graph	1	111.9	11.4	.007*
	2	90.6	25.17	
Time series graph	1	105.2	11.2	.051
	2	112.8	19.0	
Bar chart	1	102.8	19.2	.53
	2	122.0	17.4	

Figure 7 shows first-impression word use, across both groups. While there are some differences between groups, there is overall consensus. Words like *confusing*, *complex*, and *bad* were more frequently chosen, while terms like *spectacular* and *fulfilling* were less common. Since participants in **G1** engaged in more discussion, we focus on completion times for **G2**, shown in Tab. 2. Participants became faster with each use, reaching completion times of around 5 to 8 minutes. Finally, participants commented on positive and negative aspects of the CDS, and how the CDS helped them to critique their own visualisations. Twelve participants replied, saying *"it helps you see on paper what's good and what's bad"*. *"The CDS asks you to be honest and rate different properties of the program, which helps spot things that could be improved that would have been overlooked without CDS"*. Another wrote *"visually very easy to see how the design scored, and gives developers clear areas in which they can improve"*.

After analysis, we made improvements to the CDS. We removed the star plot, as it was time-consuming and not effectively completed. Participants (across **G1** and **G2**) requested clarification of some terms, possibly due to students not fully reading the explanatory descriptions. These changes are outlined in Tab. 1. Based on feedback and the words in the supplementary material, we introduced the semantic differential scale, inspired by the UEQ-style questionnaire. While we continued to emphasise the heuristic questions in the taught classes and notes, the semantic differential words helped quickly clarify the idea behind each of the 30 heuristics. Additionally, we revised the lecture material

Fig. 7: Analysis of the 20 first-impression words. The plot demonstrates a broad use of most words, apart from spectacular, fulfilling and beautiful.

to carefully explain each heuristic and expanded the supplementary notes. This led to Version 3, in which we removed the star plot and incorporated the semantic differential scale in the notes. We also rewrote the application to align with this version, and to allow users to review both current and past CDS evaluations (see Fig. 5).

5.4 CDS V.4 2023-present and its use

In our current practice, when integrating the CDS (as in Sec. 4) into assessments, we start with lectures on critical thinking and the CDS structure, and provide detailed notes (included in the supplementary material). Students then participate in a group activity [57], first collecting and discussing good and bad visualisations on a shared virtual whiteboard, then applying the CDS to two visualisations (one poor, one good). These exercises build critical thinking skills and with the CDS. Critique is inherently challenging and time-consuming to learn [33] (see Tab. 2), but the three-stage structure (overview, detail, review) and heuristic questions make it more manageable. With practice, efficiency improves. Similar to other years, students design a solution, critique it using the CDS, and submit a report with a scanned CDS. For the second assessment, they develop their tool in Processing and apply the CDS again. Writing the report enhances focus and deepens analysis.

6 Discussion and Conclusions

We asked the recent cohort (the BSc and MSc in information visualisation) about their experience, receiving 18 replies. The CDS framework significantly shaped critical thinking by providing a structured approach. One respondent said, "*It forced me to think about the nuances for my poster*"; another said "*The framework significantly enhanced my critical thinking by providing a structured approach to analyse my design*". The 30 sub-questions were mostly helpful, with one person noting, "*It became useful to find any shortcomings*". However, some found some heuristics less relevant, "Some sub-questions felt less useful, especially not relevant to my design goals". The first-impression words played a key role in shaping reflections. One participant said, "*It helped me streamline the presentation*" and another felt, "*It inspired me to think from a fresh perspective*". The "User" category was the most useful for many, with one saying, "*User environment, it made me think of the design from a different perspective*". "Components" and "Visual Marks" were also valued for tracking progress and evaluating visual elements. As one person put it, "*Components helped me keep track of my progress*". Some suggested changes, like simplifying questions. One respondent said, "*30 questions feels slightly too many*". Others suggested adding categories like "*Feedback & Iteration*" or "*Emotional Impact*" to assess user sentiment. Most participants plan to use the framework in the future. One said, "*Yes, I will continue using the framework*", while another noted, "*I would use it in my project report*". However, one wrote "*I would not continue using the framework in future design critiques because it feels somewhat rigid*". We recognise that the process takes time to complete, and is detailed, but this is intentional, as it encourages a thorough and in-depth critique of the design. Overall, the CDS framework was seen as valuable for critical thinking and design critique. As one person concluded, "*It was a helpful tool for reflection and improving my design process*".

After eight years of using the CDS in various contexts, we have gained valuable insights and identified key lessons. **Vocabulary and confidence are essential.** Many individuals struggle to critique because they don't have the vocabulary. To address this, we provide (i) a lecture on critical thinking, and emphasise specific vocabulary, particularly phrasing from Facione [33], (ii) class activities that provide a safe space to practice, and (iii) use repetition (requiring students to apply the CDS at least twice, once for each assessment). **Preparation is essential.** To effectively critique a visualisation, it is crucial to understand the data, purpose, and context. Without this foundation, evaluating its suitability becomes difficult. We support this by encouraging learners to consider the five W's and to imagine the intended user. A clear articulation of the design scenario and requirements by the instructor helps set the context, and when possible, having a client present a real-world problem adds valuable depth. Additionally, the "assign a name" task (Stage 1) sharpens focus, by creating a concise title learners can better identify the core purpose, offering a clear reference point for discussion.

Emphasise the process, not the answer. Some assessors may merely want a score to grade the design, while others want to write a fast critique. We emphasise to our students that they must spend time on the exercise and think carefully. The score is only one aspect of the strategy, and that the strategy and #30 questions are designed to prompt a deep consideration of the design. We emphasise that the CDS is a tool, to work alongside current design methods. It helps appraisers critique a variety of outputs. We do not want people to worry that they are not putting the right value on the form, rather, the CDS should be viewed as strategy to consider pros and cons with the visual design.

Write a report. Needing to write a report helps students focus on the process rather than the result. From our evaluation, we know this is occurring: e.g., one student wrote "*The CDS helped me with the final tool, and where I was going wrong on Sheet5 of the FdS design*". And another, when reflecting on a visualisation poster wrote "*the Design. Visual marks, Environment proved to be the lowest scoring sections, each with 6 points*", consequently they decided to change the colour palette, saying "*a more neutral palette may aid in rectifying this ... and more annotations*". Another wrote, "*the CDS was useful to show myself how much of a difference there is between the goal I set out to achieve and what I actually managed to do.*" and another " *It provided a very good structured way for me to critique ... [and consider] points I would not have thought of.*" **Adopt a broad-minded approach.** We encourage students to interpret the heuristics generally and apply them to their problem. E.g., question #8 about the appropriateness for the interaction, could be interpreted if it has WIMP controls; but should be generally applied to consider the concept of 'interaction' broadly.

The CDS is a three-stage critical thinking framework (overview, detail, review) with 30 core questions designed to help individuals critique their designs and visualisation tools. Our strategy fosters critical thinking skills, which are essential for making informed decisions, by guiding assessors in evaluating strengths, weaknesses, and areas for improvement. It provides a structured approach to deep thinking. To be most effective, the process should be given ample time, allowing for careful consideration of each aspect. We have used it in critical-thinking assessments, in both the design and implementation stages. Yet, it can be applied in different situations, and from several roles, including learners, researchers, and visualisation developers. We believe it is a valuable tool for tutors in teaching and assessing critical thinking. Ultimately, it empowers designers to make honest critiques and deeply reflect on their designs to create effective, engaging visualisation systems.

SUPPLEMENTAL MATERIALS

In the supplemental materials we include the full CDS explanatory notes, the final versions (with and without the differential scale) as resources for teaching. When the paper is accepted, we will make these resources open source, on ArXiv, and add links to these resources here.

ACKNOWLEDGMENTS

We acknowledge the help of the student learners across the years. Their substantial feedback (both positive and negative) has helped to improve and refine the CDS. We acknowledge the initial workshop participants, which ran as an evaluation as contribution for Alnjar's PhD [2]; while the latter evaluations received ethics clearance CSEE-2021-CG-001 and CSE-2025-0678.

REFERENCES

[1] H. M. Al-maneea and J. C. Roberts. Towards quantifying multiple view layouts in visualisation as seen from research publications. In *IEEE Visualization Conference (VIS), short papers*, 2019. doi: 10.1109/VISUAL.2019.8933655

[2] H. Alnjar. *Analysis and synthesis of critical design-thinking for data visualisation designers and learners*. PhD thesis, School of Computer Science, Bangor University, January 2017.

[3] R. Amar and J. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *Proc. InfoVis*, pp. 143–150, 2004. doi: 10.1109/INFVIS.2004.10

[4] C. Andrews and C. North. The impact of physical navigation on spatial organization for sensemaking. *IEEE TVCG*, 19:2207 – 2216, 12/2013 2013. doi: 10.1109/TVCG.2013.205

[5] K. Andrews. Evaluating information visualisations. In *Proc. AVI Workshop, BELIV*, pp. 1–5. ACM, 2006. doi: 10.1145/1168149.1168151

[6] F. J. Anscombe. Graphs in statistical analysis. *The American Statistician*, 27(1):17–21, 1973. doi: 10.2307/2682899

[7] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), Valletta, Malta, May 2010.

[8] B. Bach, M. Keck, F. Rajabiyazdi, T. Losev, I. Meirelles, J. Dykes, R. S. Laramee, M. AlKadi, C. Stoiber, S. Huron, C. Perin, L. Morais, W. Aigner, D. Kosminsky, M. Boucher, S. Knudsen, A. Manataki, J. Aerts, U. Hinrichs, J. C. Roberts, and S. Carpendale. Challenges and opportunities in data visualization education: A call to action. *IEEE Transactions on Visualization & Computer Graphics*, 30(01):649–660, jan 2024. doi: 10.1109/TVCG.2023.3327378

[9] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *J Hum-Comput Int*, 24(6):574–594, 2008. doi: 10.1080/10447310802205776

[10] J. Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983.

[11] J. B. Biggs. *Teaching for quality learning at university: What the student does*. McGraw-Hill Education (UK), 2011.

[12] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE TVCG*, 19(12):2306–2315, 2013. doi: 10.1109/TVCG.2013.234

[13] D. Borland and R. M. T. Ii. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27(2):14–17, March 2007. doi: 10.1109/MCG.2007.323435

[14] T. Borton. *Reach, Touch and Teach*. McGraw-Hill Inc., US, December 1970.

[15] R. Brath and E. Banissi. Evaluation of visualization by critiques. In *Proc. BELIV*, p. 19–26. ACM, 2016. doi: 10.1145/2993901.2993904

[16] B. Buxton. *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann, 2010.

[17] A. Calma and V. Cotronei-Baird. Assessing critical thinking in business education: Key issues and practical solutions. *The International Journal of Management Education*, 19(3):100531, 2021. doi: 10.1016/j.ijme.2021.100531

[18] C. Chen and M. P. Czerwinski. Empirical evaluation of information visualizations: an introduction. *Int. J. Hum. Comput.*, 53(5):631–635, 2000. doi: doi.org/10.1006/ijhc.2000.0421

[19] C. Chen and Y. Yu. Empirical studies of information visualization. *International Journal of Human-Computer Studies*, 53(5):851–866, Nov. 2000. doi: 10.1006/ijhc.2000.0422

[20] X. Chen, W. Zeng, Y. Lin, H. M. AI-maneea, J. Roberts, and R. Chang. Composition and configuration patterns in multiple-view visualizations. *IEEE TVCG*, 27(2):1514–1524, 2021. doi: 10.1109/TVCG.2020.3030338

[21] W. S. Cleveland. A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics*, 2(4):323–343, 1993. doi: 10.1080/10618600.1993.10474616

[22] S. Conversy, S. Chatty, and C. Hurter. Visual scanning as a reference framework for interactive representation design. *Information Visualization*, 10(3):196–211, 2011. doi: 10.1177/1473871611415988

[23] D. Council. Eleven lessons: managing design in eleven global companies desk research report, 2007. www.designcouncil.org.uk.

[24] M. R. de Oliveira and C. G. da Silva. Adapting heuristic evaluation to information visualization - a method for defining a heuristic set by heuristic grouping. In *International Conference on Information Visualization Theory and Applications*, vol. 4, pp. 225–232. SCITEPRESS, 2017.

[25] J. Dewey. *How we think*. D. C. Heath & Co., Boston, New York, Chicago, 1910.

[26] D. Dowding and J. A. Merrill. The development of heuristics for evaluation of dashboard visualizations. *Applied clinical informatics*, 9(03):511–518, 2018. doi: 10.1055/s-0038-1666842

[27] K. Duncker and L. S. Lees. On problem-solving. *Psychological monographs*, 58(5):i, 1945. doi: 10.1037/h0093599

[28] J. Dykes, J. Wood, and A. Slingsby. Rethinking map legends with visualization. *IEEE TVCG*, 16(6):890–899, 2010. doi: 10.1109/TVCG.2010.191

[29] L. Engelbrecht, A. Botha, and R. Alberts. Information visualisation view design: principles and guidelines. In *Proceedings of the International Conference on Computer Science, Computer Engineering, and Social Media, Thessaloniki, Greece*, 2014.

[30] R. Ennis. Critical thinking: Reflection and perspective part ii. *Inquiry: Critical thinking across the Disciplines*, 26(2):5–19, 2011.

[31] R. H. Ennis. A concept of critical thinking. *Harvard educational review*, 1962.

[32] M. J. Eppler and R. A. Burkhard. Visual representations in knowledge management: framework and cases. *Journal of knowledge management*, 11(4):112–122, 2007. doi: 10.1108/13673270710762756

[33] P. A. Facione. Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. research findings and recommendations. *ERIC*, ED315423, 1990.

[34] C. Forsell. Evaluation in information visualization: Heuristic evaluation. In *2012 16th International Conference on Information Visualisation*, pp. 136–142, 2012. doi: 10.1109/IV.2012.33

[35] C. Forsell and J. Johansson. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '10, p. 199–206. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1842993.1843029

[36] D. F. Halpern. Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New directions for teaching and learning*, 1999(80):69–74, 1999. doi: 10.1002/tl.8005

[37] M. Harrower and C. A. Brewer. Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 2013. doi: 10.1002/9780470979587.ch34

[38] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller. A systematic review on the practice of evaluating visualization. *IEEE TVCG*, 19(12):2818–2827, Dec 2013. doi: 10.1109/TVCG.2013.126

[39] B. Jackson, D. Coffey, L. Thorson, D. Schroeder, A. M. Ellingson, D. J. Nuckley, and D. F. Keefe. Toward mixed method evaluations of scientific visualizations and design process as an evaluation tool. In *Proc BELIV*, pp. 4:1–4:6. ACM, New York, NY, USA, 2012. doi: 10.1145/2442576.2442580

[40] N. W. Kim, S. C. Joyner, A. Riegelhuth, and Y. Kim. Accessible visualization: Design space, opportunities, and challenges. In *Computer Graphics Forum*, vol. 40, pp. 173–188. Wiley Online Library, 2021. doi: 10.1111/cgf.14298

[41] D. A. Kolb. *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[42] R. Kosara. Visualization criticism the missing link between information visualization and art. In *Proceedings of the 2007 International Conference on Information Visualization, IV '07.*, pp. 631–636, July 2007. doi: 10.1109/IV.2007.130

[43] R. Kosara, F. Drury, L. E. Holmquist, and D. H. Laidlaw. Visualization criticism. *IEEE Computer Graphics and Applications*, 28(3):13–15, May 2008. doi: 10.1109/MCG.2008.63
[44] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE TVCG*, 18(9):1520–1536, Sept. 2012. doi: 10.1109/TVCG.2011.279
[45] B. Laugwitz, T. Held, and M. Schrepp. *Construction and Evaluation of a User Experience Questionnaire*, pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-89350-9_6
[46] G. L. Lohse. A cognitive model for understanding graphical perception. *Human Computer Interaction*, 8(4):353–388, Dec. 1993. doi: 10.1207/s15327051hci0804_3
[47] J. Lohse, H. Rueter, K. Biolsi, and N. Walker. Classifying visual knowledge representations: a foundation for visualization research. In *Proceedings of the First IEEE Conference on Visualization*, pp. 131–138. IEEE, 1990. doi: 10.1109/VISUAL.1990.146374
[48] T. Munzner. A nested model for visualization design and validation. *IEEE TVCG*, 15(6):921–928, Nov 2009. doi: 10.1109/TVCG.2009.111
[49] T. Munzner. *Visualization Analysis and Design*. A K Peters/CRC Proess, 2014.
[50] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pp. 249–256. ACM, New York, NY, USA, 1990. doi: 10.1145/97243.97281
[51] C. O'Reilly, A. Devitt, and N. Hayes. Critical thinking in the preschool classroom - a systematic literature review. *Thinking Skills and Creativity*, 46:101110, 2022. doi: 10.1016/j.tsc.2022.101110
[52] R. Paul and L. Elder. *The miniature guide to critical thinking concepts and tools*. Dillon Beach, CA: Foundation for Critical Thinking Press, 2009.
[53] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, pp. 109–116. ACM, New York, NY, USA, 2004. doi: 10.1145/989863.989880
[54] G. Pólya. *How to solve it: A new aspect of mathematical method*. Princeton university press, 1973. republished 2014.
[55] D. Rams. Ten principles for good design. www.vitsoe.com, 1970.
[56] J. Roberts, H. Alnjar, A. Owen, and P. Ritsos. A method for critical and creative visualisation design-thinking. In *In IEEE VIS 2023 Posters: Visualization & Visual Analytics*, 2023.
[57] J. C. Roberts, B. Bach, M. Boucher, F. Chevalier, A. Diehl, U. Hinrichs, S. Huron, A. Kirk, S. Knudsen, I. Meirelles, et al. Reflections and considerations on running creative visualization learning activities. In *4th IEEE Workshop on Visualization Guidelines in Research, Design, and Education*, 2022. doi: 10.1109/VisGuides57787.2022.00009
[58] J. C. Roberts, C. Headleand, and P. D. Ritsos. Sketching designs using the five design-sheet methodology. *IEEE TVCG*, January 2016. doi: 10.1109/TVCG.2015.2467271
[59] J. C. Roberts, C. J. Headleand, and P. D. Ritsos. *Five Design-Sheets – Creative design and sketching in Computing and Visualization*. Springer-Nature, 2017. doi: 10.1007/978-3-319-55627-7
[60] J. C. Roberts and P. D. Ritsos. Critical Thinking Sheet (CTS) for Design Thinking in Programming Courses. In M. Romero and B. Sousa Santos, eds., *Eurographics 2020 - Education Papers*. The Eurographics Association, 2020. doi: 10.2312/eged.20201029
[61] J. C. Roberts, P. D. Ritsos, J. R. Jackson, and C. Headleand. The explanatory visualization framework: An active learning framework for teaching creative computing using explanatory visualizations. *IEEE TVCG*, 24(1):791–801, Jan 2018. doi: 10.1109/TVCG.2017.2745878
[62] Y. Rogers, H. Sharp, and J. Preeece. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, Inc., 2011.
[63] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE TVCG*, 11(4):443–456, July 2005. doi: 10.1109/TVCG.2005.53
[64] J. Scholtz. Developing guidelines for assessing visual analytics environments. *Information Visualization*, 10(3):212–231, 2011. doi: 10.1177/1473871611407399
[65] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE TVCG*, 18(12):2431–2440, Dec. 2012. doi: 10.1109/TVCG.2012.213
[66] J. Selzer. *Rhetorical analysis: Understanding how texts persuade readers*, pp. 279–307. Lawrence Erlbaum Associates, publishers, Mahwah, New Jersey, 2004.
[67] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pp. 336–343. IEEE, 1996. doi: 10.1016/B978-155860915-0/50046-9
[68] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proc AVI, BEyond time and errors*, pp. 1–7. ACM, 2006. doi: 10.1145/1168149.1168158
[69] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, and N. Elmqvist. *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Sixth Edition*. Pearson, May 2016.
[70] J. Stasko. Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pp. 46–53, 2014. doi: 10.1145/2669557.2669579
[71] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988. doi: 10.1207/s15516709cog1202_4
[72] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307, 2011. doi: 10.1162/COLI_a_00049
[73] A. Tarrell, A. Fruhling, R. Borgo, C. Forsell, G. Grinstein, and J. Scholtz. Toward visualization-specific heuristic evaluation. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pp. 110–117, 2014. doi: 10.1145/2669557.2669580
[74] M. Tavakol and R. Dennick. Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2:53–55, 2011. doi: 10.5116/ijme.4dfb.8dfd
[75] M. Tory and T. Moller. Evaluating visualizations: do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5):8–11, 2005. doi: 10.1109/MCG.2005.102
[76] D. F. Treagust and A. G. Harrison. In search of explanatory frameworks: an analysis of richard feynman's lecture 'atoms in motion'. *International Journal of Science Education*, 22(11):1157–1170, 2000. doi: 10.1080/09500690050166733
[77] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT, 1983.
[78] H. Väätäjä, J. Varsaluoma, T. Heimonen, K. Tiitinen, J. Hakulinen, M. Turunen, H. Nieminen, and P. Ihantola. Information visualization heuristics in practical expert evaluation. In *Proc. BELIV*, pp. 36–43, 2016. doi: 10.1145/2993901.2993918
[79] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A Heuristic Approach to Value-Driven Evaluation of Visualizations. *IEEE TVCG*, 25(1):491–500, Jan 2019. doi: 10.1109/TVCG.2018.2865146
[80] C. Ware. *Information visualization (Interactive Technologies)*. MK, 2012.
[81] C. Ware. *Information Visualization: Perception for Design*. Elsevier (Morgan Kaufmann), Amsterdam, 2012.
[82] T. Zuk and S. Carpendale. Theoretical analysis of uncertainty visualizations. In *Visualization and data analysis 2006*, vol. 6060, pp. 66–79. SPIE, 2006. doi: 10.1117/12.643631